

A NONPARAMETRIC MULTIDIMENSIONAL LATENT CLASS IRT MODEL IN A BAYESIAN FRAMEWORK

FRANCESCO BARTOLUCCI 

UNIVERSITÀ DI PERUGIA

ALESSIO FARCOMENI 

SAPIENZA - UNIVERSITÀ DI ROMA

LUISA SCACCIA 

UNIVERSITÀ DI MACERATA

We propose a nonparametric item response theory model for dichotomously-scored items in a Bayesian framework. The model is based on a latent class (LC) formulation, and it is multidimensional, with dimensions corresponding to a partition of the items in homogenous groups that are specified on the basis of inequality constraints among the conditional success probabilities given the latent class. Moreover, an innovative system of prior distributions is proposed following the encompassing approach, in which the largest model is the unconstrained LC model. A reversible-jump type algorithm is described for sampling from the joint posterior distribution of the model parameters of the encompassing model. By suitably post-processing its output, we then make inference on the number of dimensions (i.e., number of groups of items measuring the same latent trait) and we cluster items according to the dimensions when unidimensionality is violated. The approach is illustrated by two examples on simulated data and two applications based on educational and quality-of-life data.

Key words: cluster analysis, encompassing priors, item response theory, unidimensionality, Markov chain Monte Carlo, reversible-jump algorithm, stochastic partitions.

1. Introduction

A crucial aspect in the analysis of item responses is that of dimensionality, corresponding to the number of latent traits measured by the test items. Typically, item response theory (IRT) models (Hambleton & Swaminathan, 1985; Bartolucci et al., 2015) assume unidimensionality, that is, the test items measure a single dimension that in education corresponds to the ability in a certain domain. Obviously, this assumption may lead to misleading results when it is not realistic. This is particularly true for the Rasch model (Rasch, 1961), the main IRT model for dichotomous items, under which the raw score is a sufficient statistic for the ability level of an examinee. In fact, if unidimensionality does not hold, summarizing his/her performance by this simple statistic is not correct.

The problem of testing unidimensionality has attracted a great interest in the psychometric literature. One of the main contributions in this literature is due to Martin-Löf (1973), who proposed a test of unidimensionality against bidimensionality under the assumptions of the Rasch model. The main advantage of the test is its simple use, as it is based on conditional maximum likelihood estimates of the item parameters (under the unidimensional and bidimensional hypotheses), which may be easily obtained by using different softwares. Furthermore, this approach does not require

Correspondence should be made to Francesco Bartolucci, Dipartimento di Economia, Università di Perugia, Via A. Pascoli 20, 06123, Perugia, Italy. Email: francesco.bartolucci@unipg.it

to formulate assumptions on the distribution of the latent traits. On the other hand, the assumptions that are formulated on the conditional distribution of the item responses given the latent traits may be restrictive. Mainly, it is required that all items have the same discriminating power, leading the test to over-rejecting unidimensionality when this assumption does not hold. Moreover, it does not allow us to measure the correlation between the two latent traits assumed to exist under the most general model. See Verhelst (2001) for further comments about the Martin-Löf (1973)'s test and an analysis of its performance.

More recently, two interesting approaches were proposed by Christensen et al. (2002) and Bartolucci (2007). The first may be seen as an extension of the approach of Martin-Löf (1973) to the case of polytomous items. In their proposal, Christensen et al. (2002) also considered marginal maximum likelihood (MML) estimation based on the assumption of normality of the latent traits. In particular, under the bidimensional model, a bivariate normal distribution with arbitrary correlation is assumed, which may be estimated on the basis of the data. On the other hand, the approach of Bartolucci (2007), extended to the case of polytomous items by Bacci et al. (2014), is based on MML estimation under the assumption that the latent traits have a discrete distribution that defines a certain number of latent classes of subjects (Lazarsfeld & Henry, 1968; Goodman, 1974). Then, the approach may be seen as semiparametric, in the sense of Lindsay et al. (1991). Moreover, in Bartolucci (2007) a two-parameter logistic (2PL) parametrization (Birnbaum, 1968) is adopted and an algorithm for discovering the number of dimensions, and clustering the items in separate groups corresponding to these dimensions, is provided. See also von Davier (2008) for a related class of IRT models with a high degree of flexibility that may be used to analyze multidimensional item responses.

The 2PL parametrization, on which the approach of Bartolucci (2007) relies, is more flexible than the Rasch parametrization of the item response distribution given the underlying latent trait; in fact, it allows for different discrimination levels among items. However, the assumption that a group of items measures the same latent trait may be defined in a nonparametric fashion through inequality constraints under the assumption that the latent distribution is discrete (Bartolucci & Forcina, 2005), which amounts to assume a latent class (LC) model. In this article, and for the case of dichotomously-scored items, we develop this result to define an IRT model that allows us to discover the number of dimensions measured by the questionnaire, without making any parametric assumption on either the distribution of the latent traits or the conditional distribution of the item responses given such traits.

In order to estimate the proposed model, we rely on a Bayesian inference framework that we consider convenient given the uncertainty on both the number of latent classes and model dimensionality. Therefore, we specify a new system of priors and develop suitable algorithms for parameter estimation. In particular, for the prior distributions on the model parameters we rely on the encompassing approach of Klugkist et al. (2005). Accordingly, we formulate priors on the parameters of the most general model, which is an unconstrained LC model with a certain number of classes. Then, priors are "automatically" defined for any nested model formulated by suitable inequalities on the conditional success probabilities in order to specify the number of dimensions. Regarding estimation, we rely on an algorithm of reversible-jump Markov chain Monte Carlo (RJ-MCMC) type (Green, 1995; Green & Richardson, 2001) applied to the LC model and followed by a suitable post-processing procedure to cluster items. As a result, the proposed approach allows us to select the number of latent classes and the number of dimensions, and also provides a clustering of the items into disjoint groups measuring different latent traits.

Overall, the framework here proposed is related to the Bayesian inference approach of Hoijtink and Molenaar (1997) that may be used to test different assumptions about IRT models. Their approach is based on a suitably constrained LC model to express the hypothesis of multidimensionality in a way similar to that we follow in the present paper, whereas the IRT assumptions are tested on the basis of suitably formulated statistics. The approach of Hoijtink and Molenaar (1997),

as well as Bayesian inference for the constrained LC models at issue, has been developed in several directions; see, among others, Béguin and Glas (2001), Van Onna (2002), and Bolt and Lall (2003), and the recent work by Kuo and Sheng (2015). With respect to these Bayesian approaches, our approach is based on a innovative and flexible set of priors on the model parameters of the unrestricted LC model, which permits to arbitrarily specify the prior probability of the hypothesis of unidimensionality. Moreover, we present an innovative approach of model selection, in terms of number of latent classes and number of dimensions, which is directly connected with the encompassing approach of Klugkist et al. (2005). On the other hand, as already mentioned above, the proposed approach is also innovative with respect to parametric or semiparametric multidimensional IRT approaches typically based on maximum likelihood inference, as those described in Christensen et al. (2002) and Bartolucci (2007), and related papers. This is because we reduce the number of assumptions and avoid the specification of a parametric item response function. In this regard, it is also worth considering the approaches described in Junker and Sijtsma (2001), Vermunt (2001), and Karabatsos (2001).

The article is organized as follows. In the next section we formulate the nonparametric IRT model, corresponding to a constrained LC model, which is used to evaluate the dimensionality of a set of items. Then, we illustrate Bayesian inference for this model based on the RJ-MCMC algorithm, paying particular attention to the formulation of the prior distributions of the model parameters. This approach to assess dimensionality is illustrated through empirical examples based on both simulated and real data. The final section contains main conclusions.

We implemented the proposed approach in a series of MATLAB functions that we make available to the reader upon request.

2. Model Formulation

Let Y_{ij} , $i = 1, \dots, n$, $j = 1, \dots, r$, denote the random variable corresponding to the binary response provided by the i -th subject to the j -th item. In the following, we briefly review the LC model for the distribution of these variables and then we introduce constrained versions of this model, which are formulated through suitable inequality constraints and may be seen as nonparametric IRT models. Furthermore, we compare this approach with that proposed by Bartolucci (2007), showing that the present one is more general.

2.1. Unconstrained Latent Class Model

The LC model assumes that the sample of respondents is drawn from a population divided into k latent classes, with individuals in the same class sharing the same distribution of the response variables (Lazarsfeld & Henry, 1968). Thus, for each subject i , the latent traits of interest, which in the educational setting correspond to certain types of ability, are represented by a discrete latent variable C_i having k support points denoted, without loss of generality, by $1, \dots, k$. The model also assumes *local independence* (LI), that is, the response variables Y_{i1}, \dots, Y_{ir} are conditionally independent given C_i . Consequently, the model parameters are the class weights π_1, \dots, π_k , with $\pi_c = p(C_i = c)$, and $\lambda_{jc} = p(Y_{ij} = 1 | C_i = c)$ corresponding to the probability of success for the j -th item given class c , with $c = 1, \dots, k$ and $j = 1, \dots, r$. These parameters are common to all individuals, and then the previous definitions are independent of the specific i .

Before introducing constrained formulations of the LC model defined above, we recall the observed and complete data log-likelihood functions of this unconstrained model, which are required for inference. With reference to an observed matrix of data \mathbf{Y} , with elements y_{ij} , $i = 1, \dots, n$, $j = 1, \dots, r$, corresponding to realizations of the random variables Y_{ij} , the *observed log-likelihood* is defined as

$$\ell(\mathbf{Y}|\mathbf{\Lambda}, \boldsymbol{\pi}) = \sum_{i=1}^n \log \left[\sum_{c=1}^k \pi_c \prod_{j=1}^r \lambda_{jc}^{y_{ij}} (1 - \lambda_{jc})^{1-y_{ij}} \right], \quad (1)$$

where $\mathbf{\Lambda}$ is the $r \times k$ dimensional matrix of the conditional probabilities λ_{jc} and $\boldsymbol{\pi}$ is the k -dimensional vector of the class weights π_c . Note that, due to the LI assumption, $\prod_{j=1}^r \lambda_{jc}^{y_{ij}} (1 - \lambda_{jc})^{1-y_{ij}}$ in (1) corresponds to the conditional probability of the responses provided by subject i given that he/she belongs to latent class c ; these responses are collected in the vector \mathbf{y}_i .

In order to define the *complete data log-likelihood*, we introduce the latent class indicators z_{ic} , $i = 1, \dots, n$, $c = 1, \dots, k$, where $z_{ic} = 1$ if the i -th subject belongs to latent class c and $z_{ic} = 0$ otherwise; see for instance [Diebolt and Robert \(1994\)](#). Then, augmenting the data with these indicator variables, we write the complete log-likelihood as

$$\ell_C(\mathbf{Y}, \mathbf{Z}|\mathbf{\Lambda}, \boldsymbol{\pi}) = \ell(\mathbf{Y}|\mathbf{Z}, \mathbf{\Lambda}) + \ell(\mathbf{Z}|\boldsymbol{\pi}).$$

The first addend at the right hand side is equal to

$$\sum_{c=1}^k \sum_{i=1}^n \sum_{j=1}^r z_{ic} [y_{ij} \log \lambda_{jc} + (1 - y_{ij}) \log(1 - \lambda_{jc})],$$

as only the Bernoulli log-likelihoods corresponding to $z_{ic} = 1$ must be considered. The second addend corresponds to the usual log-likelihood of multinomial distributions. Hence the complete, or augmented, log-likelihood can be written as

$$\ell_C(\mathbf{Y}, \mathbf{Z}|\mathbf{\Lambda}, \boldsymbol{\pi}) = \sum_{c=1}^k \sum_{i=1}^n \sum_{j=1}^r z_{ic} [y_{ij} \log \lambda_{jc} + (1 - y_{ij}) \log(1 - \lambda_{jc})] + \sum_{c=1}^k z_{ic} \log \pi_c. \quad (2)$$

In the following, an LC model based on k classes is denoted by LC(k) when its parameters λ_{jc} are unconstrained.

2.2. Constrained Latent Class Model

In the proposed nonparametric IRT setting and using the previous notation, two items, j_1 and j_2 , are said to measure the same dimension if there exists a permutation of the indices $1, \dots, k$, denoted by $\mathbf{c} = (c_1, \dots, c_k)$, such that

$$\lambda_{jc_1} < \dots < \lambda_{jc_k}, \quad j = j_1, j_2. \quad (3)$$

In other words, the success probabilities of the two items are ordered in the same way. This property may be referred to a set of items with size larger than two and may be easily understood considering the case of $k = 2$ latent classes. In this case, items j_1 and j_2 measure the same latent trait if the classes are ordered in the same way in terms of probability of success with respect to both items, that is, if

$$\lambda_{j_11} < \lambda_{j_12} \quad \text{and} \quad \lambda_{j_21} < \lambda_{j_22}$$

or

$$\lambda_{j_11} > \lambda_{j_12} \quad \text{and} \quad \lambda_{j_21} > \lambda_{j_22}.$$

On the basis of the above arguments, we define an s -dimensional nonparametric IRT model as an LC model for which a partition $\mathcal{P} = \{\mathcal{J}_1, \dots, \mathcal{J}_s\}$ of the full set of items exists such that:

- A1. condition (3) holds for every pair of items $(j_1, j_2) \in \mathcal{J}_d$, $d = 1, \dots, s$, based on a unique permutation $\mathbf{c}^{(d)} = (c_1^{(d)}, \dots, c_k^{(d)})$;
- A2. $\mathbf{c}^{(d_1)} \neq \mathbf{c}^{(d_2)}$, for all $d_1, d_2 = 1, \dots, s$, with $d_1 \neq d_2$.

In practice, any pair of items (j_1, j_2) is assumed to measure the same dimension, so that the latent classes are ordered in the same way according to the conditional success probabilities $\lambda_{j_1 c}$ and $\lambda_{j_2 c}$, if the items are in the same subset \mathcal{J}_d . On the other hand, these conditional success probabilities are ordered in a different way when the items belong to two different subsets \mathcal{J}_{d_1} and \mathcal{J}_{d_2} , and then they measure two different latent traits (or dimensions). An obvious consequence is that, under the unidimensional model, the latent classes are ordered in the same way according to all items. Moreover, there is a relation between the number of latent classes k and the model dimension s . In particular, it may be simply proved that $s \leq \min(k!, r)$, where $k! = k(k-1) \dots 1$ is the number of possible permutations of k elements.

Under assumptions A1 and A2, the observed log-likelihood for the constrained LC model has the same expression as in (1) and the corresponding complete log-likelihood is as in (2). Moreover, it is clear that any nonparametric IRT model defined above depends on k and a specific partition of items \mathcal{P} made of s subsets $\mathcal{J}_1, \dots, \mathcal{J}_s$. This model is then denoted by $\text{LC}(k, \mathcal{P})$ and we also use the symbol $|\mathcal{P}|$ for the number of subsets in partition \mathcal{P} . In this regard it is worth noting that our definition, based on condition (3), rules out the case of equality between two probability parameters $\lambda_{j c_1}$ and $\lambda_{j c_2}$ for $c_1, c_2 = 1, \dots, k$, with $c_1 \neq c_2$. This constraint is necessary to avoid a possible ambiguity due to the lack of uniqueness of $\mathbf{c}^{(d)}$ considered in assumption A1 for some dimension d . However, such a constraint has no practical relevance in our Bayesian inferential framework as the probability that $\lambda_{j c_1} = \lambda_{j c_2}$ is equal to 0 for any j and any pair (c_1, c_2) .

2.3. Comparison with the Semiparametric Approach

The previous characterization of a model measuring s dimensions is completely nonparametric, in contrast with the one in Bartolucci (2007) that is based on a parametric formulation for the success probabilities $\lambda_{j c}$ of 2PL type. The latter amounts to assume that

$$\text{logit}(\lambda_{j c}) = \gamma_j \left(\sum_{d=1}^s \delta_{j d} \theta_{c d} - \phi_j \right), \quad c = 1, \dots, k, \quad j = 1, \dots, r, \quad (4)$$

where $\delta_{j d}$ is an indicator of item grouping, which is equal to 1 if item j measures dimension d ($j \in \mathcal{J}_d$) and to 0 otherwise. Moreover, $\theta_{c d}$ corresponds to the ability of the subjects in class c in answering correctly to the items measuring dimension d , while γ_j and ϕ_j may be interpreted as the discriminating power and the difficulty of the j -th item, respectively.

In order to prove that, being based on parametric assumption (4), the formulation in Bartolucci (2007) is less general than ours, suppose that $\delta_{j_1 d} = \delta_{j_2 d} = 1$, that is, items j_1 and j_2 belong to the same group \mathcal{J}_d , and for both of them such an assumption is satisfied. Suppose, without loss of generality, that $k = 2$ and that $\theta_{1 d} < \theta_{2 d}$. A simple consequence is that

$$\gamma_j (\theta_{1 d} - \phi_j) < \gamma_j (\theta_{2 d} - \phi_j), \quad j = j_1, j_2,$$

implying that

$$\text{logit}(\lambda_{j_1}) < \text{logit}(\lambda_{j_2}) \iff \lambda_{j_1} < \lambda_{j_2}, \quad j = j_1, j_2;$$

that is, the items belong to the same group also according to our characterization based on (3). On the other hand, it is obvious that even if for two items this inequality holds, this does not imply that the corresponding conditional success probabilities satisfy (4).

Note that the same reasoning applies, for instance, to the approaches of Martin-Löf (1973) and Christensen et al. (2002), which are even more constrained, being based on a Rasch parametrization of the conditional success probabilities, which is a particular case of (4) with all parameters γ_j equal to 1.

3. Bayesian Inference

In this section, we introduce a flexible class of prior distributions for the parameters of the LC models defined in the previous section, with particular attention to the conditional success probabilities given the latent class, λ_{jc} . Consequently, we illustrate the RJ-MCMC algorithm that is adopted to simulate the joint posterior distribution of the model parameters and that is used for Bayesian estimation and testing, after a suitable post-processing of its output.

3.1. Prior Distributions

It is clear that any constrained LC model (or equivalently nonparametric IRT model) of type $LC(k, \mathcal{P})$ is nested in the model $LC(k)$ having the same number of classes and in which the probabilities λ_{jc} are left unconstrained. Then, once the priors have been specified for the latter model, we can “automatically” specify those of any nested model through the *encompassing approach* (Klugkist et al., 2005): prior distributions for nested models are derived by truncating the parameter space according to the constraints of interest; see also Bartolucci et al. (2012). In practice, let $p(\Lambda, \boldsymbol{\pi} | LC(k))$ represent the encompassing prior, that is the prior specified for the encompassing $LC(k)$ model. The prior distribution of any nested model $LC(k, \mathcal{P})$ follows directly as

$$p(\Lambda, \boldsymbol{\pi} | LC(k, \mathcal{P})) = \frac{p(\Lambda, \boldsymbol{\pi} | LC(k)) I\{\Lambda, \boldsymbol{\pi} | LC(k, \mathcal{P})\}}{\int \int p(\Lambda, \boldsymbol{\pi} | LC(k)) I\{\Lambda, \boldsymbol{\pi} | LC(k, \mathcal{P})\} d\Lambda d\boldsymbol{\pi}}, \quad (5)$$

where the indicator function $I\{\Lambda, \boldsymbol{\pi} | LC(k, \mathcal{P})\}$ has value 1 if its argument is true (i.e., the parameter values are in accordance with the constraints imposed by model $LC(k, \mathcal{P})$) and 0 otherwise. In other words, the prior of a certain constrained model is obtained by imposing a zero value to the prior of the encompassing model for all parameter values that do not respect the constraints of interest. This prior is then normalized to integrate to 1 over the restricted parameter space.

For the encompassing model, Bayes–Laplace priors are a natural choice for the success probabilities and class weights (Tuyl et al., 2009). This choice reduces to a uniform prior between 0 and 1 for λ_{jc} , $j = 1, \dots, r$, $c = 1, \dots, k$, while for the class weights π_c it corresponds to a Dirichlet distribution with vector of hyper-parameters having all elements equal to 1. Finally, a discrete uniform prior can be used for the number of classes k in the discrete set $1, \dots, k_{\max}$, where k_{\max} is the maximum value of k that is allowed. With small samples, large values of k could be penalized by using a Poisson prior with a small hyper-parameter value. Note however that, while we retain the standard choice for the priors of parameters k and π_c , we adopt a different system of priors for the success probabilities λ_{jc} , as illustrated in detail in the following.

3.2. A Flexible Class of Priors for the Success Probabilities

It is worth noting that, according to the encompassing approach, the prior on the success probabilities λ_{jc} “automatically” determines the prior probability of any specific partition \mathcal{P} , as well as the prior distribution of the number of groups s . In fact, the prior probability $p(\mathcal{P} | k)$

TABLE 1.

Conditional and unconditional prior distribution of the number of partitions, as determined by a uniform prior on the success probabilities, when $r = 12$ and $k_{\max} = 10$.

k	s											
	1	2	3	4	5	6	7	8	9	10	11	12
1	1.0000	–	–	–	–	–	–	–	–	–	–	–
2	0.0005	0.9995	–	–	–	–	–	–	–	–	–	–
3	0.0000	0.0000	0.0048	0.1011	0.4563	0.4378	–	–	–	–	–	–
4	0.0000	0.0000	0.0000	0.0000	0.0002	0.0035	0.0300	0.1291	0.2894	0.3323	0.1801	0.0355
5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0006	0.0095	0.0805	0.3430	0.5664
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0031	0.0849	0.9120
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0130	0.9870
8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0016	0.9984
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.9998
10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
$p(s)$	0.1000	0.1000	0.0005	0.0101	0.0456	0.0441	0.0030	0.0130	0.0299	0.0416	0.0623	0.5499

of a specific partition \mathcal{P} is simply the double integral at the denominator of (5), while $p(s|k)$ can be obtained by summing up the probabilities $p(\mathcal{P}|k)$ over all those partitions for which $|\mathcal{P}| = s$. Notice that under a uniform prior for λ_{jc} , $p(\mathcal{P}|k)$ and $p(s|k)$ can be easily calculated; see “Appendix” for details.

The above point implies that the choice of the prior on the λ_{jc} parameters deserves some more reasoning, since it might favor a certain partition or a certain number of groups. Despite the simplicity of a uniform prior on these parameters and the consequent simplifications for the RJ-MCMC algorithm, such a choice might be not completely meaningful. This is particularly true if the main purpose of the analysis is testing the hypothesis of unidimensionality, as a uniform prior on the λ_{jc} determines a prior on s which often assigns a small probability to the null hypothesis. To clarify this point, Table 1 shows the prior distribution for s , as determined by a uniform prior on the λ_{jc} , both conditionally on k and marginally, with reference to a hypothetic set of $r = 12$ items (the same number of items as one of the datasets considered in our applications) and setting $k_{\max} = 10$. The values in the table have been calculated as explained in “Appendix.”

An unconditional prior probability of 0.1 is given to the hypothesis of unidimensionality under a uniform prior on the λ_{jc} parameters. To make the point even clearer, consider that for $r = 3$ and $k = 2$, we would have $p(s = 1|k = 2) = 0.25$, meaning that, even with just three items and two latent classes, the probability that all items belong to the same dimension would be quite small.

For the reasons above, we propose a flexible class of priors for the success probabilities λ_{jc} , which is defined as

$$\lambda_{jc} \sim \text{Beta}(\alpha_c, \beta_c), \quad c = 1, \dots, k, \quad j = 1, \dots, r, \tag{6}$$

where the parameters α_c and β_c are given by

$$\alpha_c = v_\alpha c + 1, \quad \beta_c = v_\beta (k + 1 - c) + 1, \quad c = 1, \dots, k, \tag{7}$$

with v_α and v_β being hyper-parameters with values to be appropriately chosen. With such a prior, the expected value of λ_{jc} increases with c , while its variance is approximately the same for all c and rapidly decreases as k increases. In order to choose v_α and v_β , it should be noted that

TABLE 2.

Simulated conditional and unconditional prior distribution of the number of partitions, as determined by a Beta prior with $v_\alpha = v_\beta = 3$ on the success probabilities, when $r = 12$ and $k_{\max} = 10$.

k	s											
	1	2	3	4	5	6	7	8	9	10	11	12
1	1.0000	–	–	–	–	–	–	–	–	–	–	–
2	0.3255	0.6745	–	–	–	–	–	–	–	–	–	–
3	0.0454	0.3666	0.5411	0.0460	0.0009	0.0000	–	–	–	–	–	–
4	0.0041	0.0644	0.3108	0.4294	0.1647	0.0241	0.0025	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0001	0.0059	0.0580	0.2188	0.3442	0.2495	0.0986	0.0216	0.0031	0.0002	0.0000	0.0000
6	0.0000	0.0001	0.0056	0.0354	0.1340	0.2609	0.2792	0.1848	0.0767	0.0205	0.0025	0.0003
7	0.0000	0.0000	0.0001	0.0029	0.0202	0.0798	0.1911	0.2584	0.2496	0.1437	0.0463	0.0079
8	0.0000	0.0000	0.0000	0.0003	0.0019	0.0106	0.0505	0.1404	0.2431	0.2910	0.2017	0.0604
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0063	0.0381	0.1239	0.2583	0.3525	0.2205
10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008	0.0066	0.0366	0.1455	0.3630	0.4475
$p(s)$	0.1361	0.1099	0.0919	0.0743	0.0671	0.0627	0.0630	0.0650	0.0732	0.0859	0.0969	0.0742

$$E[\lambda_{jc}] = (v_\alpha c + 1) / [v_\alpha c + 2 + v_\beta(k + 1 - c)], \quad (8)$$

$$\text{Var}[\lambda_{jc}] = \alpha_c \beta_c / [(\alpha_c + \beta_c)^2 (\alpha_c + \beta_c + 1)]. \quad (9)$$

Hence, it is natural to calibrate these hyper-parameters so that the expected values of the success probabilities are reasonably centered and well separated, and these parameters have a reasonable variability a priori. Note that one might choose an expected value and variance for one fixed c (e.g., $c = 1$) and then obtain v_α and v_β by inverting functions (8) and (9).

The choice of the prior parameters obviously involves an arbitrary decision regarding $p(s)$. A possible solution to overcome this problem would be to assign hyper-priors to v_α and v_β , assuming for instance that they are Gamma distributed. This would allow extra flexibility to the model and cope with uncertainty about these parameters. In this way, in fact, the shrinkage parameters would be chosen by the data. In the present context, however, we consider the free choice of the shrinkage parameters as an advantage of the model rather than a problem to be solved. By varying the shrinkage parameters, in fact, we can assign a measure to the strength of our personal belief in the unidimensionality of the data generating model.

The prior distribution defined in (6) is particularly flexible, since with adequate choices it might both deliver an approximately uniform prior on s (Table 2), which is advisable for explorative purposes, or a prior which assigns a reasonable probability to $s = 1$ (Table 3), which can be adopted in testing the hypothesis of unidimensionality. The prior in (6) includes the uniform prior as a special case (for $v_\alpha = v_\beta = 0$). The results in Tables 2 and 3 have been obtained by running the RJ-MCMC algorithm, described in the following section, always sampling from the prior (i.e., without data). Our post-processing algorithm has then been used to evaluate the conditional and unconditional distribution of s .

As a final remark, notice that the prior in (6) is not invariant with respect to the relabeling of the classes, unless $v_\alpha = v_\beta = 0$. This determines, in turn, a joint posterior which is not invariant to relabeling and should prevent from label switching in the RJ-MCMC algorithm that simulates the posterior distribution of the model parameters.

In conclusion, it is worth recalling that Klugkist et al. (2005) deeply discussed the sensitivity of inferential results to the specification of the encompassing prior. In particular, they stated that a vague encompassing prior, conceived not to favor any particular parameter ordering, determines a virtually objective model selection. In our setting, such a prior would be a uniform prior between 0 and 1 for λ_{jc} , $j = 1, \dots, r$, $c = 1, \dots, k$, which assigns the same probability to any ordering of the λ_{jc} , conditionally on k (see ‘‘Appendix’’). However, the main purpose of this paper is

TABLE 3.

Simulated conditional and unconditional prior distribution of the number of partitions, as determined by a Beta prior with $v_\alpha = v_\beta = 14$ on the success probabilities, when $r = 12$ and $k_{\max} = 10$.

k	s											
	1	2	3	4	5	6	7	8	9	10	11	12
1	1.0000	–	–	–	–	–	–	–	–	–	–	–
2	0.9869	0.0131	–	–	–	–	–	–	–	–	–	–
3	0.9224	0.0759	0.0017	0.0000	0.0000	0.0000	–	–	–	–	–	–
4	0.7917	0.1957	0.0124	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.5937	0.3407	0.0616	0.0040	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.3960	0.4255	0.1528	0.0237	0.0019	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.2178	0.4070	0.2740	0.0886	0.0116	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.1033	0.3162	0.3446	0.1805	0.0477	0.0070	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0455	0.1929	0.3385	0.2640	0.1253	0.0297	0.0036	0.0005	0.0000	0.0000	0.0000	0.0000
10	0.0135	0.0901	0.2455	0.3122	0.2164	0.0958	0.0224	0.0038	0.0004	0.0000	0.0000	0.0000
$p(s)$	0.5071	0.2056	0.1433	0.0874	0.0403	0.0133	0.0027	0.0004	0.0000	0.0000	0.0000	0.0000

to make inference on the number of dimensions s , rather than on a particular ordering of the parameters such as in Klugkist et al. (2005). Thus, as already discussed, a uniform prior on the success probabilities is not advisable, since it would assign a very small prior probability to models of interest (for example the unidimensional model). Obviously, also the prior specified in (6) is expected to affect inferential results, since the choice of v_α and v_β may determine very different values of $p(s)$. For this reason, a sensitivity analysis to evaluate the effect on the results of different choices of v_α and v_β is highly recommended. Notice, however, that due to the type of inequality constraints considered in (3), model selection should not be as strongly affected by the encompassing prior, as it would be in the case of approximate equality constraints. For this kind of constraints, the evidence in favor of the constrained model is known to increase with the amount of vagueness of the encompassing prior, a phenomenon known as Bartlett's or Lindley's paradox (Lindley, 1957).

3.3. Reversible Jump Implementation

The complexity of the model presented above makes the joint posterior of all parameters, including k , analytically intractable, due to the high-dimensional integrals needed to evaluate its normalizing constant. Therefore, we resort to an RJ-MCMC method to approximate this posterior distribution (Green, 1995; Green & Richardson, 2001). In general, MCMC methods allow us to sample from a probability distribution, known up to its normalizing constant. They are based on the construction of an irreducible and aperiodic Markov chain that has the desired distribution as its equilibrium distribution. Details of these computational methods can be found in Tierney (1994). In particular, the RJ-MCMC is an extension of a standard MCMC algorithm that may be used to draw values from a probability distribution defined on parameter spaces of varying dimension. Thus, in our framework, it permits to simulate the posterior distribution even when the number of latent classes, k , is unknown.

The RJ-MCMC sampler performs T sweeps and, at each sweep t , all the parameters, including k , are updated in turn. This is performed by drawing the new values of certain parameters conditionally on the data and all the other parameters. The algorithm uses different fixed-dimension moves to update the model parameters, conditionally on a fixed k , further to two different variable dimension moves to update the number of latent classes k . In particular, updating the value of k implies a change of dimensionality for the success probabilities, the class weights, and the allocation variables. We accomplish this by introducing two different types of RJ move. The first one

consists of a random choice between splitting an existing class in two and merging two existing classes into one. The second one consists of a random choice between the birth of an empty class or the death of an empty class. The probabilities of the split/merge alternatives are b_k and $d_k = 1 - b_k$, respectively, and depend only on the current value of k . Of course, $d_1 = 0$ and $b_{k_{\max}} = 0$; otherwise we choose $b_k = d_k = 0.5$, for $k = 2, \dots, k_{\max} - 1$. The same probabilities b_k and d_k are used for the birth/death alternatives, while a probability of 0.5 is used for choosing between a split/combine or a death/birth move.

A schematic description of the algorithm is provided below, where $q(\cdot|\cdot)$ is used to denote an appropriate proposal distribution used in the Metropolis-Hastings steps for a certain parameter vector, while $p(\cdot|\dots)$ is the full conditional distribution in the Gibbs sampling steps, with “ \dots ” standing for “all other parameters and data”; moreover, t is the sweep counter and T is the overall number of sweeps. A more detailed illustration of the algorithm, in which all quantities used below are precisely defined, is given in “Appendix.”

```

1: while  $t < T$  do
2:   set  $t = t + 1$ ;
3:   draw  $\boldsymbol{\pi}^{(t)} \sim p(\boldsymbol{\pi}|\dots)$ ; ▷ update  $\boldsymbol{\pi}$ 
4:   draw  $\mathbf{z}_i^{(t)} \sim p(\mathbf{z}_i|\dots)$ , independently  $\forall i$ ; ▷ update  $\mathbf{Z}$ 
5:   draw  $\lambda_{jc}^* \sim q(\lambda_{jc}|\lambda_{jc}^{(t-1)})$ , independently  $\forall j, c$ ; ▷ update  $\boldsymbol{\Lambda}$ 
6:   calculate Metropolis-Hastings' acceptance probability  $p_{\lambda_{jc}^*}$  and draw  $u \sim \text{Unif}[0; 1]$ ;
7:   if  $u \leq p_{\lambda_{jc}^*}$  then  $\lambda_{jc}^{(t)} = \lambda_{jc}^*$  else  $\lambda_{jc}^{(t)} = \lambda_{jc}^{(t-1)}$  end if
8:   draw  $(u_1, u_2) \sim \text{Unif}[0; 1]$ ; ▷ start updating  $k$ 
9:   if  $u_1 \leq 0.5$  and  $u_2 \leq b_{k^{(t-1)}}$  then ▷ try split
10:    draw a class  $c' \sim \text{Unid}[0; k^{(t-1)}]$  to be split into  $c_1$  and  $c_2$ ;
11:    split parameters of class  $c'$  to create  $\boldsymbol{\pi}^*$  and  $\boldsymbol{\Lambda}^*$ , and calculate  $\alpha_c^*$  and  $\beta_c^*$ ,  $\forall c$ ;
12:    reallocate observations for which  $z_{ic'} = 1$  to  $c_1$  or  $c_2$ , to create  $\mathbf{Z}^*$ ;
13:    calculate RJ acceptance probability  $p_k$  and draw  $u \sim \text{Unif}[0; 1]$ ;
14:    if  $u \leq p_k$  then
15:      set  $k^{(t)} = k^{(t-1)} + 1$ ,  $\boldsymbol{\pi}^{(t)} = \boldsymbol{\pi}^*$ ,  $\boldsymbol{\Lambda}^{(t)} = \boldsymbol{\Lambda}^*$ ,  $\mathbf{Z}^{(t)} = \mathbf{Z}^*$ ,  $\boldsymbol{\alpha}^{(t)} = \boldsymbol{\alpha}^*$ ,  $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^*$ ;
16:    else
17:      set  $k^{(t)} = k^{(t-1)}$ ,  $\boldsymbol{\pi}^{(t)} = \boldsymbol{\pi}^{(t-1)}$ ,  $\boldsymbol{\Lambda}^{(t)} = \boldsymbol{\Lambda}^{(t-1)}$ ,  $\mathbf{Z}^{(t)} = \mathbf{Z}^{(t-1)}$ ,  $\boldsymbol{\alpha}^{(t)} = \boldsymbol{\alpha}^{(t-1)}$ ,  $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{(t-1)}$ ;
18:    end if
19:  else if  $u_1 \leq 0.5$  and  $u_2 > b_{k^{(t-1)}}$  then ▷ try combine
20:    draw two classes  $(c_1, c_2) \sim \text{Unid}[0; k^{(t-1)}]$  to be merged into  $c'$ ;
21:    merge parameters of classes  $(c_1, c_2)$  to create  $\boldsymbol{\pi}^*$  and  $\boldsymbol{\Lambda}^*$ , and calculate  $\alpha_c^*$  and  $\beta_c^*$ ,  $\forall c$ ;
22:    reallocate observations for which  $z_{ic_1} = 1$  or  $z_{ic_2} = 1$  to  $c'$  and create  $\mathbf{Z}^*$ ;
23:    repeat command lines 13 to 18
24:  else if  $u_1 > 0.5$  and  $u_2 \leq b_{k^{(t-1)}}$  then ▷ try birth
25:    draw a position from  $\text{Unid}[0; k^{(t-1)}]$  to be occupied by new class  $c'$ ;
26:    draw parameters for  $c'$  from priors to get  $\boldsymbol{\pi}^*$  and  $\boldsymbol{\Lambda}^*$  and calculate  $\alpha_c^*$  and  $\beta_c^*$ ,  $\forall c$ ;
27:    set  $z_{ic'} = 0$ ,  $i = 1, \dots, n$ , to get  $\mathbf{Z}^*$ ;
28:    repeat command lines 13 to 18
29:  else if  $u_1 > 0.5$  and  $u_2 > b_{k^{(t-1)}}$  then ▷ try death
30:    draw a class  $c'$  at random among empty classes;
31:    delete parameters for  $c'$  to get  $\boldsymbol{\pi}^*$  and  $\boldsymbol{\Lambda}^*$  and calculate  $\alpha_c^*$  and  $\beta_c^*$ ,  $\forall c$ ;
32:    delete the  $c'$ -th column of  $\mathbf{Z}$  to get  $\mathbf{Z}^*$ ;
33:    repeat command lines 13 to 18
34:  end if
35: end while

```

The algorithm presented above is quite standard. See, for example, Pan and Huang (2014) for an implementation of the RJ-MCMC algorithm to an LC model for polytomous response variables. On the other hand, the post-processing algorithm, applied to the RJ-MCMC output to assess dimensionality and illustrated in the next section, is completely innovative.

3.4. Post-processing for Assessing Model Dimension

In practice, the RJ-MCMC sampling algorithm outlined in the previous section is used to approximate the joint posterior distribution of the model parameters. It is important to underline that the algorithm also updates the success probabilities without imposing any constraint on them; therefore, the simulated joint posterior distribution we obtain is that of the unconstrained (encompassing) LC(k) model, for $k = 1, \dots, k_{\max}$. In order to obtain the posterior probability of any constrained model LC(k, \mathcal{P}) and to assess dimensionality, we then need to post-process the simulated posterior distribution. In practice, for each sweep of the RJ-MCMC algorithm, we need to verify the particular partition \mathcal{P} with which the success probabilities λ_{jc} are in accordance, following the definition given in the previous section and based on assumptions A1 and A2. This also indicates the dimension of the model visited at each sweep. The proposed post-processing method is described below.

Let $k^{(t)}$ be the number of classes of the model visited at sweep t of the RJ-MCMC algorithm and let $\mathbf{\Lambda}^{(t)}$ and $\boldsymbol{\pi}^{(t)}$ be the simulated parameters of this model, with $t = 1, \dots, T$. Then, we examine every matrix $\mathbf{\Lambda}^{(t)}$ and we obtain the corresponding partition $\mathcal{P}^{(t)}$ made of the subsets of items $\mathcal{J}_1^{(t)}, \dots, \mathcal{J}_{s^{(t)}}^{(t)}$ such that A1 and A2 hold; thus, the model measures $s^{(t)} = |\mathcal{P}^{(t)}|$ different dimensions. To avoid a sort of label-switching problem, the groups are ordered so that $\mathcal{J}_1^{(t)}$ includes the first item, $\mathcal{J}_2^{(t)}$ includes the item with the smallest index among those excluded from $\mathcal{J}_1^{(t)}$, and so on.

More in detail, at each sweep t , let $\mathbf{P}^{(t)}$ be a matrix of zeros, having a number of rows equal to $\min(k^{(t)}, r)$, which corresponds to the maximum possible model dimensions given k and r , and a number of columns equal to $k^{(t)}$. Let $\mathbf{p}_d^{(t)}$ indicate the d -th row vector of $\mathbf{P}^{(t)}$. Initialize the post-processing algorithm by considering the first item and finding the permutation $\mathbf{c}^{(1)} = (c_1^{(1)}, \dots, c_k^{(1)})$ such that $\lambda_{1c_1}^{(1)} < \dots < \lambda_{1c_k}^{(1)}$. Set $\mathbf{p}_1^{(1)} = \mathbf{c}^{(1)}$, $s^{(1)} = 1$, $j = 1$ and allocate the first item to $\mathcal{J}_1^{(1)}$. Then the proposed post-processing algorithm proceeds as follows:

- 1: **while** $j < r$ **do**
- 2: set $j = j + 1$;
- 3: determine $\mathbf{c}^{(j)} = (c_1^{(j)}, \dots, c_k^{(j)})$ such that $\lambda_{jc_1}^{(j)} < \dots < \lambda_{jc_k}^{(j)}$;
- 4: **if** $\exists d$ such that $\mathbf{c}^{(j)} = \mathbf{p}_d^{(t)}$ **then**
- 5: allocate item j to $\mathcal{J}_d^{(t)}$;
- 6: **else**
- 7: set $s^{(t)} = s^{(t)} + 1$;
- 8: set $\mathbf{p}_{s^{(t)}}^{(t)} = \mathbf{c}^{(j)}$;
- 9: allocate item j to $\mathcal{J}_{s^{(t)}}^{(t)}$;
- 10: **end if**
- 11: **end while**

Finally, the posterior probability of model LC(k, \mathcal{P}) with a certain number of classes k and a certain partition of items $\mathcal{P} = \{\mathcal{J}_1, \dots, \mathcal{J}_s\}$ is estimated as:

$$\begin{aligned}
p(k, \mathcal{P}|\mathbf{Y}) &= \frac{1}{T} \sum_{t=1}^T I \{k^{(t)} = k, \mathcal{P}^{(t)} = \mathcal{P}\} \\
&= \frac{1}{T} \sum_{t:s^{(t)}=s} I \{k^{(t)} = k, \mathcal{J}_1^{(t)} = \mathcal{J}_1, \dots, \mathcal{J}_{s^{(t)}}^{(t)} = \mathcal{J}_s\},
\end{aligned}$$

where the sum is over all sweeps for which $s^{(t)} = s$.

3.5. Testing Unidimensionality

The above RJ-MCMC output post-processing can also be used to obtain posterior probabilities or Bayes factors (BFs; Kass & Raftery, 1995), which may be used to test the unidimensionality hypothesis. The conditional posterior probability of the number of dimensions can be readily obtained as

$$p(s|k, \mathbf{Y}) = \frac{\sum_{t=1}^T I \{k^{(t)} = k, s^{(t)} = s\}}{\sum_{t=1}^T I \{k^{(t)} = k\}}. \quad (10)$$

Otherwise, using model averaging principles, we can estimate the marginal posterior distribution of s as

$$p(s|\mathbf{Y}) = \frac{1}{T} \sum_{t=1}^T I \{s^{(t)} = s\}. \quad (11)$$

From (10) and (11) we obtain, respectively, $p(s = 1|k, \mathbf{Y})$ and $p(s = 1|\mathbf{Y})$, that is, the conditional and marginal posterior probability of the unidimensional model. In order to get the corresponding BFs, we need $p(s|k)$ and $p(s)$, which can both be obtained by simulation, as already explained. Given that we use the encompassing prior approach, the BFs are in fact obtained as the ratio of posterior and prior probabilities of the restricted model.

3.6. Other Elaborations of the Simulated Posterior Distribution

The proposed approach, and in particular the simulated posterior distribution obtained from the algorithm described above, can also be used for certain elaborations that are typical of IRT applications.

First of all, for a given k and s , it is possible to characterize the latent classes in terms of probability of success and weight. For this aim, we rely on the Bayesian estimates of the parameters λ_{jc} , that is,

$$\hat{\lambda}_{jc}(k, s) = \frac{\sum_{t=1}^T \lambda_{jc}^{(t)} I \{k^{(t)} = k, s^{(t)} = s\}}{\sum_{t=1}^T I \{k^{(t)} = k, s^{(t)} = s\}}, \quad j = 1, \dots, r, \quad c = 1, \dots, k, \quad (12)$$

and the parameters π_c , that is,

$$\hat{\pi}_c(k, s) = \frac{\sum_{t=1}^T \pi_c^{(t)} I \{k^{(t)} = k, s^{(t)} = s\}}{\sum_{t=1}^T I \{k^{(t)} = k, s^{(t)} = s\}}, \quad c = 1, \dots, k. \quad (13)$$

This characterization is particularly useful, for instance, in educational assessment, where the probability of success is a measure of the ability of individuals in a certain class. On the other

hand, the weight corresponds to the proportion of individuals in the class, and then of examinees with a certain ability level. For this aim, we can also use the unconditional (with respect to the model dimension) posterior estimates, so as to evaluate the impact of imposing a certain number of dimensions on the latent class identification. This amounts to use the following expressions

$$\hat{\lambda}_{jc}(k) = \frac{\sum_{t=1}^T \lambda_{jc}^{(t)} I\{k^{(t)} = k\}}{\sum_{t=1}^T I\{k^{(t)} = k\}},$$

$$\hat{\pi}_c(k) = \frac{\sum_{t=1}^T \pi_c^{(t)} I\{k^{(t)} = k\}}{\sum_{t=1}^T I\{k^{(t)} = k\}},$$

instead of (12) and (13). Alternatively, it is possible to compute these estimates given a certain partition of items \mathcal{P} and not just conditionally on the number of dimensions.

Another relevant elaboration leads to clustering individuals on the basis of the response configuration they provide, so as to produce an estimate of their ability level. More precisely, for a certain value of k and s , it is possible to estimate the conditional probability that individual i belongs to latent class c given the response configuration \mathbf{y}_i he/she provided, with $i = 1, \dots, n$ and $c = 1, \dots, k$. This estimate is obtained as

$$\hat{\pi}_c(k, s, \mathbf{y}_i) = \frac{\sum_{t=1}^T I\{z_{ic}^{(t)} = 1, k^{(t)} = k, s^{(t)} = s\}}{\sum_{t=1}^T I\{k^{(t)} = k, s^{(t)} = s\}}, \quad (14)$$

where $z_{ic}^{(t)}$ is the value of the indicator variable z_{ic} at the t -th sweep of the RJ-MCMC algorithm. In practice $z_{ic}^{(t)} = 1$ if, for this sweep, individual i has been assigned to latent class c and $z_{ic}^{(t)} = 0$ otherwise. Then, this individual is assigned to the class corresponding the highest value of $\hat{\pi}_c(k, s, \mathbf{y}_i)$ or, equivalently, to the class with the largest number of visits. In a simple way, we can also use different conditioning arguments and, for instance, obtain the estimate of the conditional probability of belonging to class c given \mathbf{y}_i and k only, namely $\hat{\pi}_c(k, \mathbf{y}_i)$. Moreover, if more individuals have the same response configuration, we can average $\hat{\pi}_c(k, s, \mathbf{y}_i)$ or $\hat{\pi}_c(k, \mathbf{y}_i)$ over the subsample of these individuals to obtain more stable results.

Finally, if the dataset includes individual covariates, corresponding for instance to gender or describing the family background, it may be of interest to relate the probabilities of belonging to a certain class to these covariates to understand, for instance, the effect of these covariates on the ability in an educational application. This may be based on computing the previous estimates for separate covariate configurations and then making suitable comparisons. However, this comparison is outside the scope of the present paper, whose main focus is on the assessment of dimensionality of a certain questionnaire from an IRT perspective.

4. Applications

To illustrate the proposed approach we consider four examples. The first two are based on different datasets simulated from a unidimensional model and from a multidimensional model. In this way, we assess the capability of the proposed algorithm in recovering these two situations and the correct partition of items in the multidimensional case. The third application is based on educational data for assessment in Mathematics, while the fourth considers hospital anxiety and depression data.

TABLE 4.
Estimated posterior probabilities of the number of latent classes.

k	1	2	3	4	5	6	7	8	9	10
$p(k \mathbf{Y})$	0.0000	0.0000	0.0074	0.4648	0.4203	0.0946	0.0120	0.0009	0.0000	0.0000

Simulated unidimensional data

TABLE 5.
Prior and posterior probabilities of s , and BFs.

s	1	2	3	4	5	6	7	8	9	10
$p(s k=4)$	0.3734	0.4704	0.1446	0.0115	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
$p(s k=4, \mathbf{Y})$	0.6382	0.3154	0.0447	0.0017	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BF	1.7093	0.6705	0.3089	0.1484	0.0000	-	-	-	-	-

Simulated unidimensional data.

4.1. Simulated Data: Unidimensional Case

For the first example, we set $n = 500$, $r = 10$, $k = 4$, $v_\alpha = 4$, and $v_\beta = 4$, and we generated an $r \times k$ matrix $\mathbf{\Lambda}$ according to (6), ordering all conditional probabilities λ_{jc} in the same way, so that unidimensionality is attained. Then, we drew $\boldsymbol{\pi}$ from its prior and allocated each subject i to latent class c with probability π_c , $c = 1, \dots, k$; this amounts to suitably define the indicator variables z_{ic} , $i = 1, \dots, n$, $c = 1, \dots, k$. Finally, we simulated the data \mathbf{Y} by letting, for $i = 1, \dots, n$ and $j = 1, \dots, r$, $y_{ij} = 1$ with probability λ_{jc} , where c is the class to which subject i is allocated ($z_{ic} = 1$).

Based on the simulated data, we ran 150,000 sweeps of the proposed RJ-MCMC algorithm after a burn-in of 50,000 sweeps and considered models with a number of latent classes up to $k_{\max} = 10$. Table 4 shows the simulated posterior distribution $p(k|\mathbf{Y})$. Considering that the prior for k is a discrete uniform distribution between 1 and k_{\max} , the data provide evidence in favor of the model with $k = 4$ latent classes.

Regarding dimensionality, Table 5 provides the (conditional) prior and posterior probabilities of s , as well as the BFs. We note that the unidimensional model is visited around 64% of times and is the one with the highest posterior probability, confirming that, at least in this case, the proposed strategy works properly.

In order to check the convergence of the posterior distributions $p(k|\mathbf{Y})$ and $p(s|k, \mathbf{Y})$ to the real values of k and s , as the sample size increases, we considered a small Monte Carlo study. Under the same settings and the same parameters $\mathbf{\Lambda}$ and $\boldsymbol{\pi}$ as above, we considered increasing sample sizes: $n = 500, 750, 1000$. For each sample size, we generated ten different datasets. Each dataset was obtained by drawing the allocation variables z_{ic} , $i = 1, \dots, n$, $c = 1, \dots, k$, and the data \mathbf{Y} as described above. For each dataset we then ran the RJ-MCMC algorithm and for every parameter we calculated the root-mean-squared error (RMSE); the RMSE for s , $\boldsymbol{\pi}$, and $\mathbf{\Lambda}$ were calculated conditionally on $k = 4$. Table 6 shows these RMSEs, averaged over the 10 replications, for increasing values of n . As expected, for all parameters the RMSEs tend to decrease as n increases. In particular, the posterior distribution of the number of dimensions becomes more and more concentrated on its true value $s = 1$, corresponding to the unidimensionality hypothesis. Also notice that, out of the 10 replications and for $n = 500, 750, 1000$, the number of latent classes was correctly recovered 8, 10, and 10 times, respectively, and the number of dimensions was correctly recovered 7, 9, and 9 times, respectively.

TABLE 6.
RMSE for each parameter, averaged over 10 replications, for increasing values of n .

n	RMSE			
	k	$s k$	πk	Λk
500	0.6440	0.7076	0.1490	0.8431
750	0.4038	0.6842	0.1198	0.7710
1000	0.3173	0.6327	0.1187	0.6851

Simulated unidimensional data.

TABLE 7.
Estimated posterior probabilities of the number of latent classes.

k	1	2	3	4	5	6	7	8	9	10
$p(k Y)$	0.0000	0.0000	0.0001	0.7007	0.2019	0.0733	0.0230	0.0011	0.0000	0.0000

Simulated multidimensional data.

TABLE 8.
Prior and posterior probabilities of s , and BFs.

s	1	2	3	4	5	6	7	8	9	10
$p(s k=4)$	0.0021	0.0582	0.2916	0.4107	0.1931	0.0402	0.0039	0.0001	0.0001	0.0000
$p(s k=4, Y)$	0.0006	0.0376	0.2827	0.4622	0.1919	0.0239	0.0011	0.0000	0.0000	0.0000
BF	0.2970	0.6452	0.9695	1.1254	0.9937	0.5943	0.2938	0.1450	0.0000	–

Simulated multidimensional data.

4.2. Simulated Data: Multidimensional Case

For the multidimensional case, we set $r = 10$, $k = 4$, $v_\alpha = 1.5$, and $v_\beta = 1.5$, and we generated a unique $r \times k$ matrix Λ according to (6). Post-processing of this matrix revealed the following partition of the items into $s = 4$ groups: $\{(1, 6), (2, 8), (4, 5), (3, 7, 9, 10)\}$. Then we drew class weights, allocation variables, and data as in the previous example. Finally, for the simulated data we ran 150 000 sweeps of the proposed algorithm after a burn-in of 50 000 sweeps, and let $k_{\max} = 10$.

Table 7 shows the posterior distribution $p(k|Y)$, from which the model with $k = 4$ latent classes seems to be favored. Table 8 provides the prior and posterior probabilities of s , and the BFs. The unidimensional model is rarely visited, and the model with the highest posterior probability is the one with $s = 4$ groups, corresponding to the true number of dimensions. The partitions into four groups receiving the highest BF (1.74) were those having three groups of two items each and a group of four items. Among these a priori equally probable partitions, the most visited one (15% of times) is the correct one, that is, $\{(1, 6), (2, 8), (4, 5), (3, 7, 9, 10)\}$. The second most visited partition is $\{(1, 6), (7, 8), (4, 5), (2, 3, 9, 10)\}$, which only differs from the correct one for the switching between items 2 and 7, and has a simulated posterior probability (conditional on k , s , and type of partition) of 0.11.

Also for the multidimensional case, we checked the convergence of the posterior distributions to the real values of the parameters, as the sample size increases. Under the same settings and the same Λ and π as above, we considered increasing sample sizes and for each of them, we generated ten different datasets. Table 9 shows the RMSE for each parameter, averaged over the 10 replications, for increasing values of n . As for the unidimensional case, the RMSEs tend to decrease as n increases and the posterior distribution of the number of dimensions becomes more

TABLE 9.
RMSE for each parameter, averaged over 10 replications, for increasing values of n .

n	RMSE			
	k	$s k$	πk	Λk
500	0.5147	0.6770	0.1311	1.4770
750	0.4160	0.5300	0.1231	1.3298
1000	0.3202	0.4881	0.0911	1.1565

Simulated multidimensional data.

and more concentrated on its true value $s = 4$. Out of the 10 replications for $n = 500, 750, 1000$, respectively, the number of latent classes was correctly recovered 9, 10, and 10 times, the number of dimensions was correctly recovered 8, 8, and 9 times, while the correct partition was correctly recovered 5, 5, and 6 times.

4.3. An Application in Educational Assessment

The proposed approach is applied to the analysis of a dataset concerning a sample of $n = 1510$ examinees who responded to a set of $r = 12$ items on Mathematics, the same dataset analyzed in Bartolucci (2007).

The 12 items concern the following subjects:

1. Round to thousand place;
2. Write fraction that represents shaded region;
3. Multiply two negative integers;
4. Reason about sample space (number correct);
5. Find amount of restaurant tip;
6. Identify representative sample;
7. Read dials on a meter;
8. Find $(x; y)$ solution of linear equation;
9. Translate words to symbols;
10. Find number of diagonals in polygon from a vertex;
11. Find perimeter (quadrilateral);
12. Reason about betweenness.

This dataset, available in the R package `MultiLCIRT` (Bartolucci et al., 2016), is part of a larger dataset collected in 1996 by the Educational Testing Service within the NAEP project; see Bartolucci and Forcina (2005) for a more detailed description.

We considered models with a number of latent classes up to $k_{\max} = 10$ under assumption (6), with α_c and β_c , $c = 1, \dots, k$, fixed according to the rule in (7). In this regard we recall that it is necessary to fix v_α and v_β and we calibrated them for a specific, most likely k . In particular, we considered $k = 4$ latent classes as found in Bartolucci (2007) and we looked for prior parameters such that the four latent classes are reasonably separated, and variances are relatively small, with v_α and v_β taken from a grid of possible values. Our final choice is $v_\alpha = 2.5$ and $v_\beta = 1$ and leads to an expected value of the vector of the λ_{jc} parameters equal to $(0.2000, 0.5385, 0.7500, 0.8947)$, computed according to (8), and corresponding variances equal to $(0.0267, 0.0331, 0.0208, 0.0090)$, computed according to (9).

We then proceeded, as for the previous examples, with 150 000 sweeps of the RJ-MCMC algorithm after a burn-in of 50 000 sweeps. Overall, $s = 1$ occurred around 42% of times. In this

TABLE 10.
Estimated posterior probabilities of the number of latent classes for the NAEP data.

k	1	2	3	4	5	6	7	8	9	10
$p(k Y)$	0.0000	0.0000	0.2819	0.7006	0.0175	0.0000	0.0000	0.0000	0.0000	0.0000

TABLE 11.
Estimated prior and posterior probabilities of the number of dimensions and corresponding BFs for the NAEP data.

s	1	2	3	4	5	6	7	8	9	10
$p(s k = 4)$	0.0011	0.0392	0.2379	0.4092	0.2369	0.0669	0.0080	0.0008	0.0000	0.0000
$p(s k = 4, Y)$	0.1955	0.7511	0.0530	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BF	177.7127	19.1604	0.2228	0.0010	0.0000	0.0000	0.0000	0.0000	–	–

regard note that, using the encompassing approach, we have a prior probability of about 0.14 and then the resulting BF is 3.0. We, therefore, actually have evidence in favor of unidimensionality.

Table 10 shows the posterior distribution $p(k|Y)$. Considering the uniform prior on k , the model with $k = 4$ latent classes seems to be favored. Table 11 shows the estimated prior and posterior probabilities for s , as well as BFs conditionally on $k = 4$. According to the encompassing prior approach, also conditioning on k , the case $s = 1$ is receiving the largest BF, corresponding to 177.7.

We also performed a sensitivity analysis to evaluate the effect on the results of different choices of v_α and v_β . We found that, for the data at hand, the value of the BF is only slightly affected by reasonable changes in the value of these hyper-parameters, so that the evidence in favor of unidimensionality is left unaltered. We, therefore, conclude that for these data there is founded evidence in favor of unidimensionality, and that, likely, there are four groups of students of increasing mathematical ability. In this regard, it is important to recall that for these data Bartolucci (2007) concluded in favor of a model measuring $s = 3$ dimensions corresponding to the following partition: $\{(1, 2, 9, 10), (3, 5, 8, 11), (4, 6, 7, 12)\}$. This difference in the main conclusion (unidimensionality vs. multidimensionality) is likely due to the less stringent assumptions of the present approach with respect to that in Bartolucci (2007), which is based on a 2PL model, as discussed at the beginning of the present paper.

Finally, as an illustration of the practical use of the model for student' assessment, we report in Table 12 the Bayesian estimates of the λ_{jc} parameters conditional on the selected model, with $k = 4$ and $s = 1$, and only conditional on $k = 4$, together with the proportion of exact responses for each item, denoted by $\bar{y}_{.j}$, for $j = 1, \dots, r$. The corresponding estimates of the π_c parameters are also given in Table 12. These estimates are obtained on the basis of expressions (12) and (13), when conditional on k and s , and the similar expressions reported in the same section when conditional on k only.

It is evident from the results in Table 12 that the latent classes are increasingly ordered in terms of probability of success and thus of ability. This is exactly true for the unidimensional model ($s = 1$) and true, with only one exception for the second item, when there is no restriction on the model dimensionality. In any case, there is a strong agreement between the estimates of the λ_{jc} parameters under the two scenarios (conditional on $s = 1$ and unconditional on s) confirming that the data provide evidence in favor of unidimensionality. Obviously, there is also agreement between these estimates and the proportions $\bar{y}_{.j}$, confirming the validity of the proposed Bayesian estimation algorithm.

TABLE 12.
Estimates of the λ_{jc} and π_c parameters, together with proportion of correct response, for NAEP data.

j	$\hat{\lambda}_{j1}(4, 1)$	$\hat{\lambda}_{j2}(4, 1)$	$\hat{\lambda}_{j3}(4, 1)$	$\hat{\lambda}_{j4}(4, 1)$	$\hat{\lambda}_{j1}(4)$	$\hat{\lambda}_{j2}(4)$	$\hat{\lambda}_{j3}(4)$	$\hat{\lambda}_{j4}(4)$	$\bar{y}_{\cdot j}$
1	0.2902	0.7049	0.8689	0.9629	0.2975	0.7120	0.8752	0.9570	0.7358
2	0.2341	0.6898	0.9170	0.9568	0.2366	0.7029	0.9252	0.9148	0.7430
3	0.2545	0.4854	0.8056	0.9340	0.2549	0.4986	0.8166	0.9253	0.6166
4	0.4512	0.8879	0.9540	0.9848	0.4624	0.8904	0.9578	0.9720	0.8675
5	0.1953	0.3527	0.5850	0.8466	0.1960	0.3592	0.5988	0.8449	0.4543
6	0.3698	0.6907	0.8536	0.9515	0.3755	0.6954	0.8625	0.9375	0.7338
7	0.2892	0.5178	0.7779	0.9367	0.2923	0.5257	0.7894	0.9371	0.6219
8	0.0924	0.2961	0.8068	0.9153	0.0914	0.3149	0.8224	0.8918	0.5199
9	0.1682	0.5561	0.9108	0.9702	0.1731	0.5705	0.9216	0.9525	0.6808
10	0.1815	0.5667	0.8262	0.9337	0.1861	0.5768	0.8362	0.9225	0.6470
11	0.0659	0.1385	0.4249	0.8842	0.0665	0.1453	0.4427	0.8937	0.2848
12	0.0477	0.1972	0.3727	0.8078	0.0480	0.2030	0.3835	0.8430	0.2768
$\hat{\pi}_c$	0.1344	0.3820	0.4336	0.0500	0.1393	0.3994	0.4208	0.0405	

TABLE 13.
Estimated posterior probabilities of $\pi_c(y_i)$ parameters for NAEP data.

y_i	\bar{y}_i	#	$\hat{\pi}_1(y_i, 4, 1)$	$\hat{\pi}_2(y_i, 4, 1)$	$\hat{\pi}_3(y_i, 4, 1)$	$\hat{\pi}_4(y_i, 4, 1)$	$\hat{\pi}_1(y_i, 4)$	$\hat{\pi}_2(y_i, 4)$	$\hat{\pi}_3(y_i, 4)$	$\hat{\pi}_4(y_i, 4)$
000000000000	0.00	10	0.9966	0.0034	0.0000	0.0000	0.9971	0.0029	0.0000	0.0000
110100000000	0.25	2	0.4378	0.5611	0.0010	0.0000	0.4760	0.5234	0.0006	0.0000
111111000000	0.50	1	0.0348	0.9269	0.0383	0.0000	0.0385	0.9346	0.0270	0.0000
111111110000	0.75	8	0.0000	0.1012	0.8852	0.0136	0.0001	0.1231	0.8677	0.0092
111111111111	1.00	41	0.0000	0.0011	0.3825	0.6163	0.0000	0.0017	0.4964	0.5018

Regarding the class description, we also observe on the basis of the results in Table 12 that the classes of intermediate ability, which are the second and the third, have the largest size, whereas the class of the examinees with the highest ability level is the smallest. Note that, also for the Bayesian estimates of parameters π_c , conditioning or not on s does not make any relevant difference.

The proposed model can also be used to cluster individuals on the basis of the posterior probability of belonging to the latent classes. To illustrate this process, in Table 13 we consider five respondents' configurations y_i , increasingly ordered according to the proportion of correct responses, \bar{y}_i . For each of these configurations, the table shows the number of individuals in the dataset and the posterior estimates $\hat{\pi}_c(y_i, k, s)$, computed according to (14), and $\hat{\pi}_c(y_i, k)$.

We observe that the individuals in Table 13 with all wrong responses must be clearly assigned to the first class and, in general, the probabilities $\hat{\pi}_c(y_i, k, s)$ and $\hat{\pi}_c(y_i, k)$ increase for the last latent class, corresponding to the highest ability level, as the number of correct responses increases. This is in agreement with the characterization of the latent classes provided above. Also note that such estimated probabilities reflect the level of uncertainty involved in the clustering process and then, for instance, there is uncertainty in assigning individuals with all correct responses to the third or the fourth latent class.

As already noted, for the NAEP data our approach provides evidence of unidimensionality and this result is in contrast with that provided by alternative approaches and, in particular, with that of Bartolucci (2007) that finds evidence of multidimensionality. This is a crucial difference that, as already conjectured in the previous sections, we can expect considering that our approach is nonparametric, while alternative approaches formulate the conditional distribution of the response variables given the latent variables in a parametric way. In order to clarify and better justify this

conjecture, we performed the following experiment. We considered the estimated conditional response probabilities $\hat{\lambda}_{jc}(4, 1)$ and class weights $\hat{\pi}_c(4, 1)$ given in Table 12. We used the class weights to draw the indicator variables z_{ic} , for $i = 1, \dots, n$ and $c = 1, \dots, k$, with $n = 1510$. Then, we simulated the data Y as explained before. In this way, we obtained a dataset comparable to the one for which our nonparametric approach and the semiparametric approach in Bartolucci (2007) produced contrasting results. However, for this simulated dataset unidimensionality holds. We then fitted our model and the 2PL model to the simulated data, to test dimensionality. While our approach determined a BF for the unidimensional model equal to 187.3 (the BF for $s = 2$ being equal to 18.8), the 2PL model concluded in favor of two dimensions, with a strong evidence against unidimensionality (p -value equal to 0.002).

For comparison we also report some results based on alternative IRT models. We considered both classical Rasch models and a full information item factor analysis (as in Bock & Muraki, 1988; see also Reckase, 2009), which is different from that of Bartolucci (2007) and related approach because each item response is allowed to depend on more latent variables or factors (within-item multidimensionality), under the assumption that these variables have a normal distribution; see also Bacci and Bartolucci (2016) for further comments. In particular we considered the multi-factor exploratory IRT models, as implemented in the R package `mirt` (Chalmers et al., 2017). We initially fitted a 2PL specification of this model with 1, 2, and 3 factors, obtaining values of Akaike Information Criterion (AIC; Akaike, 1973) equal to 20273.59, 20262.61, and 20267.35, respectively. The corresponding values of the Bayesian Information Criterion (BIC; Schwarz, 1978) are 20401.27, 20448.81, and 20506.67, respectively. Therefore, according to BIC there is evidence of unidimensionality as according to our approach, but according to AIC there is evidence of bidimensionality; the latter conclusion is also reached according to the corrected Akaike Information Criterion (cAIC; Hurvich & Tsai, 1989), which leads to selecting the two-dimensional factor model. It is worth noting that under the three-parameter logistic (3PL) model, AIC, cAIC, and BIC lead to the same conclusions: the data generating model is unidimensional according to AIC and cAIC, whereas it is bidimensional according to BIC. Therefore, the conclusion does not depend on the specific parametrization of the conditional distribution of the responses given the latent variables but more crucially on the model selection criteria that is adopted. This point must be considered also regarding the comparison between the results of the proposed approach based on the BF and that of Bartolucci (2007), which is based on likelihood ratio testing.

To conclude this comparison, in Table 14 we report the results of a classical Rasch model, a 2-PL and 3-PL model (where guessing parameters for the 3-PL model are reported in Table 15) and the factor loadings corresponding to a two-dimensional item factor analysis. We do not report a two-dimensional item factor analysis after rotation, as the final factor correlations are very high (0.67 after oblimin rotation). For the Rasch model, 37 out of 704 response patterns have person-fit p -value smaller than 5%. With 2-PL and 3-PL models this figure does not improve, as with 2-PL we have 39 outlying person-fits and with 3-PL we have once again 37. For the unrotated two-factors item response analysis, it is difficult to interpret the second factor as loadings that are not close to zero are all negative and correspond to clearly nonzero loadings also for the first factor.

4.4. An Application to Anxiety and Depression

The Italian version of the Hospital Anxiety and Depression Scale (HADS, Zigmond & Snaith, 1983) consists of 14 polytomous items equally divided between the two dimensions: anxiety and depression. Hence, by definition, $s = 2$ in this example. The items of the questionnaire, which are an Italian validated translation (Costantini et al., 1999) of the original HADS questionnaire, are the following:

TABLE 14.
Unidimensional and two-dimensional classical item response analysis of NAEP data.

	Rasch model		2-PL		3-PL		Bidimensional analysis	
	Coeff	S.E.	Coeff	S.E.	Coeff	S.E.	F1	F2
Difficulty item 1	-1.24	0.07	-1.13	0.09	-1.13	0.09	-0.57	0.01
Difficulty item 2	-1.29	0.07	-0.97	0.07	-0.97	0.07	-0.64	-0.20
Difficulty item 3	-0.58	0.06	-0.55	0.07	-0.33	0.29	-0.47	-0.31
Difficulty item 4	-2.25	0.09	-1.74	0.12	-1.74	0.12	-0.71	-0.15
Difficulty item 5	0.23	0.06	0.28	0.08	0.55	0.32	-0.37	-0.14
Difficulty item 6	-1.23	0.07	-1.30	0.12	-1.29	0.12	-0.48	-0.02
Difficulty item 7	-0.60	0.06	-0.64	0.08	-0.24	0.35	-0.50	0.05
Difficulty item 8	-0.09	0.06	-0.07	0.05	-0.06	0.05	-0.62	-0.56
Difficulty item 9	-0.92	0.07	-0.66	0.05	-0.66	0.05	-0.70	-0.12
Difficulty item 10	-0.74	0.06	-0.64	0.07	-0.64	0.07	-0.55	-0.15
Difficulty item 11	1.13	0.07	1.00	0.08	1.03	0.09	-0.54	-0.13
Difficulty item 12	1.17	0.07	1.26	0.12	1.26	0.12	-0.48	0.00
Discrimination item 1	-	-	1.12	0.10	1.12	0.10	-	-
Discrimination item 2	-	-	1.54	0.13	1.53	0.12	-	-
Discrimination item 3	-	-	1.06	0.09	1.18	0.22	-	-
Discrimination item 4	-	-	1.45	0.14	1.45	0.14	-	-
Discrimination item 5	-	-	0.73	0.08	0.87	0.23	-	-
Discrimination item 6	-	-	0.91	0.09	0.91	0.09	-	-
Discrimination item 7	-	-	0.91	0.08	1.09	0.23	-	-
Discrimination item 8	-	-	1.67	0.13	1.66	0.14	-	-
Discrimination item 9	-	-	1.71	0.14	1.72	0.14	-	-
Discrimination item 10	-	-	1.20	0.10	1.20	0.10	-	-
Discrimination item 11	-	-	1.16	0.10	1.30	0.21	-	-
Discrimination item 12	-	-	0.89	0.09	0.89	0.09	-	-

TABLE 15.
Guessing parameters for the 3-PL model for NAEP data.

	Coeff	S.E.
Guessing item 1	0.00	0.01
Guessing item 2	0.00	0.00
Guessing item 3	0.09	0.12
Guessing item 4	0.00	0.01
Guessing item 5	0.09	0.11
Guessing item 6	0.00	0.02
Guessing item 7	0.15	0.13
Guessing item 8	0.00	0.02
Guessing item 9	0.00	0.01
Guessing item 10	0.00	0.00
Guessing item 11	0.03	0.03
Guessing item 12	0.00	0.00

TABLE 16.
Estimated posterior probabilities of the number of latent classes for the HADS data.

k	1	2	3	4	5	6	7	8	9	10
$p(k Y)$	0.0000	0.0000	0.0089	0.5838	0.3516	0.0557	0.0000	0.0000	0.0000	0.0000

TABLE 17.
Estimated prior and posterior probabilities of the number of dimensions and corresponding BFs for the HADS data.

s	1	2	3	4	5	6	7	8	9	10
$p(s k = 4)$	0.0018	0.0581	0.3511	0.4003	0.1585	0.0277	0.0022	0.0003	0.0000	0.0000
$p(s k = 4, Y)$	0.0000	0.1739	0.4349	0.3085	0.0766	0.0059	0.0001	0.0000	0.0000	0.0000
BF	0.0000	2.9932	1.2387	0.7708	0.4834	0.2134	0.0486	0.0000	–	–

1. I can laugh and see the funny side of things;
2. I get a sort of frightened feeling like butterflies in the stomach;
3. I have lost interest in my appearance;
4. I feel as if I am slowed down;
5. I look forward with enjoyment to things;
6. I get sudden feelings of panic;
7. I get a sort of frightened feeling as if something bad is about to happen;
8. Worrying thoughts go through my mind;
9. I feel cheerful;
10. I can sit at ease and feel relaxed;
11. I feel restless and have to be on the move;
12. I feel tense or wound up;
13. I still enjoy the things I used to enjoy;
14. I can enjoy a good book or radio or TV program.

Items 2, 6, 7, 8, 10, 11, 12 are classified as measuring anxiety, and the remaining ones as measuring depression. The data available to us pertain to $n = 201$ oncological Italian patients; for a detailed description see Bartolucci et al. (2015), Section 1.8. The responses have been dichotomized so that the observed binary variables correspond to the presence of a certain symptom, when equal to 1, or to its absence, when equal to 0.

We set $v_\alpha = 4$ and $v_\beta = 1$. For a model with $k = 4$ (found to be a reasonable number of latent classes in previous analyses of these data) these settings determine an expected value of the vector of the success probabilities of the four classes equal to (0.2000, 0.6250, 0.8182, 0.9286), whose elements are reasonably centered and separated, with corresponding variance values of (0.0267, 0.0260, 0.0124, 0.0044). Also, the unconditional prior probability for the number of dimensions is adequately spread over its support.

As usual, we run our algorithm for $T = 150000$ sweeps, after discarding 50 000 iterations as burn-in; regardless of k , $s = 1$ never occurred providing definitive evidence against unidimensionality. Table 16 shows the posterior distribution $p(k|Y)$. The values of k being all a priori equally probable, the model with $k = 4$ latent classes seems to be favored.

In Table 17, we show the estimated prior and posterior probabilities for s , as well as BFs conditionally on $k = 4$. According to the encompassing prior approach the case $s = 2$ is receiving the largest BF, corresponding to 2.99. The most visited partitions were those with the same number of items split in two groups (posterior probability of 49%). These partitions were also those receiving the highest BF (257.82). Among them, the most visited partition

was $\{(1, 2, 3, 4, 6, 7, 10), (5, 8, 9, 11, 12, 13, 14)\}$ (posterior probability of 78%). Since partitions consisting of two groups of seven items each are equally probable a priori, the most visited one is also the one with the highest BF and, thus, the one which is favored by the data.

We can compare our result with that obtained through the `MULTILCIRT` package, where the likelihood-based method of Bartolucci (2007) is implemented. In this case, assuming $k = 4$ latent classes and a Rasch parameterization, $s = 4$ groups of items seem to be detected on the basis of the likelihood ratio test: $\{(1, 2), (3, 4, 6, 7, 10), (5, 9, 12, 14), (8, 11, 13)\}$. The last step of the clustering algorithm produced exactly the same partition obtained through our method. However, it shall be noted again that these results are based on parametric assumptions that might not hold in general.

Finally, we remark that the algorithm we propose is quite fast. The whole analysis (sampling from the prior, sampling from the posterior, and post-processing both outputs) took around 47 minutes on an Intel Core i7 4810MQ CPU with a clock rate of 2.8 GHz, a time which is well comparable with the one required using the `MULTILCIRT` package (34 minutes).

5. Conclusions

We propose an approach to assess the number of dimensions measured by a set of dichotomously-scored items. The approach is based on nonparametric item response theory (IRT) models, which are constrained versions of the latent class (LC) model of Lazarsfeld and Henry (1968) formulated through a set of inequality constraints. This model is less restrictive than alternative models used to assess dimensionality, such as those adopted in Martin-Löf (1973), Christensen et al. (2002), and Bartolucci (2007), in which the distribution of the response variables given the latent trait is parametrized as in the Rasch or in the two-parameter logistic model (Rasch, 1961; Birnbaum, 1968).

Also considering the complexity of maximum likelihood estimation for the proposed nonparametric IRT models, we adopt a Bayesian framework based on the encompassing prior approach (Klugkist et al., 2005) to estimate such models. This amounts to assuming a system of priors for the parameters of the unconstrained LC model with a certain number of classes; then, the priors for the parameters of any constrained nonparametric IRT model with the same number of classes are “automatically” defined. These models are estimated by a reversible-jump Markov chain Monte Carlo sampler (Green, 1995; Green & Richardson, 2001), and a suitable post-processing algorithm is then used to specifically assess the model dimensionality.

Overall, an advantage of the proposed approach is that, given a certain dataset, it allows us to jointly determine the number of latent classes and the number of dimensions measured by the items, without requiring any parametric assumption on the distribution of the response variables given the latent class. This may lead to more reasonable results with respect to parametric/semiparametric approaches; see in particular the application in education here developed. Also note that the adoption of a latent class approach avoids the formulation of any parametric assumption on the distribution of the latent trait of interest in the population from which the observed sample is drawn and, at the same time, allows for a clustering of the subjects in terms of latent characteristics. Obviously, in applying the proposed approach, attention must be paid to the choice of the priors on the unconstrained LC model. In this regard, we propose a sensible system of priors, but a sensitivity analysis with respect to the parameters of these priors is mandatory in applications, as we show on the basis of the educational dataset.

Finally, the approach may be extended to the case of polytomously scored items in which every response variable may have more than two categories. However, this requires a proper definition of dimensionality that, when the response categories are ordered, is a natural extension of that given in the present article for dichotomous items. In any case, the system of priors and the inferential approach are essentially equivalent to those here proposed.

Appendix

Prior Probabilities of \mathcal{P} and s

In this section, we show how to calculate the prior probability of any partition \mathcal{P} and the prior probability of the number s of groups, conditionally on k , under independent uniform priors on the success probabilities λ_{jc} . Under more general priors of type (6), analytical derivation of the priors for \mathcal{P} and s becomes prohibitive and resorting to simulations is unavoidable.

We start by calculating the conditional prior probability of a single partition \mathcal{P} . For this aim, it is convenient to associate to each item j the ranking vector $\mathbf{q}_j = (q_{j1}, \dots, q_{jk})$, in which q_{jc} represents the rank of λ_{jc} after ordering $\lambda_{j1}, \dots, \lambda_{jk}$ in increasing order. Let $\mathbf{Q} = (\mathbf{q}'_1, \dots, \mathbf{q}'_r)$ be the $k \times r$ matrix of the vectors \mathbf{q}_j , arranged by columns. Each vector \mathbf{q}_{j_1} can have one out of $k!$ possible configurations, independently from any other \mathbf{q}_{j_2} , for $j_1 \neq j_2$. Thus, there are $(k!)^r$ possible configurations of the matrix \mathbf{Q} , all having the same probability under a uniform prior on the success probabilities λ_{jc} . Notice that \mathbf{Q} is a function of $\mathbf{\Lambda}$.

The number of matrices \mathbf{Q} determining the same partition \mathcal{P} is

$$k!(k-1) \cdots (k-s+1),$$

where $s = |\mathcal{P}|$. This is easily proved considering that the ranking vector of the first item in the first group can have any of the $k!$ possible configurations, which consequently determines the configuration for the ranking vectors of the other items in the same group; the ranking vector of the first item in the second group can have any possible configuration but that of the items in the first group, that is, $k! - 1$ possible configurations. By iterating this process, we conclude that the ranking vector of the first item in the s -th group can have any possible configuration but those of the items in the previous groups, that is, $k! - s + 1$ possible configurations. Therefore, the probability of any partition \mathcal{P} , conditionally on k is given by

$$p(\mathcal{P}|k) = \frac{k!(k-1) \cdots (k-s+1)}{(k!)^r}, \quad (15)$$

where the denominator corresponds to the number of possible configurations of \mathbf{Q} .

We now consider the prior conditional probability of the number s of groups, that is, $p(s|k)$. Since any partition into s groups has the same probability, given in (15), to obtain $p(s|k)$ we simply have to count how many partitions into s groups can be obtained and multiply this number by the probability of the single partition. The number of ways in which a set of r items can be partitioned into s nonempty groups is known as the Stirling number of the second kind (Graham et al., 1988), which is equal to

$$\left\{ \begin{matrix} r \\ s \end{matrix} \right\} = \frac{1}{s!} \sum_{h=0}^s (-1)^{s-h} \binom{s}{h} h^r,$$

with

$$\left\{ \begin{matrix} r \\ 1 \end{matrix} \right\} = \left\{ \begin{matrix} r \\ r \end{matrix} \right\} = 1.$$

Thus, the conditional probability of s can be obtained as

$$p(s|k) = \left\{ \begin{matrix} r \\ s \end{matrix} \right\} p(\mathcal{P}|k),$$

where $p(\mathcal{P}|k)$ is the probability of a partition \mathcal{P} such that $|\mathcal{P}| = s$, which is given in (15).

The RJ-MCMC Algorithm

In this section we provide a detailed description of the RJ-MCMC algorithm implemented in this paper. We first illustrate the fixed-dimension moves and then present the changing dimension moves for updating the number of latent classes.

Fixed-Dimension Moves

Model parameters are updated according to the following steps:

Update class weights. The full conditional distribution of $\boldsymbol{\pi}$ is a Dirichlet distribution with parameters $(1 + n_1, \dots, 1 + n_k)$, where $n_c = \#\{i : z_{ic} = 1\}$. The vector $\boldsymbol{\pi}$ is then updated by means of Gibbs sampling, drawing a new vector from such a distribution.

Update latent class variables. The latent class allocation variables z_{ic} , $i = 1, \dots, n$, $c = 1, \dots, k$, can be updated by means of Gibbs sampling, drawing them independently from their full conditional

$$p(z_{ic} = 1 | \dots) = \frac{\pi_c \prod_{j=1}^r (1 - \lambda_{jc})^{1 - y_{ij}} \lambda_{jc}^{y_{ij}}}{\sum_{c=1}^k \pi_c \prod_{j=1}^r (1 - \lambda_{jc})^{1 - y_{ij}} \lambda_{jc}^{y_{ij}}},$$

where “ \dots ” denotes “all other parameters and data.”

Update the success probabilities. In order to update the success probabilities, we build a proposal based on independent zero-centered normal increments of logit (λ_{jc}) , separately for each $j = 1, \dots, r$ and $c = 1, \dots, k$. The candidate λ_{jc}^* is accepted with probability equal to $\min(1, p_{\lambda_{jc}^*})$, where

$$\begin{aligned} p_{\lambda_{jc}^*} = & \prod_{i=1}^n \{(\lambda_{jc}^*/\lambda_{jc})^{y_{ij}} [(1 - \lambda_{jc}^*)/(1 - \lambda_{jc})]^{(1 - y_{ij})}\}^{z_{ic}} \\ & \times (\lambda_{jc}^*/\lambda_{jc})^{(\alpha_c - 1)} [(1 - \lambda_{jc}^*)/(1 - \lambda_{jc})]^{(\beta_c - 1)} \\ & \times (\lambda_{jc}^*)(1 - \lambda_{jc}^*) / [(\lambda_{jc})(1 - \lambda_{jc})]. \end{aligned} \quad (16)$$

The first line on the right side is the likelihood ratio, while the second line corresponds to the ratio between the prior densities. The ratio between the proposal densities cancels out, apart from the Jacobian of the logit transformation, given in the third line of (16).

Changing Dimension Moves for Updating k

Split/merge move. For the combine proposal, we pick a class at random among $2, \dots, k$, with probability $1/(k - 1)$, and denote it with c_2 . Then, we draw another class at random among $1, \dots, c_2 - 1$, with probability $1/(c_2 - 1)$, and denote it with c_1 . Classes c_1 and c_2 are then merged into a new class c' , decreasing k by 1, with the merged class c' occupying the place $c_2 - 1$, once the place c_1 has been deleted. We then create new values for $\pi_{c'}$ and $\lambda_{jc'}$, for $j = 1, \dots, r$, and we reallocate all those observations for which $z_{ic_1} = 1$ or $z_{ic_2} = 1$ to the merged class c' .

A new vector of weights is created by letting

$$\pi_{c'} = \pi_{c_1} + \pi_{c_2}.$$

The new parameters $\lambda_{jc'}$ are created in such a way that $\lambda_{jc'} = \lambda_{jc_2}$, while the λ_{jc_1} are simply deleted for all j .

The split proposal begins by choosing a class at random among $1, \dots, k$, say c' , with probability $1/k$ and splitting it into two new classes, labeled c_1 and c_2 , augmenting k by 1. Let c_2 take the place $c' + 1$ and c_1 be inserted in a place preceding c_2 and chosen at random with probability $1/c'$. We then need to create new values for π_{c_1}, π_{c_2} and for $\lambda_{jc_1}, \lambda_{jc_2}, j = 1, \dots, r$, and reallocate all those observations for which $z_{ic'} = 1$, while the hyper-parameters α_c and β_c are simply recalculated as $\alpha_c^* = \nu c + 1$ and $\beta_c^* = \nu(k^* + 1 - c) + 1$, for $c = 1, \dots, k^*$, with $k^* = k + 1$.

Let us start by splitting the weight $\pi_{c'}$ into π_{c_1} and π_{c_2} in such a way that $\pi_{c_1} + \pi_{c_2} = \pi_{c'}$. We accomplish this by generating a random value $u \sim \text{Beta}(u_1, u_2)$, where u_1 and u_2 are the parameters of the Beta density, and setting

$$\pi_{c_1} = \pi_{c'}u \quad \text{and} \quad \pi_{c_2} = \pi_{c'}(1 - u).$$

The new vector of weights is denoted by $\boldsymbol{\pi}^*$.

We then split the parameters $\lambda_{jc'}$ into λ_{jc_1} and λ_{jc_2} for all j . We accomplish this by sampling a vector $\boldsymbol{w} = (w_1, \dots, w_r)$ from the prior distribution of λ_{jc_1} , that is, $w_j \sim \text{Beta}(\alpha_{c_1}^*, \beta_{c_1}^*)$, for $j = 1, \dots, r$, and thus setting

$$\lambda_{jc_1} = w_j \quad \text{and} \quad \lambda_{jc_2} = \lambda_{jc'}, \quad j = 1, \dots, r.$$

The new matrix of success probabilities is denoted by $\boldsymbol{\Lambda}^*$.

Finally, we reallocate all those observations for which $z_{ic'} = 1$ between c_1 and c_2 in a way analogous to the standard Gibbs allocation move and we let \boldsymbol{Z}^* denote the new allocation matrix.

According to the RJ framework, the acceptance probability for the split move is $\min(1, p_k)$, where

$$\begin{aligned} p_k &= \prod_{i=1}^n \prod_{j=1}^r \frac{[(\lambda_{jc_1}^*)^{y_{ij}} (1 - \lambda_{jc_1}^*)^{(1-y_{ij})}]^{z_{ic_1}^*} [(\lambda_{jc_2}^*)^{y_{ij}} (1 - \lambda_{jc_2}^*)^{(1-y_{ij})}]^{z_{ic_2}^*}}{[(\lambda_{jc'}^*)^{y_{ij}} (1 - \lambda_{jc'}^*)^{(1-y_{ij})}]^{z_{ic'}^*}} \\ &\times \frac{p(k^*)}{p(k)} \times \frac{\mathcal{D}(\pi_1^*, \dots, \pi_{k^*}^*)}{\mathcal{D}(\pi_1, \dots, \pi_k)} \times \frac{(\pi_{c_1}^*)^{\sum_{i=1}^n z_{ic_1}^*} (\pi_{c_2}^*)^{\sum_{i=1}^n z_{ic_2}^*}}{(\pi_{c'}^*)^{\sum_{i=1}^n z_{ic'}^*}} \\ &\times \frac{\prod_{j=1}^r \prod_{c=1}^{k^*} \mathcal{B}_{\alpha_c^*, \beta_c^*}(\lambda_{jc}^*)}{\prod_{j=1}^r \prod_{c=1}^k \mathcal{B}_{\alpha_c, \beta_c}(\lambda_{jc})} \times \frac{d_{k^*}}{b_k p_{\text{alloc}} \mathcal{B}_{u_1, u_2}(u) \prod_{j=1}^r \mathcal{B}_{\alpha_{c_1}^*, \beta_{c_1}^*}(w_j)} \times \pi_{c'}. \end{aligned} \quad (17)$$

The first two lines in (17) represent, respectively, the likelihood and the priors ratio, with \mathcal{D} being the Dirichlet density with parameters all equal to 1, and \mathcal{B} being the Beta density with parameters specified in the subscript. In the third line, the first term represents the proposal ratio, with p_{alloc} being the probability of the particular allocation made in the split move, and the second term is the Jacobian of the transformation from $(\pi_{c'}, \lambda_{1c'}, \dots, \lambda_{rc'}, u, w_1, \dots, w_r)$ to $(\pi_{c_1}^*, \lambda_{1c_2}^*, \dots, \lambda_{rc_2}^*, \pi_{c_2}^*, \lambda_{1c_1}^*, \dots, \lambda_{rc_1}^*)$. The combine move is accepted with probability $\min(1, p_k^{-1})$, with some obvious substitutions in the expression for p_k .

Birth/death move. We first make a random choice between birth and death, using the same probabilities b_k and d_k as above. For a birth, we pick a position at random among $1, \dots, k^*$, with probability $1/k^*$ for the place to be occupied by the new class c' . Then, a weight and a vector of success probabilities for the new class are drawn using

$$\pi_{c'} \sim \text{Beta}(1, k), \quad \lambda_{jc'} \sim \text{Beta}(\alpha_{c'}^*, \beta_{c'}^*), \quad j = 1, \dots, r,$$

where the hyper-parameters α and β are recalculated as in the split move, leading to α^* and β^* , respectively. To “make space” for the new class, the existing weights are rescaled, so that all weights sum to 1, using $\pi_c^* = \pi_c(1 - \pi_{c'})$. The weights and the success probabilities proposed under the birth move are denoted by π^* and Λ^* , respectively. The allocation variables remain unchanged, and no subjects are allocated to the new component, that is, $z_{ic'} = 0, i = 1, \dots, n$. The increased allocation matrix is indicated as Z^* .

For a death, a random choice is made between any existing empty components; the chosen component is deleted and the remaining weights are rescaled to sum to 1. No other changes are proposed to the variables and, in particular, the allocations are unaltered.

The acceptance probabilities for birth and death are $\min(1, p_k)$ and $\min(1, p_k^{-1})$, respectively, where

$$p_k = \frac{p(k^*)}{p(k)} \times \frac{\mathcal{D}(\pi_1^*, \dots, \pi_{k^*}^*)}{\mathcal{D}(\pi_1, \dots, \pi_k)} \times \frac{\prod_{c=1}^{k^*} (\pi_c^*)^{\sum_{i=1}^n z_{ic}^*}}{\prod_{c=1}^k \pi_c^{\sum_{i=1}^n z_{ic}}} \times \frac{\prod_{j=1}^r \prod_{c=1}^{k^*} \mathcal{B}_{\alpha_{c'}^*, \beta_{c'}^*}(\lambda_{jc}^*)}{\prod_{j=1}^r \prod_{c=1}^k \mathcal{B}_{\alpha_c, \beta_c}(\lambda_{jc})} \\ \times \frac{d_{k^*} k^*}{b_k(k_0 + 1) \mathcal{B}_{1,k}(\pi_{c'}^*) \prod_{j=1}^r \mathcal{B}_{\alpha_{c_1}^*, \beta_{c_1}^*}(\lambda_{jc'}^*)} \times (1 - \pi_{c'}^*), \quad (18)$$

with k_0 being the number of empty classes, before the birth. In equation (18), the first line is the prior ratio, and the second line contains the proposal ratio and the Jacobian; the likelihood ratio is equal to 1.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium of information theory* (pp. 267–281). Budapest: Akademiai Kiado.
- Bacci, S., & Bartolucci, F. (2016). Two-tier latent class IRT models in R. *The R Journal*, 8, 139–166.
- Bacci, S., Bartolucci, F., & Gnaldi, M. (2014). A class of multidimensional latent class IRT models for ordinal polytomous item responses. *Communication in Statistics - Theory and Methods*, 43, 787–800.
- Bartolucci, F. (2007). A class of multidimensional IRT models for testing unidimensionality and clustering items. *Psychometrika*, 72, 141–157.
- Bartolucci, F., Bacci, S., & Gnaldi, M. (2015). *Statistical analysis of questionnaires: A unified approach based on Stata and R*. Boca Raton, FL: Chapman and Hall/CRC Press.
- Bartolucci, F., Bacci, S., & Gnaldi, M. (2016). MultiLCIRT: multidimensional latent class item response theory models. In R package version, version 2.10. <https://cran.r-project.org/web/packages/MultiLCIRT/index.html>.
- Bartolucci, F., & Forcina, A. (2005). Likelihood inference on the underlying structure of IRT models. *Psychometrika*, 70, 31–43.
- Bartolucci, F., Scaccia, L., & Farcomeni, A. (2012). Bayesian inference through encompassing priors and importance sampling for a class of marginal models for categorical data. *Computational Statistics and Data Analysis*, 56, 4067–4080.
- Béguin, A. A., & Glas, C. A. W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, 66, 541–561.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee’s ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395–479). Reading, MA: Addison-Wesley.

- Bock, R., Gibbons, R. D., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, *12*, 261–280.
- Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement*, *27*, 395–414.
- Chalmers, P., Pritikin, J., Robitzsch, A., Zoltak, M., Kim, K., Falk, C. F., & Meade, A. (2017). MIRT: Multidimensional item response theory. In *R package version, version 1.23*. <https://cran.r-project.org/web/packages/mirt/index.html>.
- Christensen, K. B., Björner, J. B., Kreiner, S., & Petersen, J. H. (2002). Testing unidimensionality in polytomous Rasch models. *Psychometrika*, *67*, 563–574.
- Costantini, M., Musso, M., Viterbori, P., Bonci, F., Del Mastro, L., Garrone, O., et al. (1999). Detecting psychological distress in cancer patients: Validity of the Italian version of the hospital anxiety and depression scale. *Support Care Cancer*, *7*, 121–127.
- Diebolt, J., & Robert, C. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Series B*, *56*, 363–375.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*, 215–231.
- Graham, R. L., Knuth, D. E., & Patashnik, O. (1988). *Concrete mathematics: A foundation for computer science*. Reading, MA: Addison-Wesley.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*, 711–732.
- Green, P. J., & Richardson, S. (2001). Hidden Markov models and disease mapping. *Journal of the American Statistical Association*, *97*, 1055–1070.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer Nijhoff.
- Hojtink, H., & Molenaar, I. W. (1997). A multidimensional item response model: Constrained latent class analysis using the Gibbs sampler and posterior predictive checks. *Psychometrika*, *62*, 171–189.
- Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307.
- Junker, B. W., & Sijtsma, K. (2001). Nonparametric item response theory in action: An overview of the special issue. *Applied Psychological Measurement*, *25*, 211–220.
- Karabatsos, G. (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement*, *2*, 389–423.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.
- Klugkist, I., Kato, B., & Hoijtink, H. (2005). Bayesian model selection using encompassing priors. *Statistica Neerlandica*, *59*, 57–69.
- Kuo, T., & Sheng, Y. (2015). Bayesian estimation of a multi-unidimensional graded response IRT model. *Behaviormetrika*, *42*, 79–94.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Lindley, D. V. (1957). A statistical paradox. *Biometrika*, *44*, 187–192.
- Lindsay, B., Clogg, C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, *86*, 96–107.
- Martin-Löf, P. (1973). *Statistiska modeller*. Stockholm: Institutet för Försäkringsmatematik och Matematisk Statistisk vid Stockholms Universitet.
- Pan, J. C., & Huang, G. H. (2014). Bayesian inferences of latent class models with an unknown number of classes. *Psychometrika*, *79*, 621–646.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability*, *4*, 321–333.
- Reckase, M. D. (2009). *Multidimensional item-response theory*. New York: Springer.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, *22*, 1701–1762.
- Tuyl, F., Gerlach, R., & Mengersen, K. (2009). Posterior predictive arguments in favor of the Bayes–Laplace prior as the consensus prior for binomial and multinomial parameters. *Bayesian Analysis*, *4*, 151–158.
- Van Onna, M. J. H. (2002). Bayesian estimation and model selection in ordered latent class models for polytomous items. *Psychometrika*, *67*, 519–538.
- Verhelst, N. D. (2001). Testing the unidimensionality assumption of the Rasch model. *Methods of Psychological Research Online*, *6*, 231–271.
- Vermunt, J. K. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Applied Psychological Measurement*, *25*, 283–294.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*, 287–307.
- Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*, 361–370.

Manuscript Received: 11 JAN 2015

Final Version Received: 2 MAY 2017

Published Online Date: 12 SEP 2017