



LASSO-Based Pattern Recognition for Replenished Items With Graded Responses in Multidimensional Computerized Adaptive Testing

Jianan Sun^{1*}, Ziwen Ye², Lu Ren³ and Jingwen Li¹

¹ College of Science, Beijing Forestry University, Beijing, China, ² Department of Statistics, Columbia University, New York, NY, United States, ³ School of Mathematical Sciences, Beihang University, Beijing, China

OPEN ACCESS

Edited by:

Jin Eun Yoo,
Korea National University
of Education, South Korea

Reviewed by:

Jason C. Immekus,
University of Louisville, United States
Xiaojian Sun,
Southwest University, China

*Correspondence:

Jianan Sun
jnsun@bjfu.edu.cn

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 23 February 2022

Accepted: 09 May 2022

Published: 17 June 2022

Citation:

Sun J, Ye Z, Ren L and Li J (2022)
LASSO-Based Pattern Recognition
for Replenished Items With Graded
Responses in Multidimensional
Computerized Adaptive Testing.
Front. Psychol. 13:881853.
doi: 10.3389/fpsyg.2022.881853

As a branch of statistical latent variable modeling, multidimensional item response theory (MIRT) plays an important role in psychometrics. Multidimensional graded response model (MGRM) is a key model for the development of multidimensional computerized adaptive testing (MCAT) with graded-response data and multiple traits. This paper explores how to automatically identify the item-trait patterns of replenished items based on the MGRM in MCAT. The problem is solved by developing an exploratory pattern recognition method for graded-response items based on the least absolute shrinkage and selection operator (LASSO), which is named LPRM-GR and facilitates the subsequent parameter estimation of replenished items and helps maintaining the effectiveness of item replenishment in MCAT. In conjunction with the proposed approach, the regular BIC and weighted BIC are applied, respectively, to select the optimal item-trait patterns. Simulation for evaluating the LPRM-GR in pattern recognition accuracy of replenished items and the corresponding item estimation accuracy is conducted under multiple conditions across different numbers with respect to dimensionality, response-category numbers, latent trait correlation, stopping rules, and item selection criteria. Results show that the proposed method with the two types of BIC both have good performance in pattern recognition for item replenishment in the two- to four-dimensional MCAT with the MGRM, for which the weighted BIC is generally superior to the regular BIC. The proposed method has relatively high accuracy and efficiency in identifying the patterns of graded-response items, and has the advantages of easy implementation and practical feasibility.

Keywords: multidimensional graded response model, multidimensional computerized adaptive testing, replenished items, item-trait pattern recognition, LASSO, weighted BIC

INTRODUCTION

With the rapid development of modern assessment theory and computer technology, computerized adaptive testing (CAT; e.g., Wainer, 2000) has become a hot issue during the past several decades. The purpose of CAT is to understand the potential characteristics of examinees as accurately as possible. Constructing the optimal test for each examinee requires modern assessment theory

such as item response theory (IRT) and well-established latent variable models (e.g., Embretson and Reise, 2000; Reckase, 2009; Bartholomew et al., 2011). In psychometrics, a wide variety of psychological scales are designed to measure multiple latent traits within the framework of multidimensional item response theory (MIRT). Since the IRT-based CAT has received a lot of attention from psychometricians, the proposed method in this paper is developed within the framework of the MIRT-based CAT, which is briefly referred to as multidimensional CAT (MCAT) for simplicity. A non-negligible aspect of multidimensional items fitted by MIRT models is the corresponding relationship between the items and multiple traits measured by the overall scale, which can be simply named item-trait patterns (Sun and Ye, 2019). Specifically, item-trait pattern represents the appropriate set of the latent traits that are closely associated with an item. For instance, if a scale measures K traits, the true item-trait pattern of the j th item can be formulated as a vector: $\mathbf{Q}_j = (q_{j1}, \dots, q_{jK})^T$. If the k th trait is measured by the item, $q_{jk} = 1$; else, $q_{jk} = 0$ ($k = 1, \dots, K$). Misspecification of the item-trait patterns of the new or replenished items fitted by MIRT models may lead to serious lack-of-fit and faulty assessment.

Item replenishment is an essential part of MCAT in view of test security and reliability. In fact, the MCAT developed on MIRT models has attracted growing attention in psychometrics. An important component of MCAT is the quality of the item pool of MCAT, the statistical perspective of which is greatly affected by the calibration of item parameters. As pointed out by Wainer and Mislevy (1990), CAT is usually administered to examinees at frequent time intervals, some operational items in the item pool may become obsolete or overexposed over time, and the frequently used items should be replaced by replenished items considering the following reasons such as the item exposure control, safety, fairness, and reliability of the test. Also, for the MCAT, it is often encountered that some operational items in the item pool will no longer be applicable due to the similar reasons as above, so the item pool needs updating in a timely manner. Item-trait pattern and parameter estimate jointly convey the core statistical information of a multidimensional replenished item, and the identification of the item-trait pattern undoubtedly affects estimation accuracy and application of the item. For the multidimensional paper-and-pencil test, a latent variable selection method for the MIRT models *via* the least absolute shrinkage and selection operator (LASSO) is proposed by Sun et al. (2016). The research assumed the latent traits were entirely unknown and formulated the item-trait pattern recognition problem as the latent variable selection. For replenished items of MCAT, although domain experts can design which latent traits are measured by the overall MCAT, the patterns in the pool are difficult to know precisely and efficiently. Moreover, the parameter scaling consistency among all items should be ensured. Therefore, manually identifying the patterns of replenished items may face the problem of not only the heavy workload but also the risk of inconsistent scaling of item parameters. For the above-mentioned reasons, one topic of item pool replenishment for the MCAT is the timely and accurate identification of item-trait patterns for replenished items (Sun and Ye, 2019). The specific dimensions measured by an item can be briefly represented as

the item-trait pattern, which reflects the goodness-of-fit for the item based on the designed MIRT model and closely relates to the estimation accuracy of latent traits. The misspecification of item-trait patterns may lead to serious lack-of-fit and faulty assessment (e.g., Reckase, 2009; Sun et al., 2016; Sun and Ye, 2019). For the MCAT based on MIRT models, it is therefore appropriate to conduct the study on identifying the patterns for replenished items.

A variety of models have been proposed in the MIRT framework. Compensatory and non-compensatory models are seen as two major types of parametric MIRT models. The most commonly used compensatory MIRT models in MCAT are multidimensional two parameter logistic model (M2PLM; McKinley and Reckase, 1982), multidimensional three parameter logistic model (Reckase and McKinley, 1991), multidimensional partial credit model (Kelderman and Rijkes, 1994), multidimensional graded response model (MGRM; Muraki and Carlson, 1995). For the compensatory MIRT model, M2PLM, Sun and Ye (2019) propose a pattern recognition method based on LASSO to detect the optimal item-trait patterns of the replenished items. The method is named LPRM, which is applicable to the MCAT with dichotomous items and has been shown to make the precise and effective detection of item-trait patterns for replenished items in the multidimensional item pool. The LPRM utilizes the online feature of the MCAT to successfully achieve the goal of automatic scaling for replenished and operational items in the item pool, so that the interpretability of replenished items in terms of goodness-of-fit can be improved. In addition, the method for dichotomous items is well viable because it is sufficiently compatible and adaptable to the conventional design of the MCAT. Consequently, the LPRM is well-suited for dichotomous items with M2PLM, which can save time cost compared to manual identification.

The research scenario of LPRM can be expanded, such as establishing a pattern recognition framework for the MCAT with polytomous items. Note that for the multidimensional graded-response data, the MGRM is an essential compensatory MIRT model for the paper-and-pencil test or the MCAT (e.g., Jiang et al., 2016; Depaoli et al., 2018; Tu et al., 2018; Wang et al., 2018a, Wang et al., 2019; Nouri et al., 2021), so it is necessary to extend the item-trait pattern recognition idea for the items with dichotomous responses to those with graded responses. Moreover, the LPRM was developed directly based on the M2PLM. With the aim of solving the pattern recognition problem for the MCAT based on the MGRM, although the general idea of the LPRM can be similarly used, it is still necessary to detail how to extend the specific approach for the model such as designing the MCAT scenarios, constructing the corresponding L_1 -regularized optimization and deriving the algorithm, and exploring how to choose the optimal patterns for replenished items. Based on these considerations, this paper highlights the pattern recognition for the replenished items with graded responses and proposes the LPRM-GR for it.

The rest of this paper is organized as follows. Section "Methodology" gives a brief introduction to the research background in several aspects: two compensatory MIRT models, the LASSO and the pattern recognition idea for dichotomous

items as well as the LPRM; then the section introduces the LPRM-GR in detail for illustrating how to detect the optimal item-trait patterns of replenished items based on the MGRM in MCAT. Section “Simulation” evaluates the performance of the proposed method in pattern recognition by simulation. Section “Conclusion and discussion” summarizes the conclusion and makes the further discussion of the study.

METHODOLOGY

Two Compensatory Multidimensional Item Response Theory Models

Multidimensional Two-Parameter Logistic Model

One widely used compensatory MIRT model for dichotomous items is the M2PLM. Sun and Ye (2019) investigate the efficacy of the M2PLM based on the LASSO for detecting item-trait patterns of the replenished items in MCAT, and showed that the LASSO significantly contributes to the detection of optimal patterns of replenished items in MCAT. For a dichotomous-response test, the M2PLM is here briefly introduced: assume that the abilities of examinees follow a normal distribution. The probability of an examinee with abilities θ correctly answering the j th item defined by the M2PLM (e.g., Reckase, 2009) is

$$P(U_j = 1|\theta) = \frac{1}{1 + \exp[-(\mathbf{a}_j^T \theta + b_j)]}, \quad (1)$$

where U_j is a binary random variable for the response of the examinee to the j th item. The discrimination parameters of the j th item are denoted as the vector: \mathbf{a}_j . The intercept parameter of the j th item is denoted as b_j .

Multidimensional Graded Response Model

For a polytomous-response test, MGRM is one of the most popular MIRT models discussed by psychometricians in the past decades. As a polytomous extension of M2PLM, MGRM can appropriately analyze and score ordered categorical data, such as in a Likert-type rating scale (e.g., Jiang et al., 2016). Assume that N examinees are taking a multidimensional test with J graded-response items measuring K latent traits of interest. According to the logistic-form of the MGRM, the probability of the i th examinee to the j th item receiving a score of m or above is defined as

$$P_{ijm}^*(\theta_i) = P^*(Y_{ij} \geq m|\theta_i) = \frac{1}{1 + \exp[-\sum_{k=1}^K a_{jk}(\theta_{ik} - d_{jm})]} \quad (m = 0, 1, \dots, M) \quad (2)$$

where Y_{ij} refers to the score taking an integer from 0 to M ; $\{a_{jk}\}$ ($k = 1, \dots, K$) and $\{d_{jm}\}$ ($m = 0, \dots, M$) are discrimination parameters and boundary parameters, respectively. By denoting $b_{jm} = (-\sum_{k=1}^K a_{jk})d_{jm}$ (i.e., the intercept parameter for j and m), $\mathbf{a}_j = (a_{j1}, \dots, a_{jK})^T$ and $\theta_i = (\theta_{i1}, \dots, \theta_{iK})^T$ ($i = 1, \dots, N$), Equation (2) is rewritten as $P_{ijm}^*(\theta_i) = \frac{1}{1 + \exp[-(\mathbf{a}_j^T \theta_i + b_{jm})]}$. Note that the constraints should be

satisfied: $P_{ij0}^*(\theta_i) = 1$, $P_{ij(M+1)}^*(\theta_i) = 0$, and $+\infty = b_{j0} > \dots > b_{jM} = -\infty$. With the definition of boundary function, the probability of the i th examinee receiving the score of m to the j th item is expressed as

$$P_{ijm}(\theta_i) = P(Y_{ij} = m|\theta_i) = P_{ijm}^*(\theta_i) - P_{ij(m+1)}^*(\theta_i). \quad (3)$$

Note that MGRM is a compensatory item factor model, so here gives a brief discussion about the identifiability of the discrimination parameters of the MGRM. If the ability parameters are unknown, the MMLE/EM algorithm (e.g., Cai, 2013) is commonly used to estimate item parameters under the condition of some or all patterns are well-known. For example, Jiang et al. (2016) and Wang et al. (2018a) both assumed that the simple structures of the overall items are satisfied for estimating the item parameters for graded-response items *via* the EM algorithm: for each item, only one element of the \mathbf{a} vector was non-zero, and every dimension was represented by an equal number of items. Following the usual confirmatory factor analysis, other reasonable structures for the overall items based on the MGRM can also be specified before estimation. However, if both the ability parameters and most of the item-trait patterns are unknown, constraints need to be imposed on the discrimination parameters to guarantee their identifiability, which follows the usual exploratory factor analysis. The ways of giving constraints for the discrimination parameters can be various, the basic idea for which is pre-specifying the patterns of a few items so that it is possible for the other items to obtain reasonable results in terms of estimating discrimination parameters *via* the EM algorithm. For instance, one approach is to select K items that can measure all the K latent traits, and then pre-specify their item-trait patterns. More detailed discussion can be found in Béguin and Glas (2001), and Sun et al. (2016). If the abilities of the MGRM are known, it is not necessary to impose any constraint on the discrimination parameters or pre-specify any item-trait pattern for the identification of the discrimination parameters.

Least Absolute Shrinkage and Selection Operator-Based Pattern Recognition for Dichotomous Items

The most widely used approaches of statistical variable selection in regression analysis before the 1990s are forward selection, backward elimination, stepwise selection, and ridge regression (e.g., Hoerl and Kennard, 2000). Ridge regression is also known as the regression based on L_2 regularization. By introducing the L_2 regularization term to the optimization problem of coefficient estimation, ridge regression produces shrinkage of the size of the regression coefficients to mitigate the estimation problem due to multicollinearity in linear regression. Ridge regression improves prediction accuracy more efficiently than the previous variable selection approaches, especially for the linear models suffering a large number of covariates and multicollinearity problems. The LASSO optimization is one of the members of the penalized least squares, the idea of which can be extended naturally to likelihood-based models (e.g., Tibshirani, 1996; Fan and Li, 2001). The LASSO penalty corresponds to a Laplace

prior, which expects many coefficients to be close to zero, and a small subset to be larger and non-zero, so the LASSO tends to pick relatively simple combination structure from the alternative predictors (Tibshirani, 1996; Friedman et al., 2007). Specifically, the LASSO is usually labeled as the L_1 -constrained fitting or the L_1 regularization for regression models. The optimizations based on the ridge regression and the LASSO can both consider the complexity of the regression models. In contrast to ridge regression, the LASSO can furtherly improve model interpretability by minimizing the sum of squared errors with a bound on the sum of the absolute values of the regression coefficients, which tends to produce some coefficients directly to be zero (Tibshirani, 1996). To be brief, the LASSO is a regression method that involves penalizing the absolute size of the regression coefficients and is considered as one of the most popular methods for variable selection in multivariate regression analysis. Two essential features of the LASSO are the shrinkage and selection of regression variables.

For MIRT models, the problems of variable selection based on the LASSO have ever been discussed in the research of latent variable selection (Sun et al., 2016). For the pattern recognition problem for replenished items in MCAT based on the M2PLM, the LPRM based on the LASSO is proposed by Sun and Ye (2019). Penalized regression methods can also be taken into consideration for analyzing the data of large-scale surveys with missing categorical predictors, which inspires a new perspective for the effective and feasible treatment of missing data in the field of large-scale testing (e.g., Yoo, 2018; Yoo and Rho, 2020, 2021).

The LPRM generally collects the information in terms of examinees' ability estimates and responses to replenished items *via* the MCAT to prepare for the item-trait pattern recognition, uses the LASSO to obtain alternative patterns of replenished items, and finally applies the regular Bayesian information criterion (BIC) to select the optimal patterns from the alternatives. Details of the roles played by the MCAT and the LASSO in LPRM can be understood by reviewing the three steps of the LPRM. The first step is to operate the MCAT measuring K latent traits of interest with N examinees. A certain number of operational items are selected from all the operational items by the given item selection method in MCAT. A fixed number of replenished items are specified among all the replenished items in the item pool. The appropriate number of administered replenished items can be designed to make sure that the length of the MCAT is workable for examinees. For instance, if the number of replenished items is J_1 , N examinees can be divided into 5 groups and every $J_1/5$ replenished items are answered by each group. Note that the replenished items are also assumed to be consistent with the operational ones so that the examinees share the same test motivation for the two item types. All answers are recorded and scored, which aims to get valuable information in terms of both examinees' ability estimates and the responses to replenished items. Secondly, the LASSO or L_1 -regularized optimization for the j th replenished item is formulated in Equation (4), which can be considered the essential part of the second step of the LPRM. Because θ_j and \mathbf{a}_j in Equation (2) play the role of covariates and coefficients, respectively, the sparsity

of \mathbf{a}_j indicates the item-trait patterns. That is, if $\hat{a}_{jk} > 0$, then $q_{jk} = 1$; else $q_{jk} = 0$.

$$\min_{b_j, a_{j1}, \dots, a_{jK}} \left\{ -l(b_j, a_{j1}, \dots, a_{jK}; \mathbf{Y}_j, \Theta) + \lambda \sum_{k=1}^K |a_{jk}| \right\}. \quad (4)$$

Here continuing the notations in Equation (1), the abilities in MCAT are denoted by $\Theta = (\theta_{ik})_{N \times K}$. In Equation (4): $l(b_j, a_{j1}, \dots, a_{jK}; \mathbf{Y}_j, \Theta)$ is the log-likelihood based on the observed data, \mathbf{Y}_j , and the latent variables, Θ . Note that for the LPRM, Θ can be substituted by ability estimates. The L_1 -norm of discrimination parameters for the j th item is denoted as $\sum_{k=1}^K |a_{jk}|$. The tuning parameter (also named regularization parameter), λ , mainly controls the sparsity of the discrimination parameters. A group of non-negative values of λ is given to produce adequate alternative patterns according to the corresponding L_1 -regularized optimizations for alternative values of λ . Coordinate descent algorithm can be used to solve the optimizations (e.g., Friedman et al., 2010). Thirdly, the regular BIC is applied to choose the optimal item-trait patterns of replenished items from the alternative ones that are obtained in the second step.

Primary findings of the LPRM that may enlighten the study of this paper are summarized as follows. Firstly, for the replenished items in the MCAT item pool based on M2PLM, the correct specification rates for which patterns detected by the LPRM are above 80% and almost above 90% in the simulation of Sun and Ye (2019). Because the true abilities are unknown in practice, the LPRM can generally get comparatively sufficient ability information from ability estimates and select good patterns for replenished items. Secondly, the item parameters of the replenished items are estimated based on the patterns identified by the LPRM with the aim of investigating how the recognition accuracy of the patterns influences the estimation accuracy of item parameters, the finding for which is that the former affect the latter positively. Thirdly, the operational items with high discrimination can help the LPRM yield comparatively good performance in identifying patterns and getting the estimates of discrimination parameters. Fourthly, in practice, there is a trade-off for the LPRM to choose which of the variable-length MCAT and the fixed-length MCAT should be used, and the possible factors include the sizes of discrimination parameters, computing efficiency of the two stopping rules and other practical considerations about test safety and item exposure.

Least Absolute Shrinkage and Selection Operator-Based Pattern Recognition for Graded-Response Items

In this section, we propose the framework of the LASSO-based pattern recognition method for the item pool with graded response items in the MCAT. The proposed method is named LPRM-GR, the aim of which is to effectively give an appropriate polytomous extension of the LPRM: specially developed for the

MCAT constructed based on the MGRM. The detailed steps of the proposed method in this paper are illustrated as follows.

Step 1. Consider a situation that K latent traits of interest are measured by an MCAT item pool, which includes fixed numbers of operational and replenished items. When an MCAT with N examinees is organized, appropriate numbers of operational and replenished items can be designed in the test for balancing the goal of individual measurement and collection of responses to replenished items. For instance, the numbers of two types of items in the item pool are J_0 and J_1 , respectively. At the same time, assume that it has been ensured that the examinees have the same motivation for both types of items in the test. Note that the test length here should also take into account a combination of considering different types of stopping rules and avoiding fatigue effects. Note that the design for the LPRM-GR here is similar to the first step of the LPRM.

Step 2. For all the replenished items in the item pool of the MCAT, a range of non-negative values of the tuning parameter for the LASSO, $\lambda_1, \dots, \lambda_W$, should be given for providing various shrinkage effects for discrimination parameters of the MGRM. Specifically, the L_1 -regularized optimization constructed for the j th item with a given λ_w is defined as

$$\min_{b_{j1}, \dots, b_{j(M-1)}, a_{j1}, \dots, a_{jK}, \mathbf{Y}_j, \hat{\Theta}} \{-l(b_{j1}, \dots, b_{j(M-1)}, a_{j1}, \dots, a_{jK}, \mathbf{Y}_j, \hat{\Theta}) + \lambda_w \sum_{k=1}^K |a_{jk}|\}, \quad (5)$$

where the true abilities are denoted as $\Theta = (\theta_{ik})_{N \times K}$, and their estimates accordingly as $\hat{\Theta}$. $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{Nj})^T$ is the vector of responses to the j th item for all examinees, and therefore $-l(b_{j1}, \dots, b_{j(M-1)}, a_{j1}, \dots, a_{jK}, \mathbf{Y}_j, \hat{\Theta})$ represents the log-likelihood of \mathbf{Y}_j and $\hat{\Theta}$ for the MGRM. The L_1 -norm penalty of \mathbf{a}_j for the j th item is expressed as $\sum_{k=1}^K |a_{jk}|$. Generally, larger tuning parameter tends to cause more sparse discrimination parameters, so λ_w directly influences the shrinkage effect of $\hat{\mathbf{a}}_j$. The convexity of the L_1 -norm penalized negative log-likelihood function in Equation (5) can be easily proven by convex function properties (e.g., Boyd and Vandenberghe, 2004; Nocedal and Wright, 2006; Hastie et al., 2015), so the coordinate descent algorithm (Friedman et al., 2007, 2010) can be used to solve the optimization. Specifically, set the relatively small value of λ for which the entire vector \mathbf{a}_j is equal to zeros. The estimated abilities, $\hat{\Theta}$, and observed response data, \mathbf{Y}_j , are inputted; the solution with respect to item parameters, $(b_{j1}, \dots, b_{j(M-1)}, a_{j1}, \dots, a_{jK})$, to the LASSO optimization for the MGRM based on the setting of λ is updated coordinate by coordinate, the details for which are illustrated in the **Appendix** section. Thus, the solutions to the optimizations with a decreasing sequence of λ can be obtained respectively. That is, for W values of the tuning parameter, the discrimination parameter estimates, $\hat{\mathbf{a}}_j^{(1)}, \dots, \hat{\mathbf{a}}_j^{(W)}$, can be obtained *via* the algorithm and the alternative item-trait patterns of the j th item is accordingly obtained. Denote the true item-trait pattern for j as $\mathbf{Q}_j = (q_{j1}, \dots, q_{jK})^T$, and $q_{jk} = 1$ represents the k th ability is measured by that item, while $q_{jk} = 0$ represents the opposite

case. Denote $\hat{\mathbf{Q}}_j^{(w)} = (\hat{q}_{j1}^{(w)}, \dots, \hat{q}_{jK}^{(w)})$ as the estimated item-trait pattern of Equation (5). If $\hat{a}_{jk} > 0$, then $q_{jk} = 1$; else $q_{jk} = 0$. The optimal item-trait pattern is denoted by $\mathbf{Q}_j^* = (q_{j1}^*, \dots, q_{jK}^*)$. For this research, because necessary ability estimates can be obtained based on the operational items in MCAT, so the patterns of replenished items for the MGRM can be directly estimated (identified) *via* optimizations of the LPRM-GR without factor rotation.

Step 3. Apply an appropriate selection index to choose the optimal item-trait patterns for the replenished items from the alternative patterns. As discussed in Sun and Ye (2019), the goodness-of-fit for the M2PLM to the data can be appropriately measured by the regular BIC and it is implemented for finding the optimal patterns from the alternatives, for which the decision rule is minimizing the alternatives in terms of BIC. In this research, the regular BIC for the replenished items with a given λ can be defined as

$$BIC = -2 \sum_{j=1}^{J_1} \sum_{i=1}^N \log(\hat{l}_{ij}) + n_p \log(N), \quad (6)$$

where n_p represents the number of item parameters obtained by the result of the LASSO optimization such as that in Equation (5); N represents the number of examinees and \hat{l}_{ij} calculates the log-likelihood based on the i th examinee's ability estimate. The decision rule is to minimize the regular BIC based on a combination of alternative patterns with different λ values.

Note that for the LRPM-GR, it could also be tried to adjust the regular BIC as the weighted BIC in order to furtherly improve the precision of pattern selection from the ones in the above steps. There have been a few studies highlighting the effects of response distribution in parameter estimation and pattern selection (e.g., King and Zeng, 2001; Hu and Zidek, 2002, 2011). Hu and Zidek (2002) addressed the challenges of imbalanced classes of the population problem in maximum likelihood estimation. They show that the unbalanced distribution of responses did influence the parameter estimation accuracy and pointed out that appropriate weights enable researchers to represent the degree to which the information from the populations should be used in fitting the likelihood. Furthermore, Hu and Zidek (2002) believe that the best choice of the weights will depend on the context, and the weighted likelihood with data-dependent weights can be referred to as adaptive weighted likelihood. Thus, the weighted BIC for a given λ is here constructed in manners of adaptive weighted likelihood to assist in improving the effects of the response distribution (or the observed information of the examinee populations) in the pattern selection process:

$$BIC = -2 \sum_{j=1}^{J_1} \sum_{i=1}^N v_{ij} \log(\hat{l}_{ij}) + n_p \log(N), \quad (7)$$

where the weights $\{v_{ij}\}$ are based on the score of the examinee as well as the corresponding frequency with which examinees scored at the same response level: $v_{ij} = \eta_{j, Y_{ij}}$. That is, if $Y_{ij} = m$ ($m = 0, \dots, M$), then $v_{ij} = \eta_{jm}$, where $\eta_{jm} = \sum_{i=1}^N \frac{I(Y_{ij}=m)}{N}$.

So the weighted BIC modifies the log-likelihood by weighting the likelihood based on the above frequencies. The characteristic of the specific form of the weighted BIC here is that the weights are easier to obtain and less computationally intensive. Similar to Hu and Zidek (2002)'s point for the weighted BIC, the practical meaning of the weights in the weighted BIC here can be understood as enabling the contribution of observed responses to the likelihood of the criterion to be enhanced; that is, the likelihood could be emphasized by means of giving the degree of the information from the examinee populations. Based on the above considerations, the regular BIC and weighted BIC were, respectively, applied in the simulation section to choose the optimal item-trait patterns from the alternatives for all the replenished items.

The shrinkage effect of the LASSO leads to controllable bias to the item parameter estimation. The estimated item parameters obtained from the L_1 -regularized optimizations should not be directly used to calculate the likelihood in the regular BIC or weighted BIC. Thus, the calibration of item parameters for all the replenished items should be re-estimated based on the alternative patterns from the optimizations with different λ values, which is similar to the treatment for the simplified relaxed LASSO (Meinshausen, 2007; Hastie et al., 2017). The response data can be refitted by the MGRM based on each of the alternative patterns *via* the algorithm of path-wise coordinate descent, the corresponding optimization for which is simply the estimation based on likelihood without involving the L_1 -norm penalty. The decision rule is to minimize the regular BIC or weighted BIC based on alternative patterns with different λ values.

SIMULATION

Studies 1, 2, 3 corresponding to the two-, three-, and four-dimensional MCAT item pools were conducted to evaluate the performance of the LPRM-GR for detecting the optimal item-trait patterns. In each study, there were 16 conditions considered, which were across different latent trait correlation (i.e., Σ), response categories for the graded-response data (i.e., $M+1$), test lengths, and item selection criteria. Specifically, independent and correlated latent traits were designed, respectively, for generating response data based on the MGRM. Two types of response categories were considered for the MGRM: $M+1 = 3$ and 4. The fixed-length and variable-length stopping rules were designed, respectively, for the MCAT. Two types of item selection criteria were considered for the MCAT procedure of the LPRM-GR: Bayesian D-optimality and Bayesian A-optimality (e.g., Berger and Veerkamp, 1994; van der Linden, 1999; Mulder and van der Linden, 2009). Besides, the regular BIC and weighted BIC were applied by the LPRM-GR, respectively, to choose the optimal item-trait patterns from the alternatives for replenished items. For each condition in the simulation study, 50 replications were conducted. The computational codes were written in the R software.

Item Pool and Data Design

Item Pool Generation

For each of the three studies, discrimination parameters ($\{a_{jk}\}$) were randomly sampled from the uniform distribution,

$U(0.7, 1.5)$; boundary parameters ($\{d_{jm}\}$) were drawn from $U[-2, 2]$ and each was uniformly distributed along with an equidistant interval within this range for each item (e.g., Jiang et al., 2016). Specifically, boundary parameters for $M+1 = 3$ were randomly sampled from $U[-2, -0.67]$, $U[-0.67, 0.67]$, and $U[0.67, 2]$, respectively. Boundary parameters for $M+1 = 4$ were randomly sampled from $U[-2, -1]$, $U[-1, 0]$, $U[0, 1]$, and $U[1, 2]$, respectively.

In Study 1, each two-dimensional item pool had $J_0 = 900$ operational items, of which 300 items measured the first trait, another 300 items measured the second trait, and the remaining 300 items measured both the two traits, so 3 types of patterns were designed for the study. Each item pool of Study 1 had $J_1 = 30$ replenished items, of which 10 items measured the first trait, another 10 items measured the second trait, and the rest measured both the two traits. Similarly, since three latent traits were measured by each item pool of Study 2, there were seven types of patterns designed. Each item pool of the study consisted of $J_0 = 910$ operational items and $J_1 = 35$ replenished items, of which every 130 operational items and every five replenished items corresponded to one of the seven types of patterns. In Study 3, each four-dimensional item pool had $J_0 = 900$ operational items and $J_1 = 30$ replenished items, of which every 60 operational items and every two replenished items had one of the 15 types of patterns.

Response Data Generation

For all of the three studies, the true abilities of $N = 4,000$ examinees are randomly sampled from two multivariate normal distributions with $N(0, \Sigma_i)$, ($i = 1, 2$), respectively. For the two-dimensional study, covariance matrices were

$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$; for the three-dimensional

study, those were $\Sigma_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} 1 & 0.2 & 0.3 \\ 0.2 & 1 & 0.5 \\ 0.3 & 0.5 & 1 \end{pmatrix}$; for

the four-dimensional study, those were $\Sigma_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and

$\Sigma_2 = \begin{pmatrix} 1 & 0.2 & 0.4 & 0.5 \\ 0.2 & 1 & 0.3 & 0.5 \\ 0.4 & 0.3 & 1 & 0.7 \\ 0.5 & 0.5 & 0.7 & 1 \end{pmatrix}$. The responses of the examinees were

generated from the MGRM based on the true values of the parameters given above.

Multidimensional Computerized Adaptive Testing Procedure in LPRM-GR

The MCAT procedure with the MGRM required by the LPRM-GR was simulated by the R package: mirtCAT (Chalmers, 2016). The estimation of examinees' abilities based on the MGRM was obtained by the maximum a posterior (MAP), which is generally proven by numerous previous research to provide more accurate results than the method of expected a posterior. As introduced

by Steps 1 and 2 of the LPRM-GR, the ability estimates from Step 1 were directly exploited for the subsequent L_1 -regularized optimization in Step 2.

Item Selection Criterion

Item selection is one of the key components of CAT, and frequently used item selection approaches are constructed based on Fisher Information or Kullback-Leibler Information (e.g., Wang and Chang, 2011; Wang et al., 2011; Wang et al., 2013; Tu et al., 2018). For MCAT, the item selection strategy based on Fisher information is frequently used. In this paper, the operational items were selected from the item pool by Bayesian D-optimality and Bayesian A-optimality, respectively. Bayesian A-optimality selects the j th item from the remaining items in the item pool by minimizing the traces of the inverse to the sum of the Fisher information matrices and a prior covariance matrix, which is formulated as

$$\arg \min \{ \text{trace} [I_{S_{j-1}}(\hat{\theta}^{j-1}) + I_j(\hat{\theta}^{j-1}) + \Phi^{-1}]^{-1} \}, \quad (8)$$

where $I_{S_{j-1}}(\hat{\theta}^{j-1})$ and $I_j(\hat{\theta}^{j-1})$ represent the information matrices based on $\hat{\theta}^{j-1}$ for the $j-1$ administered items and candidate item i_j , respectively. The prior covariance matrix for abilities is denoted as Φ . Bayesian D-optimality selects the j th item from the remaining items in the item pool by maximizing the determinants of the sum of the Fisher information matrices and a prior covariance matrix for abilities:

$$\arg \max \{ \det [I_{S_{j-1}}(\hat{\theta}^{j-1}) + I_j(\hat{\theta}^{j-1}) + \Phi^{-1}] \}. \quad (9)$$

Stopping Rules

The goal of a fixed-length stopping rule is to terminate the CAT with a predetermined length. However, it is shown that the fixed-length rule may produce less accurate results for the examinees with quite different abilities from the average difficulty level of the item pool (e.g., Wang et al., 2013). Compared with the fixed-length rule, the variable-length rule (e.g., Boyd et al., 2010) such as the standard error (SE) rule (e.g., Weiss and Kingsbury, 1984; Yao, 2013) can provide approximately equal precision for all examinees regardless of their different abilities. The SE rule simply controls that the standard error of the precision of each examinee's ability estimate does not exceed a predetermined threshold. The SE rule is simple to implement as well as previous studies have also shown its good performance (e.g., Yao, 2013). Thus, the above two types of stopping rules were used in this study, respectively. The first was the fixed-length rule: the MCAT in the simulation was stopped at $Z_0 = 50$ operational items for each examinee. The second was the SE rule with a maximum 100-length constraint, which aims to jointly consider approximate similar measurement precision as well as appropriate test length for the MCAT because using the SE rule alone could produce a too long test for a few examinees with quite low or high abilities (e.g., Wang et al., 2018b); specifically, the MCAT is terminated once a pre-specified estimated standard error, 0.3, along with 100-length was reached.

Assignment of Replenished Items

All the replenished items were divided into five groups and assigned furtherly to each group. $Z_1 = 6$ replenished items were assigned to each group for the two-dimensional study while $Z_1 = 7$ for the three-dimensional study. Note that assigned replenished items contained all possible patterns. For the four-dimensional study, each group was assigned $Z_1 = 6$ replenished items, which corresponded to six different patterns, respectively. The examinees were equally divided into five groups either and answered assigned replenished items. Thus, $n = 800$ examinees were assigned to answer different replenished items.

Specification of Tuning Parameter

As is known, the sparsity of patterns for the LASSO can be affected by λ , so its alternative values should be given in a relatively large range to ensure the adequacy of the alternative item-trait patterns; whereas too many alternative values may produce similar patterns instead of diverse ones and cause unnecessary computing load. Therefore, dozens of equidistant cut-off points of λ are set between 0 and 120 for Step 2 of the proposed method after careful trials in simulation.

Evaluation Criteria

Correct Specification Rate

To evaluate the precision effects of the LPRM-GR on pattern recognition, one intuitive index is the correct specification rate of replenished items:

$$\text{CSR} = \frac{1}{J_1 \times K} \sum_{j=1}^{J_1} \sum_{k=1}^K I(q_{jk}^* = q_{jk}), \quad (10)$$

where $\mathbf{Q}_j = (q_{j1}, \dots, q_{jK})^T$ is the true pattern of the j th replenished item, and $\hat{\mathbf{Q}}_j = (q_{j1}^*, \dots, q_{jK}^*)^T$ is the identified pattern by the LPRM-GR. As above, each element of \mathbf{Q}_j takes 1 or 0, indicating the corresponding trait is or is not measured by the j th item. The identity function is denoted by I . Note that CSR falls in $[0, 1]$, and the larger value of CSR means the higher recognition accuracy.

Ability Estimation Accuracy

The absolute mean error (AME) and root mean square error (RMSE) are calculated as follows to evaluate the ability estimation accuracy of the examinees in the MCAT procedure:

$$\text{AME}(\theta) = \frac{\sum_{i=1}^N \sum_{k=1}^K |\hat{\theta}_{ik} - \theta_{ik}|}{N \times K}, \quad (11)$$

$$\text{RMSE}(\theta) = \sqrt{\frac{\sum_{i=1}^N \sum_{k=1}^K (\hat{\theta}_{ik} - \theta_{ik})^2}{N \times K}}, \quad (12)$$

where $\hat{\theta}_{ik}$ is the estimated ability on the k th dimension of the i th examinee, and θ_{ik} is the true ability for that. Smaller values of the two indices correspond to the higher estimation accuracy of abilities.

Item Parameter Estimation Accuracy

Similar to the evaluation indices in ability vectors, the AME is applied to measure the estimation of item parameters of replenished item parameters, which is calculated based on the patterns identified by the LPRM-GR with the estimated abilities and the true abilities, individually. The two indices of item parameters of replenished items with MGRM are computed as

$$AME(a) = \frac{\sum_{j=1}^{J_1} \sum_{k=1}^K |\hat{a}_{jk} - a_{jk}|}{J_1 \times K}, \tag{13}$$

$$AME(b) = \frac{\sum_{j=1}^{J_1} \sum_{m=1}^{M-1} |\hat{b}_{jm} - b_{jm}|}{J_1 \times (M - 1)}. \tag{14}$$

Smaller values of the two indices correspond to the higher estimation accuracy of item parameters.

Result of Study 1

Pattern Recognition Accuracy of Replenished Items in Study 1

Table 1 lists the CSR values of the optimal patterns of replenished items under multiple conditions based on two types of BIC in the two-dimensional item pool. Note that the table contains the CSRs for the LPRM-GR with the estimated abilities and true abilities, and those with true abilities can be used as benchmarks.

As shown in Table 1, the CSR values with estimated abilities were generally close or even equal to their benchmarks. The LPRM-GR with estimated abilities achieved quite high CSR results, which reached even 100% for the conditions with the weighted BIC. Thus, focusing on the two types of BIC, it showed that the CSRs obtained from the LPRM-GR with the weighted BIC had better performance than those with the regular BIC.

Furthermore, Table 1 showed that for the two types of response categories, the CSR values with independent-ability conditions were equal to or slightly higher than those for

correlated-ability conditions. It also indicated that Bayesian A-Optimality and Bayesian D-Optimality performed closely and had quite high CSR values. Comparing the pattern recognition accuracy of the LPRM-GR between the two types of test lengths, it was found that the CSRs are close for most conditions; the CSRs for the variable-length conditions especially with correlated abilities and selected with the regular BIC were slightly better than those for the fixed-length test conditions.

Ability Parameter Estimation Comparison in Multidimensional Computerized Adaptive Testing of Study 1

To investigate the effects of ability estimation in pattern recognition, the parameter estimation accuracy was given in Table 2. As expected, the AME and RMSE results of ability estimation for the variable-length test showed a significant improvement in estimation accuracy over the fixed-length test. It also shows that AME and RMSE values for Bayesian A-Optimality were generally lower than those for Bayesian D-Optimality. Although the AME and RMSE values of correlated abilities were slightly lower than those for independent abilities, the CSR values presented in Table 1 indicated that the pattern recognition accuracy for the correlated abilities was slightly lower than independent ones for the regular BIC.

Item Parameter Estimation Accuracy of Replenished Items in Study 1

To furtherly measure the impact of detected patterns on the recovery of replenished items and for brevity, AME values of estimated item parameters under different conditions with two types of BIC are presented in Tables 3, 4. The two tables showed that the AME values of estimated discrimination parameters based on the patterns selected by the weighted BIC were slightly smaller than those by the regular BIC. Also, by the comparison of Tables 1, 3, 4, it was found that the recovery of intercept parameters was not affected by the

TABLE 1 | Correct specification rate of item-trait patterns identified by the LPRM-GR in Study 1.

M+1	Condition Latent trait correlation Test length type		LPRM-GR with ability estimates				LPRM-GR with true abilities	
			Regular BIC		Weighted BIC		Regular BIC (%)	Weighted BIC (%)
			Item selection criterion					
			Bayesian A-optimality (%)	Bayesian D-optimality (%)	Bayesian A-optimality (%)	Bayesian D-optimality (%)		
3	Independent	Fixed-length	99.63	99.53	100	100	99.67	100
		Variable-length	99.63	99.63	100	100		
	Correlated	Fixed-length	98.33	98.40	100	100	99.60	100
		Variable-length	99.13	99.07	100	100		
4	Independent	Fixed-length	99.73	99.73	100	100	99.60	100
		Variable-length	99.63	99.73	100	100		
	Correlated	Fixed-length	98.70	98.73	100	100	99.70	100
		Variable-length	99.10	99.10	100	100		

TABLE 2 | Recovery of ability parameters in Study 1.

<i>M</i> +1	Condition		Item selection criterion			
			Bayesian A-optimality		Bayesian D-optimality	
	Latent trait correlation	Test length type	AME (θ)	RMSE (θ)	AME (θ)	RMSE (θ)
3	Independent	Fixed-length	0.3081	0.3865	0.3215	0.4036
		Variable-length	0.2464	0.3091	0.2570	0.3226
	Correlated	Fixed-length	0.2981	0.3741	0.3094	0.3887
		Variable-length	0.2414	0.3029	0.2512	0.3152
4	Independent	Fixed-length	0.2978	0.3733	0.3066	0.3848
		Variable-length	0.2398	0.3006	0.2440	0.3061
	Correlated	Fixed-length	0.2875	0.3607	0.2952	0.3703
		Variable-length	0.2345	0.2940	0.2384	0.2991

TABLE 3 | Recovery of item parameters based on patterns from the LPRM-GR with ability estimates and regular BIC in Study 1.

<i>M</i> +1	Condition		LPRM-GR with ability estimates			
			Item selection criterion			
	Latent trait correlation	Test length type	Bayesian A-optimality		Bayesian D-optimality	
			AME (<i>a</i>)	AME (<i>b</i>)	AME (<i>a</i>)	AME (<i>b</i>)
3	Independent	Fixed-length	0.0510	0.0867	0.0516	0.0874
		Variable-length	0.0487	0.0830	0.0488	0.0835
	Correlated	Fixed-length	0.0599	0.0869	0.0609	0.0877
		Variable-length	0.0525	0.0835	0.0532	0.0838
4	Independent	Fixed-length	0.0477	0.0867	0.0486	0.0871
		Variable-length	0.0460	0.0829	0.0461	0.0833
	Correlated	Fixed-length	0.0556	0.0855	0.0556	0.0855
		Variable-length	0.0501	0.0820	0.0503	0.0823

pattern recognition accuracy as sensitively as the discrimination parameters.

Result of Study 2

Pattern Recognition Accuracy of Replenished Items in Study 2

In the three-dimensional study, the CSR values for the LPRM-GR under multiple conditions are listed in **Table 5**. It indicated that the results in CSR were very close to their benchmarks, except for those in the correlated-ability condition for the regular BIC. Similar to the performance in pattern recognition accuracy in Study 1, the LPRM-GR in Study 2 generally produced relatively high CSRs for the overall conditions with the two types of BIC, and it especially achieved quite high CSR for the conditions with the weighted BIC.

In addition, **Table 5** indicated that the CSRs for the conditions with two types of response categories and the independent abilities were higher than those with the correlated abilities. Results also showed that for most cases, the tests with variable-length by either Bayesian A-Optimality or Bayesian D-Optimality for the LPRM-GR had relatively higher CSR values than tests with the length of 50 items. The two

types of item selection criteria showed a slight difference in CSR.

Ability Parameter Estimation Comparison in Multidimensional Computerized Adaptive Testing of Study 2

Table 6 listed the AME and RMSE of the estimated abilities. It showed that the ability parameter estimation accuracy for Bayesian A-Optimality was slightly better than that for Bayesian D-Optimality. Also, it was clear that the AME and RMSE for the correlated-ability conditions were lower than those for the independent-ability conditions. **Table 6** showed that the abilities could be estimated more precisely for the variable-length MCAT than that for the fixed-length case.

Item Parameter Estimation Accuracy of Replenished Items in Study 2

The AME values of the estimated item parameters based on the patterns selected by two types of BIC, individually, were listed in **Tables 7, 8**. It can be inferred from the two tables and **Table 5** that the estimation accuracy of intercept parameters was slightly influenced by the pattern recognition accuracy of replenished

TABLE 4 | Recovery of item parameters based on patterns from the LPRM-GR with ability estimates and weighted BIC in Study 1.

Condition			LPRM-GR with ability estimates			
<i>M</i> +1	Latent trait correlation	Test length type	Item selection criterion			
			Bayesian A-optimality		Bayesian D-optimality	
			AME (a)	AME (b)	AME (a)	AME (b)
3	Independent	Fixed-length	0.0502	0.0867	0.0506	0.0875
		Variable-length	0.0479	0.0830	0.0480	0.0836
	Correlated	Fixed-length	0.0550	0.0870	0.0559	0.0878
		Variable-length	0.0502	0.0835	0.0507	0.0837
4	Independent	Fixed-length	0.0472	0.0867	0.0481	0.0872
		Variable-length	0.0453	0.0829	0.0455	0.0833
	Correlated	Fixed-length	0.0520	0.0853	0.0530	0.0855
		Variable-length	0.0479	0.0820	0.0481	0.0823

TABLE 5 | Correct specification rate of item-trait patterns identified by the LPRM-GR in Study 2.

Condition			LPRM-GR with ability estimates				LPRM-GR with true abilities	
<i>M</i> +1	Latent trait correlation	Test length type	Regular BIC		Weighted BIC		Regular BIC (%)	Weighted BIC (%)
			Item Selection Criterion					
			Bayesian A-optimality (%)	Bayesian D-optimality (%)	Bayesian A-optimality (%)	Bayesian D-optimality (%)		
3	Independent	Fixed-length	99.62	99.64	99.94	99.94	99.67	100
		Variable-length	99.41	99.50	100	100		
	Correlated	Fixed-length	94.34	94.34	99.62	99.52	99.60	100
		Variable-length	96.61	96.59	99.85	99.89		
4	Independent	Fixed-length	99.33	99.28	100	100	99.47	100
		Variable-length	99.43	99.37	100	100		
	Correlated	Fixed-length	94.17	94.11	99.28	99.49	99.50	100
		Variable-length	96.63	96.50	99.83	99.81		

TABLE 6 | Recovery of ability parameters in Study 2.

Condition			Item selection criterion			
<i>M</i> +1	Latent trait correlation	Test length type	Bayesian A-optimality		Bayesian D-optimality	
			AME (θ)	RMSE (θ)	AME (θ)	RMSE (θ)
3	Independent	Fixed-length	0.4081	0.5122	0.4163	0.5223
		Variable-length	0.3334	0.4182	0.3385	0.4251
	Correlated	Fixed-length	0.3838	0.4815	0.3897	0.4888
		Variable-length	0.3204	0.4019	0.3250	0.4078
4	Independent	Fixed-length	0.3956	0.5832	0.4012	0.5034
		Variable-length	0.3195	0.4005	0.3243	0.4066
	Correlated	Fixed-length	0.3729	0.4681	0.3782	0.4748
		Variable-length	0.3088	0.3874	0.3126	0.3923

items, while the recovery of discrimination parameters was more sensitive to the pattern recognition accuracy. Besides, similar to the results of **Tables 3, 4**, the AMEs of the estimated

discrimination parameters based on the patterns selected by the LPRM-GR with weighted BIC were smaller than those with the regular BIC.

TABLE 7 | Recovery of item parameters based on patterns from the LPRM-GR with ability estimates and regular BIC in Study 2.

Condition			LPRM-GR with ability estimates			
M+1	Latent trait correlation	Test length type	Item selection criterion			
			Bayesian A-optimality		Bayesian D-optimality	
			AME (a)	AME (b)	AME (a)	AME (b)
3	Independent	Fixed-length	0.0532	0.1286	0.0536	0.1299
		Variable-length	0.0470	0.1121	0.0470	0.1129
	Correlated	Fixed-length	0.0834	0.1208	0.0856	0.1223
		Variable-length	0.0654	0.1089	0.0654	0.1096
4	Independent	Fixed-length	0.0515	0.1245	0.0518	0.1254
		Variable-length	0.0447	0.1086	0.0450	0.1094
	Correlated	Fixed-length	0.0807	0.1163	0.0813	0.1169
		Variable-length	0.0613	0.1054	0.0623	0.1055

TABLE 8 | Recovery of item parameters based on patterns from the LPRM-GR with ability estimates and weighted BIC in Study 2.

Condition			LPRM-GR with ability estimates			
M+1	Latent trait correlation	Test length type	Item selection criterion			
			Bayesian A-optimality		Bayesian D-optimality	
			AME (a)	AME (b)	AME (a)	AME (b)
3	Independent	Fixed-length	0.0527	0.1287	0.0531	0.1300
		Variable-length	0.0458	0.1122	0.0460	0.1129
	Correlated	Fixed-length	0.0635	0.1220	0.0652	0.1235
		Variable-length	0.0536	0.1094	0.0538	0.1001
4	Independent	Fixed-length	0.0499	0.1246	0.0500	0.1255
		Variable-length	0.0435	0.1087	0.0436	0.1095
	Correlated	Fixed-length	0.0622	0.1174	0.0614	0.1180
		Variable-length	0.0506	0.1058	0.0509	0.1059

Result of Study 3

Pattern Recognition Accuracy of Replenished Items in Study 3

For the four-dimensional item pool, the pattern recognition performance in CSR for the LPRM-GR is listed in **Table 9**. It showed that the CSR values with estimated abilities were close to their benchmarks in most scenarios. Furthermore, when the weighted BIC was used to choose the optimal item-trait patterns, the accuracy of the LPRM-GR was significantly improved, which is particularly evident for the conditions with correlated abilities. In general, comparing the results from **Tables 1, 5, 9**, it can be observed that the weighted BIC assisted the LPRM-GR in producing higher recognition accuracy than the regular BIC, especially for the correlated-ability conditions.

Similar to the results in **Tables 1, 5, Table 9** also showed that the CSRs for the correlated-ability conditions were lower than those for the independent-ability conditions. In addition, the CSR values for the variable-length tests were generally higher than those for the fixed-length tests, which was especially noticeable for the correlated-ability conditions. In contrast to the results with similar conditions in **Tables 1, 5**, the CSRs of the LPRM-GR were not significantly sensitive to the increase of dimensionality.

Ability Parameter Estimation Comparison in Multidimensional Computerized Adaptive Testing of Study 3

The recovery of ability parameters in Study 3 is listed in **Table 10**. It was found that the ability estimation accuracy for Bayesian A-Optimality was slightly better than that for Bayesian D-Optimality; the AME and RMSE values for the estimated correlated abilities were better than those for the independent abilities. Also, the variable-length tests in this study produced more precise estimates with respect to abilities than the fixed-length tests.

Item Parameter Estimation Accuracy of Replenished Items in Study 3

Tables 11, 12 list the AME values of item parameters of replenished items in study 3, which were estimated based on the optimal item-trait patterns selected by the LPRM-GR. For the four-dimensional replenished items, it can be inferred from **Tables 9, 11, 12** that the estimation accuracy of intercept parameters was less sensitive to the pattern recognition accuracy than the discrimination parameters. Furthermore, the recovery of estimated discrimination parameters based on the patterns

TABLE 9 | Correct specification rate of item-trait patterns identified by the LPRM-GR in Study 3.

Condition			LPRM-GR with ability estimates				LPRM-GR with true abilities	
			Regular BIC		Weighted BIC		Regular BIC (%)	Weighted BIC (%)
M+1	Latent trait correlation	Test length type	Item selection criterion					
			Bayesian A-optimality (%)	Bayesian D-optimality (%)	Bayesian A-optimality (%)	Bayesian D-optimality (%)		
3	Independent	Fixed-length	99.32	99.48	99.83	99.68	99.35	99.98
		Variable-length	99.32	99.38	99.95	99.92		
	Correlated	Fixed-length	87.65	87.05	93.93	93.68	98.57	98.70
		Variable-length	90.73	90.35	95.38	95.20		
4	Independent	Fixed-length	99.20	99.12	99.93	99.90	99.42	100
		Variable-length	99.28	99.30	99.97	99.97		
	Correlated	Fixed-length	86.87	87.28	93.13	93.13	98.38	98.70
		Variable-length	90.32	89.93	94.58	94.57		

TABLE 10 | Recovery of ability parameters in Study 3.

Condition			Item selection criterion			
			Bayesian A-optimality		Bayesian D-optimality	
M+1	Latent trait correlation	Test length type	AME (θ)	RMSE (θ)	AME (θ)	RMSE (θ)
			3	Independent	Fixed-length	0.4650
Variable-length	0.3858	0.4839			0.3888	0.4876
Correlated	Fixed-length	0.4189		0.5258	0.4220	0.5298
	Variable-length	0.3602		0.4519	0.3627	0.4551
4	Independent	Fixed-length	0.4528	0.5679	0.4576	0.5738
		Variable-length	0.3719	0.4664	0.3749	0.4020
	Correlated	Fixed-length	0.4099	0.5144	0.4133	0.5188
		Variable-length	0.3494	0.4383	0.3515	0.4409

TABLE 11 | Recovery of item parameters based on patterns from the LPRM-GR with ability estimates and regular BIC in Study 3.

Condition			LPRM-GR with ability estimates			
			Item selection criterion			
M+1	Latent trait correlation	Test length type	Bayesian A-optimality		Bayesian D-optimality	
			AME (<i>a</i>)	AME (<i>b</i>)	AME (<i>a</i>)	AME (<i>b</i>)
3	Independent	Fixed-length	0.0630	0.1962	0.0634	0.1981
		Variable-length	0.0534	0.1604	0.0532	0.1615
	Correlated	Fixed-length	0.1401	0.1646	0.1428	0.1654
		Variable-length	0.1097	0.1438	0.1113	0.1442
4	Independent	Fixed-length	0.0594	0.1917	0.0602	0.1939
		Variable-length	0.0486	0.1542	0.0488	0.1550
	Correlated	Fixed-length	0.1373	0.1544	0.1371	0.1558
		Variable-length	0.1057	0.1338	0.1069	0.1339

identified by the LPRM-GR performs better with weighted BIC than that with the regular BIC, while the recovery of estimated intercept parameters performs similarly for the LPRM-GR with the two types of BIC.

Moreover, the comparison of the AMEs among the three simulation studies illustrated that the discrimination parameter estimation accuracy was obviously affected by the pattern recognition accuracy of the LPRM-GR, especially for the

TABLE 12 | Recovery of item parameters based on patterns from the LPRM-GR with ability estimates and weighted BIC in Study 3.

Condition			LPRM-GR with ability estimates			
M+1	Latent trait correlation	Test length type	Item selection criterion			
			Bayesian A-optimality		Bayesian D-optimality	
			AME (a)	AME (b)	AME (a)	AME (b)
3	Independent	Fixed-length	0.0620	0.1970	0.0637	0.1989
		Variable-length	0.0522	0.1606	0.0523	0.1618
	Correlated	Fixed-length	0.1128	0.1679	0.1139	0.1690
		Variable-length	0.0908	0.1453	0.0916	0.1458
4	Independent	Fixed-length	0.0577	0.1921	0.0585	0.1944
		Variable-length	0.0472	0.1544	0.0475	0.1553
	Correlated	Fixed-length	0.1120	0.1570	0.1131	0.1582
		Variable-length	0.0890	0.1351	0.0889	0.1353

conditions with correlated abilities. Results of the three studies also showed that the AMEs of item parameters based on the MCAT under the variable-length rule were lower than those for the fixed-length rule.

CONCLUSION AND DISCUSSION

To manually identify how the replenished items in MCAT based on MGRM measure the latent traits is not only time-consuming but also affecting the accuracy and quality of the MCAT. In contrast, to automatically detect the item-trait patterns for replenished items with graded responses in the MCAT item pool can facilitate the item calibration of the items and benefit for the security of the subsequent diagnostic information mining and feedback. At the basis of Sun and Ye (2019), but not limited to, this research proposes a data-driven pattern recognition method for graded-response items, LPRM-GR, to find the optimal item-trait patterns of replenished items fitted by the MGRM in MCAT, which extends the framework of the LPRM to the case of graded responses for the item pool replenishment. The LPRM-GR formulates the L_1 -regularized optimization and gives the coordinate descent algorithm to solve it, and investigate the performance of the regular BIC and the weighted BIC, respectively, for selecting the optimal patterns of replenished items from the alternatives. Three simulation studies were conducted to evaluate the LPRM-GR with the two types of BIC in multiple cases across different numbers of response categories, ability correlation, stopping rules, and item selection criteria.

Conclusion

Conclusions can be drawn from the results of three studies in simulation. In general, if one focuses on the CSR of item-trait patterns and the AME of estimated item parameters, it can be found: the LPRM-GR for variable-length scenarios performed better than that for fixed-length scenarios; the proposed method for independent cases performed better than that for correlated cases in terms of ability correlation; the proposed method

with the Bayesian A-optimality in MCAT performed as similar as that with the Bayesian D-optimality; the proposed method for the MGRM with three response categories and that with four response categories had similar results; the higher pattern recognition accuracy of the LPRM-GR yielded comparatively higher estimation accuracy of discrimination parameters in MGRM; the weighted BIC performed better than the regular BIC for assisting the LPRM-GR in finding the optimal patterns. In addition, the estimation accuracy of discrimination parameters in MGRM was higher than that of intercept parameters.

Details of the findings of the three studies are given as follows. Firstly, the CSRs of the selected optimal patterns for the LPRM-GR with ability estimates in the MCAT procedure reached 86.87–100% for the overall cases for the three numbers of dimensionality. Specifically, the CSRs of the LPRM-GR with the ability estimates under the overall conditions are above 86%; for the two- and three-dimensional item pools, the CSRs are above 98% and 94%. By comparison with the benchmarks, the CSRs indicate that it is feasible for the LPRM-GR to help the LASSO optimization procedure by using ability estimates from the MCAT. Since it is not possible to know examinees' true abilities in practice, with the appropriate sample size designed in simulation, the ability estimates ensure that the LPRM-GR can detect the optimal item-trait patterns of graded-response items in relatively satisfactory accuracy and efficiency.

Secondly, a worth-noting result is that the LPRM-GR with ability estimates and the weighted BIC obtains fairly high recognition accuracy, for which the CSRs reach over 93% for the overall cases and even being close to 100% for some cases. Regarding the overall performance of the LPRM-GR with the two types of BIC in pattern recognition accuracy and item parameter recovery, the studies demonstrate that the weighted BIC can help LPRM-GR improve the recognition accuracy better than the regular BIC, especially in the conditions for correlated abilities. The weighted BIC used in simulation has conveniently calculated weights, which provides a good choice for the practical application of the LPRM-GR.

Thirdly, taking an insight into the LPRM-GR with the weighted BIC, it can be found that: for the correlated ability

conditions, the MCAT with SE rule and the constraint of up to 100 items in the LPRM-GR produces significantly better ability estimates, but slightly better CSRs, than the fixed-length rule (i.e., 50-length). The variable-length rule used by the LPRM-GR can play the positive part in ability estimation accuracy, which consequently to some extent benefits pattern recognition accuracy and discrimination parameter recovery of the replenished items with graded responses. As a matter of fact, it can be found that for most simulation cases with the variable-length rule, the length of MCAT in LPRM-GR is larger than 50, which could primarily explain the yield of better ability estimates. Nevertheless, the three studies also show that for the independent ability conditions, the fixed-length rule used in the LPRM-GR can provide similar pattern recognition accuracy to the variable-length rule, although the ability estimation accuracy for the former is not as good as the latter. In practice, if test developers cannot ignore the possible low computing efficiency yielded by the variable-length rule, the fixed-length MCAT can be a good choice for the application of the LPRM-GR.

Fourthly, for the two types of response categories of MGRM, the results in three studies suggest that the LPRM-GR can both get relatively high CSRs under the designed conditions, so by simulation, the response-category number has a relatively slight influence on ability estimation accuracy, pattern recognition accuracy, and item parameter recovery.

Discussion

Advantages of the LPRM-GR are illustrated as follows. As one of the key parts for properly maintaining the functioning of the MCAT system, item-trait pattern recognition for item replenishment of the tests with polytomous responses is necessary. As the polytomous extension of the LPRM, the LPRM-GR can solve the recognition problem by detecting the optimal item-trait patterns of replenished items with graded responses efficiently and accurately, which appropriately utilizes the MCAT operation to integrate with the LASSO and BIC to promote the data-driven pattern recognition. Additionally, the LPRM-GR addresses the potential meaning of examinees' item scoring frequencies in developing the weighted BIC for the further improvement of the effect of selecting the optimal patterns based on the LASSO. The advantage of the weighted BIC utilized in the LPRM-GR is well-supported by the results of the studies, especially for the simulation cases of correlated abilities.

As pointed out by Sun and Ye (2019), although the online calibration methods (e.g., Chen and Wang, 2016; Chen et al., 2017) and the LPRM can utilize the online feature of the MCAT, they solve different problems. An online calibration method focuses on estimating item parameters of replenished items based on well-known patterns, which can be reached by either a theory-driven or a data-driven method. The LPRM and LPRM-GR take interest in selecting the optimal pattern for each replenished item from all the possible alternatives, i.e., considering statistical goodness-of-fit for items based on the penalized-likelihood optimization like the L_1 -regularization. Similar to the advantages of the LPRM, the MCAT online function used in Step 1 of the LPRM-GR can automatically

put the item parameters of the replenished items on the same scale as the operational ones in the item pool, which saves time and labor costs for the pretest of the replenished items. For a considerable long period of MCAT operation, the number of replenished items in the item pool can be relatively large, the data-driven methods may save the recognition efficiency to some extent. Moreover, the MCAT designed for the LPRM-GR does not require an extremely large sample size. Certainly, if the practical situation permits, an appropriate increase in the sample size for the proposed method may still expectedly yield higher pattern recognition accuracy. Too long tests for the variable-length rule can increase the fatigue effect as well as time costs, so for the LPRM-GR, which type of stopping rules should be put into practice needs full consideration of the tradeoff between pattern recognition accuracy and efficiency for estimating abilities.

Taking insight into the sample size of the LPRM-GR in simulation, it can be found that 4,000 examinees are designed for the graded-response scenario. The purpose of designing 800 examinees for each MCAT of the LPRM-GR is to fully consider the relatively more item parameters for the MGRM than that for the M2PLM, so that to investigate how well the LPRM-GR performs in pattern recognition accuracy in a relatively large sample size for the MGRM. Simulation results sufficiently indicate the above sample size can help the LPRM-GR getting good results in comparison with the benchmarks. Note that the sample size here is not too large to implement for the LPRM-GR, so based on the above considerations, the LPRM-GR is expected to be feasible and have good application value.

Implications and future directions are summarized as follows. First of all, this research takes observed responses as the role of the so-called adaptive weighting of the likelihood (e.g., Hu and Zidek, 2002) in BIC, and found that the weighted BIC exploited in this paper had even better performance than the regular BIC. Those weights for the weighted BIC are also easy to compute. Note that Warm (1989) propose a weighted likelihood estimator for ability estimation in the three-parameter logistic model. The idea for defining weights of the weighted likelihood is technically to measure the amount of information by the Fisher's information function with abilities. Similar ideas were subsequently generalized to compensatory MIRT models (e.g., Tseng and Hsu, 2001; Wang, 2015). The idea of measuring the item information by the statistical variant of Fisher information matrix can also be found in constructing multidimensional item selection indices (e.g., Mulder and van der Linden, 2009). Thus, if thinking about other forms of the weights in the likelihood for constructing the adaptive weighted BIC, reasonable variants such as the determinant, trace, or other math quantities composed of examinees' Fisher information may be worth experimenting and discussing in the future. Secondly, it is possible to take into account other shrinkage methods for selecting appropriate item-trait patterns of replenished items such as the elastic net (Zou and Hastie, 2005) and the adaptive LASSO (Zou, 2006) in conjunction with the proposed approach in future. Future research can also consider further pattern recognition study for more test situations or some extended models. For instance, the proposed method can also be extended for the items fitted

by other MIRT models, such as the bi-factor models (Gibbons and Hedeker, 1992; Gibbons et al., 2007; Seo and Weiss, 2015), the two-tier full-information item factor model (Cai, 2010), and hierarchical test situations.

In addition, a limitation of the LPRM-GR is that its idea for pattern recognition constructs on the compensatory MIRT model with graded responses, so how to extend that to suit for the non-compensatory MIRT models can be considered in the future. The current pattern recognition method can also be extended to design for more variable-length stopping rules (e.g., Wang et al., 2013; Wang et al., 2018b) or new item selection and exposure control methods (e.g., Chen et al., 2020) in MCAT. As reviewers' suggestion, researchers can furtherly explore the practical topics on the MCAT item pool replenishment in future: what indices should be comprehensively considered for finding the frequently used operational items; how long items should be used before item exposure becomes a concern; what specific factors that test developers should be aware for the MCAT development.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the corresponding author upon request, without undue reservation.

REFERENCES

- Bartholomew, D., Knott, M., and Moustaki, I. (2011). *Latent Variable Models and Factor Analysis*, 3rd Edn. Chichester: John Wiley.
- Béguin, A. A., and Glas, C. A. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika* 66, 541–561. doi: 10.1007/BF02296195
- Berger, M. P. F., and Veerkamp, W. J. J. (1994). *A Review of Selection Methods for Optimal Test Design*. Enschede: University of Twente.
- Boyd, A. M., Dodd, B. G., and Choi, S. W. (2010). "Polytomous models in computerized adaptive testing," in *Handbook of Polytomous Item Response Theory Models*, eds M. L. Nering and R. Ostini (New York, NY: Routledge), 229–255.
- Boyd, S. P., and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge: Cambridge university press.
- Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika* 75, 581–612. doi: 10.1007/s11336-010-9178-0
- Cai, L. (2013). *flexMIRT: A Numerical Engine for Flexible Multilevel Multidimensional Item Analysis and Test Scoring (Version 2.0) [Computer software]*. Chapel Hill, NC: Vector Psychometric Group.
- Chalmers, R. P. (2016). Generating adaptive and non-adaptive test interfaces for multidimensional item response theory applications. *J. Stat. Softw.* 71, 1–38. doi: 10.18637/jss.v071.i05
- Chen, C., Wang, W., Chiu, M. M., and Ro, S. (2020). Item selection and exposure control methods for computerized adaptive testing with multidimensional ranking items. *J. Educ. Meas.* 57, 343–369. doi: 10.1111/jedm.12252
- Chen, P., and Wang, C. (2016). A new online calibration method for multidimensional computerized adaptive testing. *Psychometrika* 81, 674–701. doi: 10.1007/s11336-015-9482-9
- Chen, P., Wang, C., Xin, T., and Chang, H. (2017). Developing new online calibration methods for multidimensional computerized adaptive testing. *Br. J. Math. Stat. Psychol.* 70, 81–117. doi: 10.1111/bmsp.12083

AUTHOR CONTRIBUTIONS

JS was in charge of the literature review, original idea and derivation of methodology, simulation design and computing codes, and drafting and revising of the manuscript. ZY contributed to the literature review, application of the weighted BIC, computing codes, program running, tabular data collection, and drafting of the manuscript. LR contributed to the literature review, computing codes, and drafting and revising of the manuscript. JL contributed to the literature review and revising of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Natural Science Foundation of China (No. 11701029) and National Social Science Fund of China (20CSH075).

ACKNOWLEDGMENTS

We would like to thank the editors and the reviewers for their insightful comments and valuable suggestions.

- Depaoli, S., Tiemensma, J., and Felt, J. M. (2018). Assessment of health surveys: fitting a multidimensional graded response model. *Psychol. Health Med.* 23, 13–31. doi: 10.1080/13548506.2018.1447136
- Embretson, S. E., and Reise, S. P. (2000). *Psychometric Methods: Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fan, J., and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* 96, 1348–1360. doi: 10.1198/016214501753382273
- Friedman, J., Hastie, T., Höfling, H., and Tibshirani, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.* 1, 302–332. doi: 10.1214/07-AOAS131
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi: 10.18637/jss.v033.i01
- Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., et al. (2007). Full-information item bifactor analysis of graded response data. *Appl. Psychol. Meas.* 31, 4–19. doi: 10.1177/0146621606289485
- Gibbons, R. D., and Hedeker, D. (1992). Full-information item bi-factor analysis. *Psychometrika* 57, 423–436. doi: 10.1007/BF02295430
- Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Portland, OR: Chapman and Hall.
- Hastie, T., Tibshirani, R., and Tibshirani, R. J. (2017). Extended comparisons of best subset selection, forward stepwise selection, and the lasso. *arXiv [Preprint]*. Available online at: <https://arxiv.org/abs/1707.08692> (accessed July 29, 2017).
- Hoerl, A. E., and Kennard, R. W. (2000). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 42, 80–86. doi: 10.2307/1271436
- Hu, F., and Zidek, J. V. (2002). The weighted likelihood. *Can. J. Stat.* 30, 347–371. doi: 10.2307/3316141
- Hu, F., and Zidek, J. V. (2011). "The relevance weighted likelihood with applications," in *Empirical Bayes and Likelihood Inference*, eds S. E. Ahmed and N. Reid (New York, NY: Springer), 211–235. doi: 10.1007/978-1-4613-0141-7_13
- Jiang, S., Wang, C., and Weiss, D. J. (2016). Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Front. Psychol.* 7:109. doi: 10.3389/fpsyg.2016.00109

- Kelderman, H., and Rijkes, C. P. M. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika* 59, 149–176. doi: 10.1007/BF02295181
- King, G., and Zeng, L. (2001). Logistic regression in rare events data. *Polit. Anal.* 9, 137–163. doi: 10.1093/oxfordjournals.pan.a004868
- McKinley, R. L., and Reckase, M. D. (1982). *The Use of the GENERAL RASCH MODEL with Multidimensional Item Response Data*. Iowa City, IA: American College Testing.
- Meinshausen, N. (2007). Relaxed lasso. *Comput. Stat. Data Anal.* 52, 374–393. doi: 10.1016/j.csda.2006.12.019
- Mulder, J., and van der Linden, W. J. (2009). Multidimensional adaptive testing with optimal design criteria for item selection. *Psychometrika* 74, 273–296. doi: 10.1007/s11336-008-9097-5
- Muraki, E., and Carlson, J. E. (1995). Full-information factor analysis for polytomous item responses. *Appl. Psychol. Meas.* 19, 73–90. doi: 10.1177/014662169501900109
- Nocedal, J., and Wright, S. (2006). *Numerical Optimization*, 2nd Edn. New York, NY: Springer.
- Nouri, F., Feizi, A., Roohafza, H., Sadeghi, M., and Sarrafzadegan, N. (2021). How different domains of quality of life are associated with latent dimensions of mental health measured by GHQ-12. *Health Qual. Life Outcomes* 19:255. doi: 10.1186/s12955-021-01892-9
- Reckase, M. (2009). *Multidimensional Item Response Theory*. New York, NY: Springer.
- Reckase, M. D., and McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Appl. Psychol. Meas.* 15, 361–373. doi: 10.1177/014662169101500407
- Seo, D. G., and Weiss, D. J. (2015). Best design for multidimensional computerized adaptive testing with the bifactor model. *Educ. Psychol. Meas.* 75, 1–25. doi: 10.1177/0013164415575147
- Sun, J., Chen, Y., Liu, J., Ying, Z., and Xin, T. (2016). Latent variable selection for multidimensional item response theory models via L_1 regularization. *Psychometrika* 81, 921–939. doi: 10.1007/s11336-016-9529-6
- Sun, J., and Ye, Z. (2019). A lasso-based method for detecting item-trait patterns of replenished items in multidimensional computerized adaptive testing. *Front. Psychol.* 10:1944. doi: 10.3389/fpsyg.2019.01944
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x
- Tseng, F., and Hsu, T. (2001). Multidimensional adaptive testing using the weighted likelihood estimation: a comparison of estimation methods. *Paper Presented at the Annual Meeting of National Council on Measurement in Education*, Seattle, WA, 1–63.
- Tu, D., Han, Y., Cai, Y., and Gao, X. (2018). Item selection methods in multidimensional computerized adaptive testing with polytomously scored items. *Appl. Psychol. Meas.* 42, 677–694. doi: 10.1177/0146621618762748
- van der Linden, W. J. (1999). Multidimensional adaptive testing with a minimum error-variance criterion. *J. Educ. Behav. Stat.* 24, 398–412. doi: 10.2307/1165370
- Wainer, H., and Mislevy, R. J. (1990). “Item response theory, item calibration, and proficiency estimation,” in *Computerized Adaptive Testing: A Primer*, ed. H. Wainer (Hillsdale, NJ: Erlbaum), 65–102.
- Wainer, H. (2000). *Computer Adaptive Testing: A Primer*. Hillsdale, NJ: Lawrence Erlbaum.
- Wang, C. (2015). On latent trait estimation in multidimensional compensatory item response models. *Psychometrika* 80, 428–449. doi: 10.1007/s11336-013-9399-0
- Wang, C., and Chang, H.-H. (2011). Item selection in multidimensional computerized adaptive testing—Gaining information different angles. *Psychometrika* 76, 363–384. doi: 10.1007/s11336-011-9215-7
- Wang, C., Chang, H.-H., and Boughton, K. A. (2011). Kullback-Leibler information and its applications in multi-dimensional adaptive testing. *Psychometrika* 76, 13–39. doi: 10.1007/s11336-010-9186-0
- Wang, C., Chang, H.-H., and Boughton, K. A. (2013). Deriving stopping rules for multidimensional computerized adaptive testing. *Appl. Psychol. Meas.* 37, 99–122. doi: 10.1177/0146621612463422
- Wang, C., Su, S., and Weiss, D. J. (2018a). Robustness of parameter estimation to assumptions of normality in the multidimensional graded response model. *Multivariate Behav. Res.* 53, 403–418. doi: 10.1080/00273171.2018.1455572
- Wang, C., Weiss, D. J., and Shang, Z. (2018b). Variable-length stopping rules for multidimensional computerized adaptive testing. *Psychometrika* 84, 749–771. doi: 10.1007/s11336-018-9644-7
- Wang, C., Weiss, D. J., and Su, S. (2019). Modeling response time and responses in multidimensional health measurement. *Front. Psychol.* 10:51. doi: 10.3389/fpsyg.2019.00051
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika* 54, 427–450. doi: 10.1007/BF02294627
- Weiss, D. J., and Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *J. Educ. Meas.* 21, 361–375. doi: 10.1111/j.1745-3984.1984.tb01040.x
- Yao, L. (2013). Comparing the performance of five multidimensional CAT selection procedures with different stopping rules. *Appl. Psychol. Meas.* 37, 3–23. doi: 10.1177/0146621612455687
- Yoo, J. E. (2018). TIMSS 2011 Student and teacher predictors for mathematics achievement explored and identified via elastic net. *Front. Psychol.* 9:317. doi: 10.3389/fpsyg.2018.00317
- Yoo, J. E., and Rho, M. (2020). Exploration of predictors for Korean teacher job satisfaction via a machine learning technique. *Group Mnet. Front. Psychol.* 11:441. doi: 10.3389/fpsyg.2020.00441
- Yoo, J. E., and Rho, M. (2021). Large-scale survey data analysis with penalized regression: a Monte Carlo simulation on missing categorical predictors. *Multivariate Behav. Res.* 1–29. doi: 10.1080/00273171.2021.1891856 [Epub ahead of print].
- Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* 101, 1418–1429. doi: 10.1198/016214506000000735
- Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sun, Ye, Ren and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Derivation of solving the L_1 -regularized optimization in Equation (5) via the coordinate descent algorithm and Fisher's scoring iteration is given as follows. For simplicity, new symbols of the i th examinee's responses to the j th item (i.e., Y_{ij} in Equation (2) are here redefined as $(y_{ij0}, \dots, y_{ijM})$, where $y_{ijm} \in \{0, 1\}$ and $\sum_{m=0}^M y_{ijm} = 1$. If $Y_{ij} = m$, then $y_{ijm} = 1$ and $y_{ij\tilde{m}} = 0$ ($\tilde{m} \neq m$). For simplicity, the symbol θ_i is omitted from $P_{ijm}(\theta_i)$ and $P_{ijm}^*(\theta_i)$, which are hereafter denoted as P_{ijm} and P_{ijm}^* . The negative log-likelihood of the observed data and the known abilities based on the MGRM is

$$-l(\mathbf{b}_j, \mathbf{a}_j, \mathbf{Y}_j, \Theta) = -\sum_{i=1}^N \log \left(\prod_{m=0}^M P_{ijm}^{y_{ijm}} \right) = -\sum_{i=1}^N \sum_{m=0}^M y_{ijm} \log(P_{ijm}). \quad (15)$$

Note that for the MGRM, the intercept parameters should be updated are $b_{j1}, \dots, b_{j(M-1)}$. By the definition of $\Delta P_{ijm}^* = P_{ijm}^*(1 - P_{ijm}^*)$ and $\Delta^2 P_{ijm}^* = P_{ijm}^*(1 - P_{ijm}^*)(1 - 2P_{ijm}^*)$, the first-order derivative of Equation (15) with respect to b_{jm} ($m = 1, \dots, M - 1$) is represented as $g_{jm} = -\frac{\partial l(\mathbf{b}_j, \mathbf{a}_j, \mathbf{Y}_j, \Theta)}{\partial b_{jm}}$; after the t th iteration for item parameters, we have

$$g_{jm}^{(t)} = \sum_{i=1}^N \left[\frac{y_{ij(m-1)}}{P_{ij(m-1)}^{(t)}} \Delta P_{ijm}^{*(t)} - \frac{y_{ijm}}{P_{ijm}^{(t)}} \Delta P_{ijm}^{*(t)} \right]. \quad (16)$$

For the Fisher's scoring update, the second-order derivative of Equation (15) with respect to b_{jm} (i.e., $h_{jm} = -\frac{\partial^2 l(\mathbf{b}_j, \mathbf{a}_j, \mathbf{Y}_j, \Theta)}{\partial b_{jm}^2}$) is substituted with its expectation, and then

$$E(h_{jm}^{(t)}) = -\sum_{i=1}^N \left[\left(\frac{1}{P_{ij(m-1)}^{(t)}} (\Delta P_{ijm}^{*(t)})^2 + \Delta^2 P_{ijm}^{*(t)} \right) + \left(\frac{1}{P_{ijm}^{(t)}} (\Delta P_{ijm}^{*(t)})^2 - \Delta^2 P_{ijm}^{*(t)} \right) \right]. \quad (17)$$

The Fisher's scoring update of b_{jm} for the $(t+1)$ th iteration has the form:

$$\hat{b}_{jm}^{(t+1)} = \hat{b}_{jm}^{(t)} - \frac{g_{jm}^{(t)}}{E(h_{jm}^{(t)})}. \quad (18)$$

The first-order derivative of Equation (15) with respect to a_{jk} ($k = 1, \dots, K$) is represented as $g_{jk} = -\frac{\partial l(\mathbf{b}_j, \mathbf{a}_j, \mathbf{Y}_j, \Theta)}{\partial a_{jk}}$, and then

$$g_{jk}^{(t)} = -\sum_{i=1}^N \sum_{m=0}^M \frac{y_{ijm}}{P_{ijm}^{(t)}} [\Delta P_{ijm}^{*(t)} - \Delta P_{ij(m+1)}^{*(t)}] \theta_{ik}. \quad (19)$$

The expected second-order derivative of Equation (15) of a_{jk} is expressed as

$$E(h_{jk}^{(t)}) = \sum_{i=1}^N \sum_{m=0}^M \left\{ \frac{1}{P_{ijm}^{(t)}} [\Delta P_{ijm}^{*(t)} - \Delta P_{ij(m+1)}^{*(t)}]^2 - [\Delta^2 P_{ijm}^{*(t)} - \Delta^2 P_{ij(m+1)}^{*(t)}] \right\} \theta_{ik}^2. \quad (20)$$

The coordinate-wise update of a_{jk} for the $(t+1)$ th iteration has the form:

$$\hat{a}_{jk}^{(t+1)} = S \left(\hat{a}_{jk}^{(t)} - \frac{g_{jk}^{(t)}}{E(h_{jk}^{(t)})}, \lambda \right), \quad (21)$$

where the soft-thresholding operator (e.g., Friedman et al., 2010) is

$$S(z, \gamma) = \begin{cases} z - \gamma & \text{if } z > 0 \text{ and } \gamma < |z| \\ z + \gamma & \text{if } z < 0 \text{ and } \gamma < |z| \\ 0 & \text{if } \gamma \geq |z| \end{cases}. \quad (22)$$