Statistics and Its Interface Volume 12 (2019) 537–547

A method of Q-matrix validation based on symmetrised Kullback-Leibler divergence for the DINA model

Jinxin Guo*, Xin Xu, Jian Tao, and Ningzhong Shi

Q-matrix validation is one of the most vital parts in cognitive diagnosis, as the misspecification of Q-matrix may seriously influence the model fit and lead to incorrect classifications of examinees. In this paper, we propose a symmetrised Kullback-Leibler divergence- (SKLD-) based method to validate misspecified Q-matrix with a combination of K-means clustering. Three simulation studies are conducted to evaluate the sensitivity and specificity of the proposed method compared with that based on log odds ratio (LOR) and item discrimination index (IDI). The results show that the SKLD-based method could efficiently identify and validate misspecified elements in Q-matrix, and at the same time retain those correct ones. What's more, two real data sets are employed to further illustrate the performance of SKLD-based method.

Keywords and phrases: Q-matrix, Misspecification, DINA model, SKLD, IDI, LOR, K-means clustering.

1. INTRODUCTION

The traditional test theories, such as Classical Test Theory (CTT) and Item Response Theory (IRT), usually aim to measure and evaluate the unidimensional trait of examinees, which are more applicable for selection examination, ignoring the examinees' cognitive structure, skills, strategies or knowledge. In contrast, cognitive diagnosis models (CDMs) are the combination of cognitive psychology and modern measurement theory which can overcome the shortcomings of traditional test theories, providing a fine-grained assessment of examinees' skills profiles. The purpose of cognitive diagnosis is to make accurate classifications of examinees based on the response data and thus offers proper and detailed suggestions. Accordingly, CDMs can provide detailed information about strengths and weaknesses of students, which can make it possible to teach them in accordance with their aptitude.

In this procedure, one of the key elements in CDMs is Q-matrix, which describes the relations between items and attributes. Specifically, Q-matrix is a loading matrix that indicates which attributes are required for each item in the test. In general, Q-matrix is assumed to be correct and usually defined by domain experts according to their experiences and knowledge. However, there are various opinions on the specification of Q-matrix for the same items. So it is difficult to specify the Q-matrix exactly. In most cases, the misspecification of Q-matrix can seriously influence the model fit and thereby lead to incorrect classifications of examinees.

Realizing the significance of correctly specifying Q-matrix, several methods have been proposed by scholars and researchers. One of the most general methods is δ-method proposed in [9], that maximizes the difference of probabilities of correct response between examinees who master all the required attributes and those who don't. The performance of δ-method was evaluated by a simulation study and two real data examples. The results of simulation study showed that δ-method is effective. Although the results performed well, the performance of the method is affected by the cut-off points in practice. Meanwhile it is a little difficult to choose a reasonable cut-off point. Moreover, there are some other methods of Q-matrix validation (e.g. [1, 4, 7, 12, 18]). But the robustness and applicability of those methods are still problems.

In this paper, we propose a SKLD-based method to validate misspecified Q-matrix. The proposed method is established on the rational that correct Q-matrix can distinguish latent groups with different ideal response patterns to the maximum extent. The rest of this paper is organized as follows. The proposed method and related calculation procedure are introduced in Section 2. In Section 3, three simulation studies are conducted to evaluate the performance of the proposed method. Two real experimental data sets are used to illustrate the proposed methods in Section 4. Finally, some issues that need to be resolved are addressed and further research directions are discussed in Section 5.

2. SKLD-BASED METHOD

In this section, we introduce SKLD-based method for validating Q-matrix. Assume that K is the number of attributes in the test. And there are I examinees, each of whom responds to J items. Let $\boldsymbol{\alpha_i} = (\alpha_{i1}, \cdots, \alpha_{ik}, \cdots, \alpha_{iK})$ denote the ith examinee's attribute pattern, where $\alpha_{ik} \in$

*Corresponding author.

$\{0,1\}$. And $\alpha_{ik}=1$ indicates that the ith examinee masters the kth attribute. There are 2^K possible attribute patterns for all the examinees.

Q is a $J\times K$ matrix with binary entries q_{jk}, called Q-matrix. The element q_{jk} is used to describe whether or not the kth attribute is measured by jth item. For each j and k, $q_{jk}=1$ means that the jth item requires the kth attribute and $q_{jk}=0$ otherwise. Actually, there are 2^K-1 alternative q-vectors for each item, which are less 1 than attribute patterns, because $(0,0,\cdots,0)$ can be an attribute pattern rather than an alternative q-vector. And there are two common kinds of misspecification for q-entries, over-specification, that q-entries of 0 are modified into 1, and underspecification, that q-entries of 1 are misspecified as 0.

2.1 DINA model

Deterministic, input, noisy "and" gate (DINA) model [15] is one of the most important and widely used CDMs for there are only two parameters in this model and it significantly fits real data well [8].

For DINA model, it usually assumes that an examinee can not answer an item correctly unless all the required attributes have been mastered.

Based on these assumptions, DINA model is one of the noncompensatory CDMs, which means that one attribute can't be compensated by any other attributes.

Let X_{ij} be the response of the ith examinee to the jth item. X_{ij} is a dichotomous variable, that is, an examinee gets the item either right or wrong. For DINA model, item response function takes the following form,

$$P(X_{ij}=1\,|\boldsymbol{\alpha_i})=(1-s_j)^{\eta_{ij}}{g_j}^{1-\eta_{ij}},$$

where η_{ij} is the ideal response of the ith examinee to the jth item,

$$\eta_{ij}=\prod_{k=1}^{K}\alpha_{ik}^{q_{jk}}.$$

η_{ij} takes the two values, 0 and 1. Because the product is defined over all attributes, $\eta_{ij}=1$ occurs only when all product terms are 1, which means that all required attributes for the jth item have been mastered by the ith examinee. Through above analysis, we can see that DINA model divides examinees into two mastery classes for each item: those who have mastered all required attributes (group $\eta_j=1$) and those who are lacking at least one required attribute (group $\eta_j=0$). What's more, $s_j=P(X_{ij}=0\mid\eta_{ij}=1)$ and $g_j=P(X_{ij}=1\mid\eta_{ij}=0)$ are the slipping and guessing parameters of the jth item, respectively. In details, s_j is a specific item parameter to illustrate the probability of incorrectly answering the jth item for examinees in group $\eta_j=1$, when they should have correctly answered the item, that is, slipping. And it is always assumed that this probability is only indexed by item and is independent of the individuals. In other words, for any examinee in group $\eta_j=1$, they have the same slipping probability. Similarly, g_j is used to specify the probability of answering correctly when in fact, examinees lack at least one required attribute and are supposed to answer that incorrectly, that is, guessing.

2.2 SKLD-based method

Kullback-Leibler divergence (KLD) is used to measure the difference between two probability distributions. Specially, for two discrete probability distributions F and G defined on the same probability space, the KLD between F and G is defined as

$$KLD(F,G)=\sum_n F(n)\log\frac{F(n)}{G(n)}. \tag{1}$$

In fact, the larger this KLD is, the more easily the two distributions could be distinguished.

Actually, KLD is not rare in cognitive diagnosis. Tatsouka and Ferguson used KLD as item selection index for computer-adaptive tests in CDMs [22]. What's more, KLD was discussed by Henson and Douglas for test construction in [13]. And it was also studied for application in cognitive diagnosis by Xu, Chang, and Douglas [25]. Moreover, KLD is also viewed as relative (Shannon) entropy in information systems and called Kullback-Leibler information (KLI). Compared with other informations, such as Fisher information, it can deal with both continuous and discrete distributions. Further more, it is a global information while Fisher information is local [13]. So KLD is more commonly used in cognitive diagnosis.

The purpose of cognitive diagnosis measurement is making diagnostic classifications for examinees based on response data and Q-matrix. Naturally, only when Q-matrix is correctly specified, the diagnostic classifications can be accurate. As previously mentioned, for each item, DINA model divides examinees into group $\eta_j=1$ and $\eta_j=0$. Basically, when Q-matrix is correctly specified, DINA model should maximize the difference between the two groups. That is, the KLD of conditional distributions of response given $\eta_{ij}=1$ and $\eta_{ij}=0$ should reach its highest when the Q-matrix is correct.

However, the KLD is asymmetric, which means that, $KLD(F,G)$ and $KLD(F,G)$ are not always equal. Considering the symmetry of group $\eta_j=1$ and $\eta_j=0$, SKLD is taken into account, which is the original divergence proposed by [17]. In fact, SKLD is often used for feature selection in classification problems. Formally, it has the following form for the two distributions F and G:

$$SKLD(F,G)=KLD(F,G)+KLD(G,F) \tag{2}$$

The SKLD satisfies the following two properties:

(1) $SKLD(F,G)\geq 0$, for $\forall\, F$ and G, and $SKLD(F,G)=0$ if and only if $F=G$ almost everywhere.

(2) $SKLD(F,G)=SKLD(G,F)$.

In the proposed method, SKLD is employed to describe the difference between group $\eta_j = 1$ and $\eta_j = 0$. Specifically, the conditional distributions of response given the two groups are considered. If the q-vector of item j is $\boldsymbol{q_j}$, SKLD between the two distributions can be written as:

(3)
$$SKLD_j(g_0, g_1|\boldsymbol{q_j}) = KLD_j(g_0, g_1|\boldsymbol{q_j}) + KLD_j(g_1, g_0|\boldsymbol{q_j}),$$

where g_0 and g_1 denote the group corresponding to $\eta_j = 1$ and $\eta_j = 0$, respectively. In Equation (3), $KLD_j(g_0, g_1|\boldsymbol{q_j})$ has the following form:

$$\begin{aligned} KLD_j(g_0, g_1|\boldsymbol{q_j}) = P(X_{ij} = 1|\eta_{ij} = 1) \log \frac{P(X_{ij} = 1|\eta_{ij} = 1)}{P(X_{ij} = 1|\eta_{ij} = 0)} \\ + P(X_{ij} = 0|\eta_{ij} = 1) \log \frac{P(X_{ij} = 0|\eta_{ij} = 1)}{P(X_{ij} = 0|\eta_{ij} = 0)}, \end{aligned}$$

Accordingly, Equation (3) can be simplified as

(4)
$$\begin{aligned} SKLD_j(g_0, g_1|\boldsymbol{q_j}) = [P(X_{ij} = 1|\eta_{ij} = 1) - P(X_{ij} = 1|\eta_{ij} = 0)] \\ \cdot \log \frac{P(X_{ij} = 1|\eta_{ij} = 1)P(X_{ij} = 0|\eta_{ij} = 0)}{P(X_{ij} = 1|\eta_{ij} = 0)P(X_{ij} = 0|\eta_{ij} = 1)}. \end{aligned}$$

Equation (4) leads to an interesting phenomena that SKLD is the product of IDI and LOR. In details,

$$P(X_{ij} = 1|\eta_{ij} = 1) - P(X_{ij} = 1|\eta_{ij} = 0)$$

happens to be IDI defined in δ-method. And

$$\log \frac{P(X_{ij} = 1|\eta_{ij} = 1)P(X_{ij} = 0|\eta_{ij} = 0)}{P(X_{ij} = 1|\eta_{ij} = 0)P(X_{ij} = 0|\eta_{ij} = 1)}$$

is the LOR of X_{ij} and η_{ij} exactly, which is used to quantitatively describe the association between X_{ij} and η_{ij}. The positive LOR indicates the positive correlation of that two variables. Moreover, the larger positive LOR is, the stronger the positive association is.

Actually, SKLD is obviously superior to the two single indexes, LOR and IDI. An example is given in Table 1 to illustrate this point. For $\boldsymbol{q_1}$ and $\boldsymbol{q_2}$, they have the same IDI, 0.7. If $\boldsymbol{q_1}$ and $\boldsymbol{q_2}$ both are the alternative q-vectors for the same item, the δ-method couldn't determine which one is better. However, if the SKLD is used, LOR will be taken into account to describe the divergence from another angle. The LORs of $\boldsymbol{q_1}$ and $\boldsymbol{q_2}$ are $\log 36$ and $\log 54$. Thus the latter is more likely to be chosen as the correct q-vector. And similarly, the LORs of two different alternative q-vectors may be equal, such as $\boldsymbol{q_3}$ and $\boldsymbol{q_4}$ in Table 1, with the same LOR $\log 21$. At this point, IDI needs to be considered. So $\boldsymbol{q_4}$ is more reasonable than $\boldsymbol{q_3}$. Accordingly, the SKLD is a perfect combination of IDI and LOR. From this perspective, using SKLD as the validation index seems more reasonable.

Table 1. Example

Alternative q-vectors	p_{11}	p_{10}	IDI	LOR	SKLD
$\boldsymbol{q_1}$	0.9	0.2	0.7	$\log 36$	$0.7 \log 36$
$\boldsymbol{q_2}$	0.95	0.25	0.7	$\log 54$	$0.7 \log 54$
$\boldsymbol{q_3}$	0.9	0.3	0.6	$\log 21$	$0.6 \log 21$
$\boldsymbol{q_4}$	0.875	0.25	0.625	$\log 21$	$0.625 \log 21$

Note. p_{11} and p_{10} refer to the probabilities of not slipping and guessing when $\boldsymbol{q_l}$ is the alternative q-vector for $l = 1, 2, 3, 4$.

Formally, the rationale of SKLD-based method is that the correct q-vector should give the largest SKLD between the response distributions of group $\eta_j = 1$ and $\eta_j = 0$. That is, the q-vector for jth item is correctly specified if

(5)
$$\begin{aligned} \boldsymbol{q_j} &= \underset{\{\boldsymbol{\alpha_l}, l=1,\cdots,2^K-1\}}{\arg\max} [SKLD_j(g_0, g_1|\boldsymbol{\alpha_l})] \\ &= \underset{\{\boldsymbol{\alpha_l}, l=1,\cdots,2^K-1\}}{\arg\max} \{[P(X_{ij} = 1|\eta_{il} = 1) - P(X_{ij} = 1|\eta_{il} = 0)] \\ &\quad \cdot \log \frac{P(X_{ij} = 1|\eta_{il} = 1)P(X_{ij} = 0|\eta_{il} = 0)}{P(X_{ij} = 1|\eta_{il} = 0)P(X_{ij} = 0|\eta_{il} = 1)}\}, \end{aligned}$$

for $i = 1, 2, \cdots, I$. Note that η_{il} is the ideal response of ith examinee to jth item when the q-vector is $\boldsymbol{\alpha_l}$. That is, $\eta_{il} = \prod_{k=1}^{K} \alpha_{ik}^{\alpha_{lk}}$. So η_{il} is similar but not identical to previously mentioned η_{ij}. And $\eta_{il} = 1$ means that examinee i belongs to group $\eta_j = 1$ when $\boldsymbol{\alpha_l}$ is the alternative q-vector.

Especially, for DINA model, only when the q-vector for the jth item is correctly specified, $P(X_{ij} = 1|\eta_{il} = 1) = 1 - s_j$ and $P(X_{ij} = 1|\eta_{il} = 0) = g_j$.

In the proposed method, we choose the q-vector $\boldsymbol{q_1}$ over $\boldsymbol{q_2}$ as correct q-vector, when one of the following conditions is satisfied:

(1) $SKLD_j(g_0, g_1|\boldsymbol{q_1}) > SKLD_j(g_0, g_1|\boldsymbol{q_2})$.

(2) $\sum_{k=1}^{k=K} q_{1k} < \sum_{k=1}^{k=K} q_{2k}$ and $SKLD_j(g_0, g_1|\boldsymbol{q_1}) = SKLD_j(g_0, g_1|\boldsymbol{q_2})$.

In other words, the correct q-vector should satisfy the following conditions:

(1) Its corresponding SKLD should approximate the maximum value.

(2) There are fewer attributes in this q-vector.

Unfortunately, the computation in Equation (5) is time-consuming for large value of K. In order to resolve this problem, cluster analysis is introduced to the method, which will be shown in the next section.

2.3 The K-means clustering

There are a few clustering algorithms for classification. Compared with other clustering algorithms, such as hierarchical clustering method, K-means method has lower computation complexity but with similar accuracy. K-means clustering is very popular in data classification, which aims to partition n observations into K clusters according to spatial extent. In most K-means algorithms, the number of clusters is pre-determined. Once the cluster centers are decided, data is assigned to corresponding clusters.

K-means clustering algorithm is used to search the correct q-vector in SKLD-based method. Using K-means clustering, all the alternative q-vectors could be divided into two clusters based on their SKLDs. One is underspecified and the other one includes correct or overspecified ones. The latter cluster has higher SKLDs. Within this cluster, correct one has less required attributes than overspecified ones, so the correct one could be selected.

The detailed algorithm to find out the final q-vector is as follows:

Step 1: Estimate parameter $s_{jl} = P(X_{ij} = 0|\eta_{il} = 1)$ and $g_{jl} = P(X_{ij} = 1|\eta_{il} = 0)$ based on EM algorithm for DINA model [10].

Step 2: Figure out the corresponding $SKLD_j(g_0, g_1|\boldsymbol{\alpha_l})$ for all $2^K - 1$ alternative q-vectors based on examinees' response data.

Step 3: Generate two initial cluster centers randomly within the set $C = \{SKLD_j(g_0, g_1|\boldsymbol{\alpha_l}), l = 1, \cdots, 2^K - 1\}$, labeled as $SKLD_1^{(0)}$ and $SKLD_2^{(0)}$ ($SKLD_1^{(0)} > SKLD_2^{(0)}$).

Step 4: Create two clusters by associating every SKLD value with the nearest centers, $C_1^{(0)}$ and $C_2^{(0)}$.

Step 5: Update the cluster centers.

$$SKLD_1^{(1)} = \frac{1}{\#C_1^{(0)}} \sum_{C_1^{(0)}} SKLD_j(g_0, g_1|\boldsymbol{\alpha_1}),$$
$$SKLD_2^{(1)} = \frac{1}{\#C_2^{(0)}} \sum_{C_2^{(0)}} SKLD_j(g_0, g_1|\boldsymbol{\alpha_1})$$

where "#" denotes the number of elements in the set.

Step 6: Repeat Step4 and Step5 until convergence has been reached. The number of iterations is marked as s.

Step 7: Obtain the final q-vector for jth item.

$$\boldsymbol{q_j} = \arg\min_{\boldsymbol{\alpha_l}} \{\sum_{k=1}^{K} \alpha_{lk}, SKLD_j(g_0, g_1|\boldsymbol{\alpha_1}) \in C_1^{(s)}\}$$

3. SIMULATION STUDIES

Three simulation studies are conducted to compare the performances of the Q-matrix validation methods based on SKLD, LOR, IDI (δ-method) indexes. As the δ-method relies on the cut-off points, 5 different cut-off points are considered in the simulation studies.

3.1 Simulation Study 1

Simulation Study 1 is used to investigate the sensitivities of those three methods. In the context of Q-matrix validation, sensitivity is the precondition of using related validation methods, which is evaluated via the true positive rate.

3.1.1 Design

In the simulation studies, the number of items is set as $J = 30$. The slipping and guessing parameters are set to 0.20 for all items. And the number of attributes is set as $K = 3$, 4, 5. The true Q-matrices with different number of attributes in the simulations are shown in Table 2.

The number of examinees takes four values, $I = 500$, 1000, 2000, 4000. We sample random vectors $\boldsymbol{\alpha_1}, \cdots, \boldsymbol{\alpha_I}$ independently and identically, where α_{ik} is from $B(1, p_k)$, for $i = 1, \cdots, I$, $k = 1, \cdots, K$. To make DINA model identifiable, p_k is set as 0.5 for all k.

Based on the examinees' attribute patterns and true Q-matrix, the simulated item responses under DINA model are generated. And then the true Q-matrix is also set as initial Q-matrix in the validation methods. At last, the true positive rates at Q-matrix or q-entries level would be calculated to evaluate the recovery. Simply speaking, let $Q = \{q_{jk}\}_{J\times K}$ be the true Q-matrix and $\widehat{Q}^{(t)}(Q) = \{\hat{q}_{jk}^{(t)}(Q)\}_{J\times K}$ be its validation at replication t, $t = 1, \cdots, T$, where T is the number of replications. The true positive rates at Q-matrix and q-entries levels could be estimated by

$$\mathrm{TP}_Q = \frac{\sum_{t=1}^{T} \mathrm{I}(\widehat{Q}^{(t)}(Q) = Q)}{T},$$
$$\mathrm{TP}_q = \frac{\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{k=1}^{K} \mathrm{I}(\hat{q}_{jk}^{(t)}(Q) = q_{jk})}{T * J * K},$$

where $\mathrm{I}(\cdot)$ is the indicator function. Actually, the true positive rate at Q-matrix level measures the percentage of entirely recovered Q-matrix across all replications. And the true positive rate at q-entries level indicates the proportions of correct q-entries, which is more fine-grained.

In addition, $T = 1000$ is set in the simulation studies.

3.1.2 Results

The sensitivity results of the three validation methods are presented in Table 3.

For the SKLD-based method, its performance on sensitivity is near-perfect across all 12 conditions. More concretely, at q-entries level, its true positive rates are always higher than 0.995. It is hardly surprising that the true positive rate at Q-matrix level is much lower than that at q-entries level under the same conditions. One possible explanation is that the recovery at Q-matrix level may be too strict. Because if just only one q-entry in the Q-matrix isn't recovered, the Q-matrix will be viewed as not recovered. What's more, at the same level, much lower true positive rates are found when the number of attributes is more and the number of examinees is less, especially for $(K = 4, I = 500)$, $(K = 5, I = 500)$ and $(K = 5, I = 1000)$. Based on the nature of CDMs, the sample size of some latent classes is not enough and sufficient to analysis under those 3 conditions. Taking $(K = 5, I = 500)$ as an example, there are $2^5 = 32$ latent classes for all 500 examinees, so each class has about 16 examinees on average. In some extreme cases, there would exist some empty latent classes, which may even cause the unidentifiability of model [2].

For the LOR-based method, a similar conclusion could be drawn from Table 3. A small difference is that its

Table 2. True Q-matrices in the simulation studies

Item	K = 3			K = 4				K = 5				
1	1	0	0	1	0	0	0	1	0	0	0	0
2	0	1	0	0	1	0	0	0	1	0	0	0
3	0	0	1	0	0	1	0	0	0	1	0	0
4	1	0	0	0	0	0	1	0	0	0	1	0
5	0	1	0	1	0	0	0	0	0	0	0	1
6	0	0	1	0	1	0	0	1	0	0	0	0
7	1	0	0	0	0	1	0	0	1	0	0	0
8	0	1	0	0	0	0	1	0	0	1	0	0
9	0	0	1	1	1	0	0	0	0	0	1	0
10	1	0	0	1	0	1	0	0	0	0	0	1
11	0	1	0	1	0	0	1	1	1	0	0	0
12	0	0	1	0	1	1	0	1	0	1	0	0
13	1	1	0	0	1	0	1	1	0	0	1	0
14	1	0	1	0	0	1	1	1	0	0	0	1
15	0	1	1	1	1	0	0	0	1	1	0	0
16	1	1	0	1	0	1	0	0	1	0	1	0
17	1	0	1	1	0	0	1	0	1	0	0	1
18	0	1	1	0	1	1	0	0	0	1	1	0
19	1	1	0	0	1	0	1	0	0	1	0	1
20	1	0	1	0	0	1	1	0	0	0	1	1
21	0	1	1	1	1	1	0	1	1	1	0	0
22	1	1	0	1	1	0	1	1	1	0	1	0
23	1	0	1	1	0	1	1	1	1	0	0	1
24	0	1	1	0	1	1	1	1	0	1	1	0
25	1	1	1	1	1	1	0	1	0	1	0	1
26	1	1	1	1	1	0	1	1	0	0	1	1
27	1	1	1	1	0	1	1	0	1	1	1	0
28	1	1	1	0	1	1	1	0	1	1	0	1
29	1	1	1	1	1	1	0	0	1	0	1	1
30	1	1	1	1	1	0	1	0	0	1	1	1

Table 3. The sensitivity results

K	I	TP_Q							TP_q						
		SKLD	LOR	IDI (Cut-off points)					SKLD	LOR	IDI (Cut-off points)				
				0	0.01	0.05	0.10	0.20			0	0.01	0.05	0.10	0.20
3	500	0.936	0.813	0.814	0.881	0.996	0.873	0	0.999	0.998	0.998	0.999	1.000	0.997	0.870
	1000	0.998	0.964	0.982	0.990	1.000	0.981	0	1.000	1.000	1.000	1.000	1.000	1.000	0.868
	2000	1.000	0.980	1.000	1.000	1.000	0.999	0	1.000	1.000	1.000	1.000	1.000	1.000	0.867
	4000	1.000	0.987	1.000	1.000	1.000	1.000	0	1.000	1.000	1.000	1.000	1.000	1.000	0.867
4	500	0.562	0.353	0.063	0.146	0.643	0.795	0	0.995	0.990	0.979	0.985	0.996	0.997	0.839
	1000	0.950	0.681	0.217	0.428	0.941	0.980	0	1.000	0.997	0.988	0.993	0.999	1.000	0.835
	2000	0.999	0.885	0.544	0.767	0.995	1.000	0	1.000	0.999	0.995	0.998	1.000	1.000	0.833
	4000	0.998	0.870	0.528	0.779	0.998	1.000	0	1.000	0.999	0.995	0.998	1.000	1.000	0.833
5	500	0.278	0.098	0.008	0.048	0.550	0.739	0	0.990	0.973	0.966	0.976	0.995	0.996	0.873
	1000	0.762	0.388	0.102	0.295	0.936	0.977	0	0.998	0.993	0.981	0.990	0.999	1.000	0.868
	2000	0.965	0.651	0.447	0.762	0.999	0.999	0	1.000	0.997	0.993	0.998	1.000	1.000	0.867
	4000	0.975	0.645	0.477	0.747	0.995	1.000	0	1.000	0.997	0.994	0.998	1.000	1.000	0.867

true positive rates are lower than that of SKLD-based method under the same conditions, especially at the q-entries level.

Table 3 also shows that the sensitivity of δ-method differs greatly for different cut-off points. Specially when its cut-off point is 0.2, the true positive rates at Q-matrix level are always 0, and that at q-entries level are almost no more than 0.9, which means that its sensitivity is relatively low. For cut-off points at 0.05 and 0.1, their true positive rates at both q-entries and Q-matrix level are usually higher than that under other cut-off points. When the cut-off points are 0 or 0.01, the sensitivity is slightly lower. It indicates

Table 4. The specificity results

K	I	R	TN_Q							TN_q						
			SKLD	LOR	IDI (Cut-off points)					SKLD	LOR	IDI (Cut-off points)				
					0	0.01	0.05	0.10	0.20			0	0.01	0.05	0.10	0.20
3	500	10%	0.851	0.742	0.748	0.838	0.986	0.881	0	0.998	0.997	0.997	0.998	1.000	0.997	0.868
		20%	0.741	0.670	0.716	0.825	0.980	0.890	0	0.997	0.995	0.996	0.998	1.000	0.997	0.863
	1000	10%	0.973	0.945	0.949	0.974	1.000	0.988	0	1.000	0.999	0.999	1.000	1.000	1.000	0.865
		20%	0.852	0.869	0.935	0.967	0.999	0.983	0	0.998	0.998	0.999	1.000	1.000	1.000	0.862
	2000	10%	0.991	0.981	0.995	0.999	1.000	1.000	0	1.000	1.000	1.000	1.000	1.000	1.000	0.866
		20%	0.934	0.962	0.996	0.999	1.000	1.000	0	0.999	1.000	1.000	1.000	1.000	1.000	0.867
	4000	10%	0.993	0.977	0.998	0.999	1.000	0.999	0	1.000	1.000	1.000	1.000	1.000	1.000	0.866
		20%	0.939	0.958	0.989	0.998	1.000	1.000	0	0.999	1.000	1.000	1.000	1.000	1.000	0.862
4	500	10%	0.436	0.128	0.004	0.012	0.412	0.778	0	0.991	0.978	0.960	0.970	0.993	0.996	0.835
		20%	0.283	0.044	0	0	0.094	0.506	0	0.984	0.965	0.927	0.943	0.981	0.992	0.832
	1000	10%	0.853	0.304	0.012	0.042	0.711	0.972	0	0.998	0.987	0.971	0.979	0.997	1.000	0.830
		20%	0.780	0.135	0	0	0.235	0.869	0	0.997	0.976	0.937	0.952	0.989	0.998	0.826
	2000	10%	0.970	0.483	0.011	0.079	0.927	1.000	0	1.000	0.992	0.977	0.985	0.999	1.000	0.830
		20%	0.960	0.169	0	0	0.397	0.979	0	0.999	0.980	0.942	0.954	0.993	1.000	0.819
	4000	10%	0.971	0.487	0.014	0.079	0.922	1.000	0	1.000	0.992	0.977	0.985	0.999	1.000	0.830
		20%	0.960	0.176	0	0	0.357	0.976	0	1.000	0.980	0.942	0.954	0.992	1.000	0.820
5	500	10%	0.068	0.014	0	0	0.042	0.039	0	0.982	0.965	0.929	0.943	0.977	0.977	0.861
		20%	0.074	0.012	0	0	0.066	0.111	0	0.978	0.957	0.940	0.953	0.981	0.980	0.865
	1000	10%	0.481	0.180	0	0	0.050	0.142	0	0.995	0.985	0.941	0.952	0.983	0.986	0.855
		20%	0.442	0.106	0	0.001	0.121	0.248	0	0.994	0.978	0.951	0.961	0.986	0.986	0.862
	2000	10%	0.843	0.364	0	0	0.072	0.299	0	0.999	0.991	0.943	0.954	0.986	0.990	0.854
		20%	0.830	0.242	0	0	0.197	0.358	0	0.999	0.986	0.957	0.964	0.989	0.990	0.860
	4000	10%	0.855	0.362	0	0	0.059	0.313	0	0.999	0.991	0.943	0.954	0.986	0.991	0.854
		20%	0.852	0.245	0	0	0.170	0.331	0	0.999	0.986	0.957	0.964	0.988	0.989	0.860

Note. R refers to the proportion of misspecified q-entries.

that this method could specify correct q-entries efficiently only if an appropriate cut-off point is chosen. On the whole, the true positive rate increases as the number of examinees increases and decreases as the number of attributes increases.

Through above analysis, to some extent, all the three methods could effectively retain the correct q-entries in the Q-matrix. However, the sensitivity of δ-method may highly rely on its cut-off point. So choosing an appropriate cut-off point becomes particularly important when δ-method is used. Unlike δ-method, the sensitivities of the validation method based on SKLD and LOR are independent of any additional condition, which means that they could work well and are easy to implement. So these two methods are more sensitive and applicable.

3.2 Simulation Study 2

Simulation Study 2 is used to explore the specificities of the three methods. Specificity is the core of validating misspecified Q-matrices, which is measured by the true negative rates. And it indicates the proportions of misspecified Q-matrices and q-entries that are correctly identified and calibrated.

3.2.1 Design

The design is very similar but not identical to Simulation Study 1. Specially, the setting of true parameters is same. The only difference is about the input Q-matrices. For specificity, misspecified Q-matrices are regarded as the initial matrix. Misspecified Q-matrices are conducted by altering 10% or 20% of q-entries from 1 to 0 or 0 to 1 in true Q-matrices. Similar to Simulation Study 1, the recovery rates—true negative rates at Q-matrix or q-entries level would be recorded. Let $Q^m = \{q^m_{jk}\}_{J\times K}$ be a misspecified Q-matrix and $\widehat{Q}^{(t)}(Q^m) = \{\hat{q}^{(t)}_{jk}(Q^m)\}_{J\times K}$ be its validation at replication t, $t = 1, \cdots, T$. The true negative rates at Q-matrix and q-entries levels could be estimated by

$$\mathrm{TN}_Q = \frac{\sum_{t=1}^{T} \mathrm{I}(\widehat{Q}^{(t)}(Q^m) = Q)}{T},$$

$$\mathrm{TN}_q = \frac{\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{k=1}^{K} \mathrm{I}(\hat{q}^{(t)}_{jk}(Q^m) = q_{jk})}{T * J * K}.$$

3.2.2 Results

Table 4 summaries the specificity results of the three methods.

For the SKLD-based method, the true negative rates at q-entries level range from 0.978 to 1.000, indicating the encouraging validating performance. Similar to Simulation Study 1, the true negative rates at Q-matrix level are lower than that at q-entries level, especially when $(K = 4, I = 500)$, $(K = 5, I = 500)$ and $(K = 5, I = 1000)$. Under those conditions, there are more attributes and less examinees which make it more difficult to detect the misspecification in Q-matrix. What's more, if the numbers of attributes and examinees are fixed, the specificity of SKLD-based method is little affected by the proportions of misspecified q-entries.

For the LOR-based method, its true negative rates are always lower than that of the SKLD-based method at both q-entries and Q-matrix level. In particular, when K is 5, the true negative rates of LOR-based method at Q-matrix level are always no more than 0.4, even when the number of examinees is 4000. Maybe LOR-based method couldn't apply to the condition that the number of attributes is too large.

For the δ-method, when K is 3, the true negative rates at both q-entries and Q-matrix level turn out satisfactory except the condition when its cut-off point is 0.2. However, the specificity is affected by the number of attributes greatly. Specially, when K is large, such as $K = 5$, the true negative rates of δ-method at Q-matrix level are very low and even close to 0. In comparison, when the cut-off point is 0.10, the specificity is higher than that under other cut-off points, nevertheless, not better than SKLD-based method.

When the Q-matrix is misspecified, the above results show that the validation methods based on LOR and IDI may be no longer valid when K is large. However, the SKLD-based method can still deal with this more challenging Q-matrix misspecification, even though the number of attribute is much larger.

3.3 Simulation Study 3

Simulation Study 3 is used to explore the relationship between the performances of the three methods and noise level.

3.3.1 Design

Following the idea of Culpepper [5], five different noise levels are considered in the simulation: Case 1, a low noise level, $s_j = g_j = 0.1$; Case 2, the slipping parameter is lower than the guessing parameter, $s_j = 0.1$, $g_j = 0.3$; Case 3, a medium noise level, $s_j = g_j = 0.2$; Case 4, the slipping parameter is higher than the guessing parameter, $s_j = 0.3$, $g_j = 0.1$; Case 5, a high noise level, $s_j = g_j = 0.3$. And only the Q-matrix with $K = 4$ is taken in the simulation study. And the Q-matrix with 20% misspecified q-entries is used to evaluate the specificity. The settings of examinees are the same as Simulation Study 1.

3.3.2 Results

Figure 1 shows the performances of the three methods under different noise levels. From the results, under Case 1, 2, 3, 4, δ-method with cut-off point 0.1 performs better than other cut-off points basically. And the performance of SKLD-based method is very close to that condition. However, under Case 5, δ-method with cut-off point 0.1 performs badly, while SKLD-based method performs stably and well. Even though the true positive rates at Q-matrix level of all the methods is 0, SKLD-based method do well at q-entries level. And under all cases, LOR-based method performs worse than SKLD-based method mainly.

Additionally, it's hardly surprising that all three methods performed better under lower noise level than under higher noise level. Actually, $1 - s_j - g_j$ is exactly the global item discrimination index under DINA model where an item with higher this index could discriminate different examinees better [20].

On the whole, δ-method highly depends on its cut-off points, and LOR-based method has worse sensitivity and specificity on some conditions. K-means algorithm is employed to avoid choosing cut-off points. In addition, SKLD, the product of IDI in δ-method and LOR, is considered to improve the recovery. So the SKLD-based method is a perfect combination of IDI and LOR, which is more stable and flexible.

4. REAL-DATA EXAMPLES

Besides the simulation studies, two real-data examples are considered to further evaluate the performance of SKLD-based method.

4.1 Example 1: fraction subtraction data

4.1.1 The data set

The data set used in this section is binary responses of 536 middle school students to 11 fraction subtraction items. The data set is a subset of data originally described and used by K.Tatsuoka in [23], and more recently by C.Tautsuoko in [21], de la Torre and Douglas in [8], Henson, Templin and Wilse in [14], and Decarlo in [6]. For this data set, several Q-matrices are available. The Q-matrices used in this article are Q_1 and Q_2, which are shown in Table 5. Specifically, Q_1 was defined by Henson, Templin and Wilse in [14], which

Table 5. The Q-matrices of fraction subtraction items

Item number	Item	Q_1	Q_2
1	$3\frac{1}{2} - 2\frac{3}{2}$	1 1 0	0 1 0
2	$3 - 2\frac{1}{5}$	1 0 1	0 0 1
3	$3\frac{7}{8} - 2$	1 0 1	0 0 1
4	$4\frac{4}{12} - 2\frac{7}{12}$	1 0 0	1 0 0
5	$4\frac{1}{3} - 2\frac{4}{3}$	1 1 0	0 1 0
6	$\frac{11}{8} - \frac{1}{8}$	1 1 0	0 1 0
8	$2 - \frac{1}{3}$	1 0 1	0 0 1
9	$4\frac{5}{7} - 1\frac{7}{4}$	1 1 1	0 0 1
10	$7\frac{3}{5} - \frac{4}{5}$	1 0 0	1 0 0
11	$4\frac{1}{10} - 2\frac{8}{10}$	1 0 0	1 0 0
13	$4\frac{1}{3} - 1\frac{5}{3}$	1 1 0	0 1 0

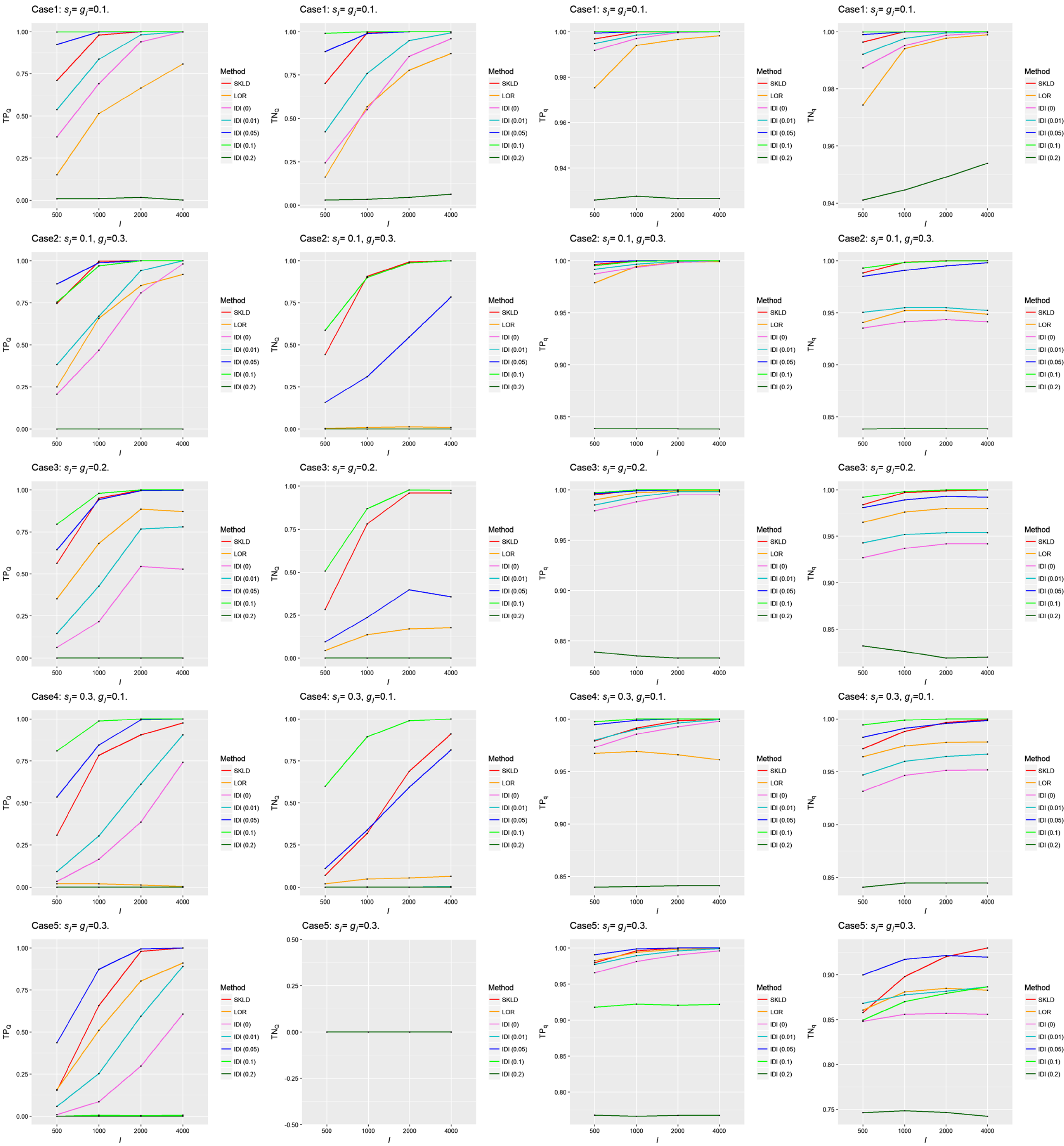

Figure 1. The performances of the three methods under different noise levels.

contains 3 attributes: (1) borrowing from a whole number, (2) separating a whole number from a fraction and (3) determining a common denominator. Q_2 is a modified Q-matrix of Q_1, contained in R package CDM. Unfortunately, Q_1 is not a complete Q-matrix as the completeness requires that at least one item just requiring one attribute is contained in the test for any attributes. This incompleteness would lead to the model unidentifiablity and further influence on

Table 6. The parameter estimations for DINA model based on Q_1 and Q_2

Item number	Q_1			Q_2		
	guessing	slipping	RMSEA	guessing	slipping	RMSEA
1	0.216	0.121	0.033	0.217	0.121	0.046
2	0.104	0.177	0.015	0.042	0.199	0.023
3	0.508	0.153	0.073	0.501	0.188	0.100
4	0.031	0.251	0.115	0.035	0.232	0.070
5	0.067	0.069	0.009	0.070	0.070	0.014
6	0.530	0.047	0.059	0.532	0.047	0.118
8	0.131	0.088	0.016	0.051	0.099	0.025
9	0.521	0.054	0.131	0.482	0.064	0.136
10	0.032	0.149	0.112	0.035	0.126	0.056
11	0.110	0.158	0.003	0.128	0.154	0.085
13	0.008	0.176	0.024	0.008	0.175	0.022

the performance of the validation method. So it is usually recommended to use a complete Q-matrix [2].

4.1.2 Method and results

The idea about the validations of Q_1 and Q_2 is following [4]. That is, both Q_1 and Q_2 are assumed to be correct. Before validation, model parameters need to be estimated firstly using the fraction subtraction data set by EM algorithm. And then generate the new response data sets based on the parameters estimates. Next, we construct the misspecified Q-matrix as follows. For matrix Q_1, randomly alter q-entries from 0 to 1 or from 1 to 0, whereas for Q_2, only alter q-entries from 0 to 1 because the q-vectors in Q_2 only have one q-entry of 1. Finally, based on the above steps, the proposed method can be employed to validate the misspecified Q-matrices similar to Simulation Study 2.

The results of parameter estimation under the two Q-matrices are present in Table 6. According to the results, the guessing parameters of Item 3 exceed 0.5 under both the two Q-matrices, leading to $g_3 > 1 - g_3$. Apparently, it is unreasonable that the correct response probabilities are higher than the incorrect response probabilities for examinees who have not mastered all required attributes. What's more, a good fit among the diagnostic assessment design, the response data, and the postulated DINA model is attained when the estimates of the slipping and guessing parameters both are small [24]. From this point of view, DINA model fits the data bad for some items, such as Item 3 and 4. Moreover, the root mean square error of approximations (RMSEAs) of some items are over 0.10, which also indicates the model could not fit data very well.

The results of validation Q-matrix show that about 53% misspecified q-entries have been validated for Q_1 and about 79% for Q_2. As mentioned earlier, the performance may be affected by the incompleteness of Q_1. Though Q_2 is complete, it is somewhat too simple for every q-vector just has only one q-entry of 1. To some extent, the poor fitness of some items would result in the slightly worse recovery.

Table 7. The Q-matrix for SDA6 dataset

Item	Q-matrix			
	α_1	α_2	α_3	α_4
1	1	0	0	0
2	0	1	0	0
3	1	0	0	0
4	0	1	0	0
5	0	1	0	0
6	0	1	0	0
7	0	0	1	0
8	0	0	1	0
9	0	0	0	1
10	0	0	1	0
11	0	0	0	1
12	0	0	1	0
13	1	0	0	0
14	0	0	1	0
15	0	0	0	1
16	0	0	1	0
17	1	0	0	0

4.2 Example 2: SDA6 data

4.2.1 The data set

The Sociocultural Dimension Assessment version 6 (SDA6) dataset is provided by Jurich and Bradshaw in [16]. The dataset contains 17 items to test 1710 examinees. For the dataset, Q-matrix requires 4 attributes: (1) critique whether a study's methods address the research questions (short for "critique methods"), (2) identify improvements that could strengthen inferences made from a study (short for "identify improvements"), (3) evaluate whether research methods protect a participant's well-being (short for "protect participants"), and (4) discern the study's generalizability (short for "discern generalizability"). The Q-matrix is shown in Table 7.

4.2.2 Method and results

DINA model is used to fit the response data provided in the SDA6 dataset. And SKLD-based method is employed to validate the Q-matrix in Table 7. This Q-matrix is used as the initial Q-matrix in the SKLD-based method. The results show that the proposal Q-matrix is the same as the initial Q-matrix. At the same time, we alter the q-entries of 0 in the Q-matrix to 1 randomly as the misspecified Q-matrices repeating 1000 times. And then SKLD-based method is used to validate those altered Q-matrices. About 68% of misspecified Q-matrices are correctly revised.

5. DISCUSSION

Q-matrix validation has become a vital part in cognitive diagnosis for misspecified Q-matrix may involve many fateful consequences in model fit. We propose a SKLD-based method to validate misspecified Q-matrix in this paper. Actually the proposed method also is a perfect combination of IDI and LOR. The process of this method is divided into the following steps: (1) estimating the guessing and slipping parameters based on EM algorithm for DINA model, (2) figuring out the SKLD values for all alternative q-vectors, (3) dividing all alternative q-vectors into two classes based on K-means cluster analysis algorithm, (4) selecting the SKLD values approximating the maximum but with fewer attributes as the correct q-vector. Simulation studies with high rates of misspecification are designed to evaluate the performance of the proposed method. The results of simulations indicate that SKLD-based method is able to identify and correctly validate misspecified q-vectors, and at the same time retain those correct ones.

Despite such promising results, there are some other issues in need of deep thinking. The first one is whether the proposed method could be extended to more complicated CDM models, such as Higher-Order DINA model proposed by [8] and generalized DINA model in [11]. Second, it's unclear whether the number of clusters can be changed to improve the performance. More simulation studies should be conducted to test and verify the effect of the number of clusters. Third is whether the proposed method could provide accurate estimation rather than validation of the Q-matrix based on the response data. Finally, the number of attributes is assumed known and correct. It is still a considerable issue to extend the proposed method to specify the number of attributes. In conclusion, to generalize the proposed method, there is a deal of work to be done.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (grant number 11571069).

Received 13 June 2018

REFERENCES

[1] Chen, J. (2017). A Residual-Based Approach to Validate Q-Matrix Specifications. *Applied Psychological Measurement.* **40** 1–17.

[2] Chen, Y., Liu, J., Xu, G. and Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association.* **110** 850–866. MR3367269

[3] Chiu, C.-Y., Douglas, J. and Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika.* **74** 633–665. MR2565331

[4] Chiu, C.-Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement.* **37** 598–618.

[5] Culpepper, S. A. (2015). Bayesian estimation of DINA model with Gibbs sampling. *Journal of Educational and Behavioral Statistics.* **40** 454–476.

[6] DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: DINA model, classification, latent class sizes, & the Q-matrix. *Applied Psychological Measurement.* **35** 8–26.

[7] DeCarlo, L. T. (2012). Recognizing uncertainty in the Q-Matrix via a Bayesian extension of DINA model. *Applied Psychological Measurement.* **36** 447–468.

[8] De la Torre, J. and Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika.* **69** 333–353. MR2272454

[9] De la Torre, J. (2008). An empirically based method of Q-matrix validation for DINA model: Development and applications. *Journal of Educational Measurement.* **45** 343–362.

[10] De la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics.* **34** 115–130.

[11] De La Torre, J. (2011). The generalized DINA model framework. *Psychometrika.* **76** 179–199. MR2788881

[12] De la Torre, J. and Chiu, C. Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika.* **81** 253–273. MR3505366

[13] Henson, R. A. and Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement.* **81** 262–277. MR2142701

[14] Henson, R. A., Templin, J. L. and Wilse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika.* **74** 191–210. MR2507377

[15] Junker, B. W. and Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement.* **25** 258–272. MR1842982

[16] Jurich, D. P. and Bradshaw, L. P. (2014). An illustration of diagnostic classification modeling in student learning outcomes assessment. *International Journal of Testing.* **14** 49–72.

[17] Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics.* **22** 79–86. MR0039968

[18] Liu, J., Xu, G. and Ying, Z. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement.* **36** 548–564.

[19] Rupp, A. A. and Templin, J. (2008). Effects of Q-matrix misspecification on parameter estimates and misclassification rates in DINA model. *Educational and Psychological Measurement.* **68** 78–98. MR2416509

[20] Rupp, A. A., Templin, J. L. and Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications.* Guilford, New York.

[21] Tatsuoka, C. (2002). Data-analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society, Series C: Applied Statistics.* **51** 337–350. MR1920801

[22] Tatsuoka, C. and Ferguson, T. (2003). Sequential classification on partially ordered sets. *Journal of the Royal Statistical Society, Series B: Statistical Methodology.* **65** 143–158. MR1959819

[23] Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. *Diagnostic Monitoring of*

Skill & Knowledge Acquisition 453–488.

[24] Templin, J. L. and Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*. **11** 287–305.

[25] Xu, X., Chang, H. and Douglas, J. (2003). Computerized adaptive testing strategies for cognitive diagnosis. Paper presented at the annual meeting of National Council on Measurement in Education, Montreal, Canada.

Jinxin Guo
School of Mathematics and Statistics
Northeast Normal University
5268 Renmin Street, Changchun, Jilin Province
China
E-mail address: guojx235@nenu.edu.cn

Xin Xu
School of Mathematics and Statistics
Northeast Normal University
5268 Renmin Street, Changchun, Jilin Province
China
E-mail address: xux636@nenu.edu.cn

Jian Tao
School of Mathematics and Statistics
Northeast Normal University
5268 Renmin Street, Changchun, Jilin Province
China
E-mail address: taoj@nenu.edu.cn

Ningzhong Shi
School of Mathematics and Statistics
Northeast Normal University
5268 Renmin Street, Changchun, Jilin Province
China
E-mail address: shinz@nenu.edu.cn