An MCMC Algorithm for Estimating the Reduced RUM

Meng-ta Chung$^1$ and Matthew S. Johnson$^2$

$^1$ Department of Human Development, Columbia University
$^2$ Office of Medical Research, St. Joseph’s Hospital
Abstract

The RRUM is a model that is frequently seen in language assessment studies. The objective of this research is to advance an MCMC algorithm for the Bayesian RRUM. The algorithm starts with estimating correlated attributes. Using a saturated model and a binary decimal conversion, the algorithm transforms possible attribute patterns to a Multinomial distribution. Along with the likelihood of an attribute pattern, a Dirichlet distribution is used as the prior to sample from the posterior. The Dirichlet distribution is constructed using Gamma distributions. Correlated attributes of examinees are generated using the inverse transform sampling. Model parameters are estimated using the Metropolis within Gibbs sampler sequentially. Two simulation studies are conducted to evaluate the performance of the algorithm. The first simulation uses a complete and balanced Q-matrix that measures 5 attributes. Comprised of 28 items and 9 attributes, the Q-matrix for the second simulation is incomplete and imbalanced. The empirical study uses the ECPE data obtained from the CDM R package. Parameter estimates from the MCMC algorithm and from the CDM R package are presented and compared. The algorithm developed in this research is implemented in R.

Keywords
CDM, RUM, RRUM, Q-matrix, Bayesian, MCMC
Introduction

Cognitive diagnostic assessment (CDA) is a framework that aims to evaluate whether an examinee has mastered a particular cognitive process called attribute (Leighton & Gierl, 2007). In CDA, exam items are each associated with attributes that are required for mastery. Using examinees’ attribute states, CDA provides effective information for examinees to improve their learning and for educators to adjust their teaching. Studies have demonstrated that CDA is a valid application for providing useful diagnostic feedback in language assessment (e.g., Jang, 2009; Jang et al., 2013; Kim, 2011; Kim, 2014; Li & Suen, 2012; Richards, 2008).

Concerning the current thinking and future directions of CDA, Language Testing publishes a special issue (Volume 32, Issue 3, July 2015) that integrates insights from experts in the field of language assessment.

In recent years, a few cognitive diagnosis models (CDMs) have been developed, including the deterministic input, noisy-and gate (DINA) model (Junker & Sijtsma, 2001), the noisy input, deterministic-and gate (NIDA) model (Maris, 1999), and the reparameterized unified model (RUM) (Hartz, 2002; Hartz, Roussos, & Stout, 2002). All these models use the Q-matrix (Tatsuoka, 1983) to measure attribute states of examinees. Suppose there are \( I \) examinees taking the exam that measures \( K \) attributes. A binary matrix \( A_{I \times K} = (\alpha_{ik})_{I \times K} \) reveals the connection between examinees and attributes. If examinee \( i \) does not master attribute \( k \), then \( \alpha_{ik} = 0 \); if examinee \( i \) masters attribute \( k \), then \( \alpha_{ik} = 1 \).

In order to evaluate examinees with respect to their levels of competence of each attribute in an exam, the Q-matrix (Tatsuoka, 1983) is used to partition exam items into attributes. The
Q-matrix is a binary matrix that shows the relationship between exam items and attributes. Given an exam with \( J \) items that measure \( K \) attributes, the Q-matrix is represented as a \( J \) by \( K \) matrix, \( Q_{J \times K} = (q_{jk})_{J \times K} \). In a Q-matrix, if attribute \( k \) is required by item \( j \), then \( q_{jk} = 1 \). If attribute \( k \) is not required by item \( j \), then \( q_{jk} = 0 \).

Among all of the CDMs, the RUM is frequently seen in language assessment research. Extending the NIDA model, Maris (1999) proposed a model that attempts to estimate the slip and guess parameters for different items. That is, the the slip and guess parameters have subscripts for both items and attributes. To improve this model, Dibello, Stout, and Roussos (1995) advances the unified model that incorporates a unidimensional ability parameter. However, these two models are not statistically identifiable. Hartz (2002) reparameterizes the unified model so that the parameters of the model can be identified while retaining their interpretability. As is expected, this reparameterized unified model is a more complicated conjunctive CDMs (Roussos, Templin, & Hensen, 2007). The RUM defines the probability of a correct response to an item as

\[
\pi_j^* = \prod_{k=1}^{K} (1 - s_{jk})^{q_{jk}},
\]

and the penalty for each attribute no possessed as

\[
r_{jk}^* = g_{jk} / (1 - s_{jk}).
\]

\( \pi_j^* \) is the probability that an examinee, having acquired all the attributes required for item
will correctly apply these attributes in solving the item. That is, $\pi^*$ is interpreted as an item difficulty parameter. $r_{jk}^*$ is used to define the penalty of not mastering the $k^{th}$ attribute. Under this view, $r_{jk}^*$ can be seen as an indicator of the diagnostic capacity of item $j$ for attribute $k$. Also from the perspective of monotonicity, $1 - s_{jk}$ should be greater than $g_{jk}$. Explicitly, $r_{jk}^*$ should be constrained to the interval $(0, 1)$.

Incorporating a general ability measure, $P_{c_j}(\theta_i)$, the probability of a correct response in the RUM can be written as

$$P(X_{ij} = 1|\alpha, r^*, \pi^*, \theta) = \pi^*_j \prod_{k=1}^{K} (r_{jk}^*(1-\alpha ik))^{q_{jk}} P_{c_j}(\theta_i).$$

$P_{c_j}(\theta_i)$ is the item characteristic curve in the Rasch model, where $c_j$ is the difficulty parameter and $\theta_i$ is the general measure of an examinee’s knowledge not specified by the Q-matrix.

The RUM has larger flexibility than other CDMs in modeling the probability of correct item response for different attribute patterns. This flexibility, however, is achieved at the cost of introducing a significant degree of complexity into the estimation process. Assuming that the Q-matrix completely specifies the attributes required by the exam items, Hartz (2002) further suggests a reduced version of the RUM (RRUM) that sets $P_{c_j}(\theta_i) = 1$. The parameters of the RRUM retain the model identifiable and allow the probabilities of slipping and guessing to vary across items. The IRF of the RRUM is therefore reduced to

$$P(X_{ij} = 1|\alpha, r^*, \pi^*) = \pi^*_j \prod_{k=1}^{K} (r_{jk}^*(1-\alpha ik))^{q_{jk}}. \quad (3)$$
Based on the assumptions of local independence and independence among examinees, the joint likelihood function for all responses in the RRUM is

\[
P(X_{ij} = x_{ij}, \forall i, j | \alpha, r^*, \pi^*) = \prod_{i=1}^{I} \prod_{j=1}^{J} \left( \pi_j^{x_{ij}} \prod_{k=1}^{K} r_j^{(1-\alpha_{jk})a_{jk}} \right)^{x_{ij}} \left( 1 - \pi_j^{x_{ij}} \prod_{k=1}^{K} r_j^{(1-\alpha_{jk})a_{jk}} \right)^{1-x_{ij}}.
\]

The RRUM is a simplified yet practical model that has received considerable attention among psychometricians and educators (e.g., Chiu & Köhn, 2016; Feng, Habing, & Huebner, 2014; Henson & Templin, 2007; Jang, 2009; Jang et al., 2013; Kim, 2011; Kim, 2014; Templin, 2004; Templin et al., 2004; Templin & Douglas, 2004; Zhang, 2013). Nevertheless, the RRUM remains more complex than other CDMs. Due to its complexity, the RRUM have been mostly estimated in a Bayesian framework. Hartz (2002) uses a Bayesian method to estimate the RRUM, and Hartz, Roussos, & Stout (2002) develops the patented Arpeggio program, which is commonly applied to analyze the data in language assessment studies.

This research proposes a different MCMC algorithm for estimating the Bayesian RRUM, with the hope of reducing the complexity of computation. Specifically, a saturated model using the inverse transform sampling is used to estimate correlated attributes, and the Metropolis with Gibbs sampling is adopted to estimate the \( \pi^* \) and \( r^* \) parameters. The proposed algorithm, as well as a way to simulate data, are implemented in R (R Development Core Team, 2017). With the algorithm, it is readily flexible for researchers and practitioners to code using any programming languages.
Proposed MCMC Algorithm

The setting for the estimation is comprised of responses from \( I \) examinees to \( J \) items that measure \( K \) attributes. Given a \( J \times K \) Q-matrix, the following steps perform sequentially at iteration \( t, t = 1, \ldots, T \).

Step 1: Binary Decimal Conversion

With \( K \) attributes, there are a total of \( 2^K \) possible attribute patterns for examinee \( i \). Let \( 2^K = M \), and let the matrix \( x_{M \times K} = (x_{mk})_{M \times K} \) be the matrix of possible attribute patterns. Each of the \( M \) rows in \( x \) represents a possible attribute pattern, which is converted to a decimal number by

\[
(b_n b_{n-1} \ldots b_0)_2 = b_n(2)^n + b_{n-1}(2)^{n-1} + \cdots + b_0(2)^0,
\]

where \((b_n b_{n-1} \ldots b_0)_2\) denotes a binary number.

After the conversion, these \( M \) possible attribute patterns become a Multinomial distribution. To estimate correlated attributes, a saturated Multinomial model is used that assumes no restrictions on the probabilities of the attribute patterns (see Maris, 1999). Assuming a Dirichlet prior \( \theta \), the hierarchical model for estimating attributes is

\[
x|\theta \sim \text{Multinomial}(M, \theta),
\]

\[
\theta \sim \text{Dirichlet}(a_1, a_2, \ldots, a_M).
\]

Step 2: Updating Probability of Attribute Pattern

Let \( y \) and \( q \) be the data and the Q-matrix. The full conditional posterior distribution is

\[
p(x|y, \pi^*, r^*, q) \propto p(y|x, \pi^*, r^*, q)p(x|\theta)p(\theta).
\]

As the conjugate prior for a Multinomial distribution is also a Dirichlet distribution, \( p(x|\theta)p(\theta) \) is a Dirichlet distribution. There-
fore, use $\text{Dirichlet}(1, 1, \ldots, 1)$ as the prior, and the conditional posterior is distributed as $\text{Dirichlet}(1 + y_1, 1 + y_2, \ldots, 1 + y_M)$, where $y_\ell \ (\ell = 1, \ldots, M)$ is the number examinees possessing the $\ell^{th}$ attribute pattern. As no function in base R can be used to sample from Dirichlet distribution, Gamma distributions are used to construct the Dirichlet distribution. In step 2, suppose that $w_1, \ldots, w_M$ are distributed as $\text{Gamma}(a_1, 1), \ldots, \text{Gamma}(a_M, 1)$, and that $\tau = w_1 + \cdots + w_M$, then $(w_1/\tau, w_2/\tau, \ldots, w_M/\tau)$ is distributed as $\text{Dirichlet}(a_1, a_2, \ldots, a_M)$.

For each of the $M$ possible attribute patterns, step 2 calculates the total number of examinees $(y_1, y_2, \ldots, y_M)$ falling into an attribute pattern, and then samples from $\text{Gamma}(1 + y_1, 1) = w'_1, \text{Gamma}(1 + y_2, 1) = w'_2, \ldots, \text{Gamma}(1 + y_M, 1) = w'_M$. Let $\tau' = w'_1 + w'_2 + \cdots + w'_M$, and we can get $p(x|\theta)p(\theta) = (w'_1/\tau', w'_2/\tau', \ldots, w'_M/\tau')$. Along with the likelihood of each possible attribute pattern, which is $p(y|x, \pi^*, r^*, q)$, step 2 obtains the full conditional posterior.

**Step 3: Updating Attribute**

The full conditional posterior distribution is sampled using the discrete version of inverse transform sampling. Let the posterior $(p_1, p_2, \ldots, p_M)$ be the PMF of the $M$ possible attribute patterns. The CDF is computed by adding up the probabilities for the $M$ points of the distribution. To sample from this discrete distribution, we partition $(0, 1)$ into $M$ subintervals $(0, p_1)$, $(p_1, p_1 + p_2), \ldots, (\sum_{m=0}^{M-1} p_m, \sum_{m=0}^M p_m)$, and then generate a value $u$ from $\text{Uniform}(0, 1)$.

Updating the attribute state of examinee $i$ is achieved by checking which subinterval the value $u$ falls into. This subinterval number (a decimal number) is then converted to its corresponding binary number (see step 1) that represents the attribute state of examinee $i$. 

7
After step 3 is applied to each examinee, attribute states for all examinees, denoted as $\alpha$, are obtained for iteration $t$.

**Step 4: Updating $r^*$ and $\pi^*$ Parameters**

A Metropolis within Gibbs algorithm is used to sample $\pi^*$ and $r^*$. The non-informative $Beta(1, 1)$ prior is applied in updating $r^*$ and $\pi^*$. Candidate values for $r^*$ is sampled from $Uniform(r^{*(t-1)} - \delta, r^{*(t-1)} + \delta)$. It should be noted that candidate values for $r^*$ are restricted to the interval $(0, 1)$, and that $\delta$ is adjusted so that the acceptance rate is between $25\%$ and $40\%$ (see Gilks et al., 1996). The updated $\alpha$ from step 3 is carried to step 4. As $\pi^*$ and $r^*$ are assumed to be independent of each other, $p(\pi^*, r^*) = p(\pi^*)p(r^*)$.

In updating $r^*$ at iteration $t$, the acceptance probability $\varphi_r$ for the candidate value $r^{*({t})}$ is calculated by

$$\varphi_r = \frac{p(y|\alpha^{(t)}, r^{*({t})}, \pi^{*({t-1})}, q)p(r^{*({t})})}{p(y|\alpha^{(t)}, r^{*({t-1})}, \pi^{*({t-1})}, q)p(r^{*({t-1})})},$$

and $r^{*({t})}$ is then set by

$$r^{*({t})} = \begin{cases} r^{*({t})} & \text{with probability } \min(1, \varphi_r) \\ r^{*({t-1})} & \text{otherwise} \end{cases}$$
With the obtained $r^{s(t)}$, the acceptance probability for updating $\pi^*$ is

$$
\varphi_{\pi} = \frac{p(y|\alpha^{(t)}, r^*, \pi^{s(t)}, q)p(\pi^{s(t)})}{p(y|\alpha^{(t)}, r^*, \pi^{s(t-1)}, q)p(\pi^{s(t-1)})},
$$

and $\pi^{s(t)}$ is decided by

$$
\pi^{s(t)} = \begin{cases} 
\pi^{s(*)} & \text{with probability } \min(1, \varphi) \\
\pi^{s(t-1)} & \text{otherwise}
\end{cases}
$$

Simulation Study

Procedure for Simulating Data

To investigate the effectiveness of the proposed MCMC algorithm, simulation studies are conducted to see how well the true attribute states could be recovered. Simulated data sets are generated using the following procedure.

The first step is to generate correlated attributes. Let $\theta$ be the $N$ by $K$ underlying probability matrix of $\alpha$, and let column $k$ of $\theta$ be a vector $\theta_k$, $k = 1, \ldots, K$. That is, $\theta = (\theta_1, \ldots, \theta_K)$. A copula is used to generate intercorrelated $\theta$ (see Ross, 2013). The correlation coefficient for each pair of columns in $\theta$ takes a constant value $\rho$, and the correlation matrix $\Sigma$ is expressed as

$$
\Sigma = \begin{bmatrix} 1 & \rho \\ & \ddots \\ \rho & 1 \end{bmatrix},
$$

9
Table 1: Q-matrix for Simulation I

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

where the off-diagonal entries are $\rho$. Each entry in $\Sigma$ corresponds to the correlation coefficient between two columns in $\theta$. Symmetric with all the eigenvalues positive, $\Sigma$ is a real symmetric positive-definite matrix that can be decomposed as $\Sigma = \nu^T \nu$ using Choleski decomposition, where $\nu$ is an upper triangular matrix.

After $\nu$ is derived, create an $I \times K$ matrix $\tau$, in which each entry is generated from $N(0, 1)$. $\tau$ is then transformed to $\gamma$ by using $\gamma = \tau \nu$, so that $\gamma$ and $\Sigma$ will have the same correlation structure. Set $\Phi(\gamma) = \theta$, where $\Phi(\cdot)$ is the cumulative standard normal
Table 2: Q-matrix for Simulation II

| Item | Attribute | Item | Attribute |
|------|-----------|------|-----------|
| 1    | 1 0 1 0 0 0 0 0 0 | 20   | 0 0 0 1 0 0 0 0 1 0 |
| 2    | 0 0 1 0 0 1 0 0 0 | 21   | 0 1 0 0 0 0 0 0 0 0 |
| 3    | 0 1 1 0 0 0 0 0 0 | 22   | 0 0 1 1 0 0 1 0 0 0 |
| 4    | 1 0 1 0 1 0 0 0 0 | 23   | 0 0 0 0 0 1 0 0 0 0 |
| 5    | 0 0 0 1 1 0 1 1 0 | 24   | 0 0 1 1 0 0 0 0 0 0 |
| 6    | 0 0 0 0 1 0 0 0 0 | 25   | 0 0 0 1 0 0 0 1 0 0 |
| 7    | 0 1 0 0 0 1 1 0 0 | 26   | 0 0 1 0 1 0 0 0 0 0 |
| 8    | 0 0 0 1 0 0 0 0 0 | 27   | 0 1 0 0 0 0 0 0 0 0 |
| 9    | 0 1 0 0 0 0 0 0 0 | 28   | 0 0 0 0 0 1 1 0 0 0 |
| 10   | 0 1 0 0 0 0 0 0 0 | 29   | 0 1 0 0 0 0 0 0 0 0 |
| 11   | 1 0 0 0 0 1 0 0 0 | 30   | 0 0 0 1 0 0 0 0 0 1 |
| 12   | 0 0 1 1 0 0 0 0 0 | 31   | 0 0 0 0 0 1 0 0 0 0 |
| 13   | 0 0 0 0 0 0 0 1 0 | 32   | 1 0 0 0 0 1 0 0 0 0 |
| 14   | 1 0 0 1 0 0 0 0 0 | 33   | 1 0 1 0 0 0 0 0 0 0 |
| 15   | 0 0 0 0 0 1 0 0 0 | 34   | 0 0 0 0 1 0 0 0 0 0 |
| 16   | 0 0 0 0 0 1 0 0 0 | 35   | 0 0 0 0 1 0 0 0 0 0 |
| 17   | 0 0 0 1 0 0 0 1 0 | 36   | 0 0 1 1 0 0 0 0 0 0 |
| 18   | 0 0 0 0 1 0 0 0 0 | 37   | 0 0 0 0 0 0 0 0 0 1 |
| 19   | 0 1 0 0 0 0 0 0 0 |

distribution function. \( \alpha \) is determined by

\[
\alpha_{ik} = \begin{cases} 
1 & \text{if } \theta_{ik} \geq \Phi^{-1}\left(\frac{k}{K+1}\right), \\
0 & \text{otherwise}
\end{cases}
\]

where \( k = 1, 2, \ldots, K \) (see Chiu, Douglas, & Li, 2009). Note that the above method can also be used to generate correlated attributes for the DINA and NIDA models.
The next step is to draw $\pi^*$ and $r^*$. Set $g_{jk}$ and $s_{jk}$ to 0.2, and $\pi^*$ and $r^*$ are obtained respectively from equations (1) and (2). Probability of an examinee correctly answer an item is calculated using equation (3), thus forming a matrix $y = (y_{nj})_{N \times J}$. The data is then generated using inverse transform sampling for two points 0 and 1. Another matrix $\xi = (\xi_{nj})_{N \times J}$ is created where each element is generated from $Uniform(0, 1)$, and then $\xi$ is compared with $y$. If the element in $\xi$ is greater than the corresponding element in $y$, set $y_{nj}$ to 0; if otherwise, then set $y_{nj}$ to 1. The simulated data $y$ is thus generated.

For $M$ simulated data sets, let $\hat{\alpha}^{(m)} = (\hat{\alpha}_{nk}^{(m)})_{N \times K}$ ($m = 1, \ldots, M$) be the estimated Q-matrix from $m^{th}$ data set, and let $\alpha = (\alpha_{nk})_{N \times K}$ represents the true $\alpha$. To measure how well each method recovers the true $\alpha$, the measure of accuracy $\Delta_\alpha$, confined between 0 and 1, is defined as

$$\Delta_\alpha = \frac{1}{M} \sum_{m=1}^{M} \left( 1 - \left\lceil \frac{\hat{\alpha}^{(m)} - \alpha}{NK} \right\rceil \right), \ m = 1, 2, \ldots, M,$$

where the $\lceil \cdot \rceil$ returns the value rounded to the nearest integer and $| \cdot |$ is the absolute value.

**Q-matrix in Simulation**

The Q-matrix (Table 1) for simulation I is obtained from de la Torre (2008). 30 items that measure 5 attributes comprise this artificial Q-matrix, which is constructed in a way that each attribute appears alone, in a pair, or in triple the same number of times as other attributes. This balanced Q-matrix, with each attribute being measured by 12 items, appears to have a clear pattern that implies main effects from items 1 to 10, two-way interactions from items 11
to 20 and three-way interactions from items 21 to 30. This Q-matrix is complete, containing at least one item devoted solely to each attribute (Chen, Liu, Xu, & Ying, 2015).

The Q-matrix (Table 2) for simulation II is acquired from Jang (2009), which discusses second language speakers’ reading comprehension. This complex Q-matrix is imbalanced and incomplete, consisting of 37 items that assess 9 attributes. For both simulations, examinees in groups of 500, 1000 and 2000 are simulated with the correlation between each pair of attributes set to 0.1, 0.3 and 0.5 for simplicity, as in Feng, Habing, & Huebner (2014). 20 data sets are simulated for each concoction. Corresponding R codes are run 7000 iterations after 2000 burn-in periods.

**Results**

The $\delta$ is set to 0.052 in step 4, so that the acceptance rate is around 35%. The Raftery and Lewis diagnostic (Raftery & Lewis, 1992) from the CODA R package (Plummer et al., 2006) suggests that $\pi^*$ and $r^*$ estimates are converged. Table 3 presents the results from simulations I and II. For the complete and balanced Q-matrix in simulation I, the measure of accuracy $\Delta_\alpha$ ranges from 0.919 to 0.941. For the incomplete and imbalanced Q-matrix in simulation II, the $\Delta_\alpha$ is less than 0.9 but above 0.8, ranging from 0.822 to 0.843.

It should be noted that using the independent model for simulation I with sample size 2000 and correlation 0.5, we notice that the average $\Delta_\alpha$ of 20 data sets drops to 0.835, indicating that using the saturate model for correlated attributes is indeed improving the accuracy of attribute estimates.

**Empirical Study**
Obtained from the CDM R package, the data consists of responses of 2922 examinees to 28 multiple choice items that measure 3 attributes (morphosyntactic, cohesive, lexical) in the grammar section of the Examination for the Certificate of Proficiency in English (ECPE). A standardized English as a foreign language examination, the ECPE is recognized in several countries as official proof of advanced proficiency in English (ECPE, 2015).

The CDM R package is also used to compare with the results from the MCMC algorithm. Specifically, the function with arguments `gdina(data, q.matrix, maxit=1000, rule="RRUM")` is applied. Note that the empirical Q-matrix (Table 5) is complete but imbalanced.

Table 4 shows the classification rate of each attribute pattern. Table 5 exhibits parameter estimates from the MCMC algorithm and the CDM R package. Applying the marginal maximum likelihood estimation, the CDM R package is implemented using the EM algorithm. As can be seen in Table 5, parameter estimates from the two methods do not deviate much.

**Discussion**

The current research proposes an MCMC algorithm for estimating parameters of the RRUM
Table 4: Classification Rate

| Method | Attribute Pattern |        |        |        |        |        |        |
|--------|------------------|--------|--------|--------|--------|--------|--------|
|        | (0,0,0)          | (0,0,1) | (0,1,0) | (0,1,1) | (1,0,0) | (1,0,1) | (1,1,0) | (1,1,1) |
| MCMC   | 0.309            | 0.120  | 0.006  | 0.186  | 0.004  | 0.007  | 0.002  | 0.369   |
| CDM R  | 0.294            | 0.124  | 0.020  | 0.181  | 0.010  | 0.013  | 0.008  | 0.353   |

Note. CDM R stands for the CDM R package

in a Bayesian framework. The algorithm is summarized as follows. Using the binary decimal conversion, possible attribute patterns are transformed to a Multinomial distribution (step 1). Along with the likelihood of an attribute pattern, a Dirichlet distribution is used as the prior to sample from the posterior. The Dirichlet distribution is constructed using Gamma distributions (step 2), and attributes of examinees are updated using the inverse transform sampling (step 3). Sequentially, $r^*$ and $\pi^*$ are generated using the Metropolis within Gibbs sampler (step 4). Of note is that steps 1 to 3 can also be used in estimating correlated attributes in the DINA and NIDA models.

Like most of the studies, the first simulation uses a complete and balanced Q-matrix. The measure of accuracy is on average 0.929. However when the Q-matrix is incomplete and imbalanced as in the second simulation, the measure of accuracy drops to an average of 0.833. A similar result is also revealed using the EM algorithm in the CDM R package. Therefore, one should be cautious when using a complex Q-matrix for the RRUM.

Another issue is the correlation between each pair of attributes. As can be seen from Table 3, when the sample size increases, the measure of accuracy increases as expected. However, when the correlation between each pair of attributes is higher, the measure of accuracy is
| Item | Mor | Coh | Lex | $\pi^*$ | $r^*$ | $\pi^*$ | $r^*$ |
|------|-----|-----|-----|--------|-------|--------|-------|
|      |     |     |     | Mor    | Coh    | Lex    | Mor    | Coh    | Lex    |
| E1   | 1   | 1   | 0   | 0.926 | 0.876 | 0.853 | 0.928 | 0.875 | 0.851 |
| E2   | 0   | 1   | 0   | 0.906 | 0.813 |       | 0.905 | 0.812 |       |
| E3   | 1   | 0   | 1   | 0.780 | 0.636 | 0.833 | 0.784 | 0.640 | 0.824 |
| E4   | 0   | 0   | 1   | 0.824 | 0.564 |       | 0.825 | 0.562 |       |
| E5   | 0   | 0   | 1   | 0.956 | 0.779 |       | 0.957 | 0.779 |       |
| E6   | 0   | 0   | 1   | 0.926 | 0.760 |       | 0.927 | 0.760 |       |
| E7   | 1   | 0   | 1   | 0.940 | 0.737 | 0.704 | 0.943 | 0.738 | 0.704 |
| E8   | 0   | 1   | 0   | 0.966 | 0.841 |       | 0.966 | 0.840 |       |
| E9   | 0   | 0   | 1   | 0.787 | 0.673 |       | 0.788 | 0.672 |       |
| E10  | 1   | 0   | 0   | 0.888 | 0.574 |       | 0.892 | 0.575 |       |
| E11  | 1   | 0   | 1   | 0.924 | 0.763 | 0.701 | 0.925 | 0.769 | 0.695 |
| E12  | 1   | 0   | 1   | 0.728 | 0.522 | 0.371 | 0.733 | 0.527 | 0.362 |
| E13  | 1   | 0   | 0   | 0.905 | 0.726 |       | 0.907 | 0.727 |       |
| E14  | 1   | 0   | 0   | 0.821 | 0.660 |       | 0.826 | 0.658 |       |
| E15  | 0   | 0   | 1   | 0.957 | 0.761 |       | 0.958 | 0.761 |       |
| E16  | 1   | 1   | 0   | 0.906 | 0.751 | 0.715 | 0.909 | 0.753 | 0.714 |
| E17  | 0   | 1   | 1   | 0.943 | 0.916 | 0.923 | 0.943 | 0.919 | 0.920 |
| E18  | 0   | 0   | 1   | 0.910 | 0.785 |       | 0.910 | 0.785 |       |
| E19  | 0   | 0   | 1   | 0.838 | 0.537 | 0.839 | 0.839 | 0.538 |       |
| E20  | 1   | 0   | 1   | 0.754 | 0.500 | 0.516 | 0.759 | 0.501 | 0.511 |
| E21  | 1   | 0   | 1   | 0.917 | 0.849 | 0.705 | 0.917 | 0.854 | 0.699 |
| E22  | 0   | 0   | 1   | 0.796 | 0.371 | 0.797 | 0.797 | 0.370 |       |
| E23  | 0   | 1   | 0   | 0.936 | 0.704 |       | 0.936 | 0.699 |       |
| E24  | 0   | 1   | 0   | 0.698 | 0.479 |       | 0.696 | 0.474 |       |
| E25  | 1   | 0   | 0   | 0.771 | 0.676 |       | 0.775 | 0.674 |       |
| E26  | 0   | 0   | 1   | 0.782 | 0.691 |       | 0.783 | 0.690 |       |
| E27  | 1   | 0   | 0   | 0.689 | 0.422 |       | 0.695 | 0.421 |       |
| E28  | 0   | 0   | 1   | 0.909 | 0.701 | 0.910 | 0.910 | 0.701 |       |

Note. Mor = morphosyntactic; Coh = cohesive; Lex = lexical
counterintuitively lower. Chen, Liu, Xu, & Ying (2015) also observes a similar phenomenon in their Q-matrix research based on the DINA model. A heuristic explanation according to Chen, Liu, Xu, & Ying (2015) is that the simulated data has more observations with the attribute pattern \((0, 0, 0, 0, 0)\) when the correlation is higher. For a sample size of 1000 in simulation I, there are around 90, 60 and 20 examinees having attribute pattern \((0, 0, 0, 0, 0)\) for correlations 0.5, 0.3 and 0.1, respectively. Because it is more difficult to identify \((0, 0, 0, 0, 0)\), the algorithm needs more \((0, 0, 0, 0, 0)\) examinees to estimate accurately. Therefore when the correlation is higher, the performance of the algorithm is better.

The complete Q-matrix for the empirical study measures only 3 attributes although imbalanced. The result is consistent with that from the CDM R package and from Feng, Habing, & Huebner (2014). It is suggested that future research compare the estimated examinees’ attribute patterns with the estimate from other CDMs such as the popular DINA model.
References

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, 110(510), 850-866.

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

Chiu, C.-Y., & Köhn, H.-F. (2016). The reduced RUM as a logit model: Parameterization and constraints. *Psychometrika*, 81: 350.

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

DiBello, L., Roussos, L. A., & Stout, W. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. V. Rao & S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26, Psychometrics, pp. 979-1027). Amsterdam, the Netherlands: Elsevier.

ECPE (2015). ECPE 2015 Report (p. 1). The Examination for the Certificate of Proficiency in English (ECPE). Retrieved from [http://cambridgemichigan.org/institutions/products-services/tests/proficency-certification/ecpe/](http://cambridgemichigan.org/institutions/products-services/tests/proficency-certification/ecpe/).

Feng, Y., Habing, B. T., & Huebner, A. (2014). Parameter estimation of the Reduced RUM using the EM algorithm. *Applied Psychological Measurement*, 38, 137–150.

Hartz, S. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practicality* (Doctoral dissertation). University of Illinois, Urbana-Champaign.

Hartz, S., Roussos, L., & Stout, W. (2002). Skills diagnosis: Theory and practice. Unpublished manuscript. University of Illinois at Urbana Champaign.

Henson, R., & Templin, J. (2007, April). *Importance of Q-matrix construction and its effects cognitive diagnosis model results*. Paper presented at the annual meeting of the National Council on Measurement in Education in Chicago, Illinois.
Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for applying Fusion Model to LanguEdge assessment. *Language Testing*, 26(1), 31–73.

Jang, E. E., Dunlop, M., Wagner, M., Kim, Y. H., & Gu, Z. (2013). Elementary school ELLs’ reading skill profiles using cognitive diagnosis modeling: Roles of length of residence and home language environment. *Language Learning*, 63(3), 400–436.

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

Kim, Y. H. (2011). Diagnosing EAP writing ability using the reduced Reparameterized Unified Model. *Language Testing*, 28, 509–541.

Kim, A. Y. (2014). Exploring ways to provide diagnostic feedback with an ESL placement test: Cognitive diagnostic assessment of L2 reading ability. *Language Testing*, 32(2): 227–258.

Leighton, J.P., & Gierl, M.J. (Eds.). (2007). *Cognitive diagnostic assessment for education. Theory and applications*. Cambridge, MA: Cambridge University Press.

Li, H., & Suen, H. K. (2013). Constructing and validating a Q-matrix for cognitive diagnostic analyses of a reading test. *Educational Assessment*, 18(1), 1–25.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187–212.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC, R News, vol 6, 7-11

R Development Core Team. (2017). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Available from [http://www.r-project.org](http://www.r-project.org).

Raftery, A.E. & Lewis, S.M. (1992). One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science*, 7, 493-497.
Richards, B. (2008). Formative Assessment in Teacher Education: The Development of a Diagnostic Language Test for Trainee Teachers of German. *British Journal of Educational Studies*, 56(2), 184-204.

Ross, S. M. (2006), *Simulation*. 4th ed., Academic Press, San Diego.

Roussos, L. A., Templin, J. L., & Henson, R. A. (2007). Skills diagnosis using IRT-based latent class models. *Journal of Educational Measurement*, 44, 293-311.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345–354.

Templin, J. (2004). Estimation of the RUM without alpha tilde: a general model for the proficiency space of examinee ability. External Diagnostic Research Group Technical Report.

Templin, J., & Douglas, J. (2004). Higher order RUM. External Diagnostic Research Group Technical Report.

Templin, J., Henson, R., Templin, S., & Roussos, L. (2004). Robustness of unidimensional hierarchical modeling of discrete attribute association in cognitive diagnosis models. External Diagnostic Research Group Technical Report.