THE REDUCED RUM AS A LOGIT MODEL: PARAMETERIZATION AND CONSTRAINTS

CHIA-YI CHIU

RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY

HANS-FRIEDRICH KÖHN

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Cognitive diagnosis models (CDMs) for educational assessment are constrained latent class models. Examinees are assigned to classes of intellectual proficiency defined in terms of cognitive skills called attributes, which an examinee may or may not have mastered. The Reduced Reparameterized Unified Model (Reduced RUM) has received considerable attention among psychometricians. Markov Chain Monte Carlo (MCMC) or Expectation Maximization (EM) are typically used for estimating the Reduced RUM. Commercial implementations of the EM algorithm are available in the latent class analysis (LCA) routines of Latent GOLD and Mplus, for example. Fitting the Reduced RUM with an LCA routine requires that it be reparameterized as a logit model, with constraints imposed on the parameters. For models involving two attributes, these have been worked out. However, for models involving more than two attributes, the parameterization and the constraints are nontrivial and currently unknown. In this article, the general parameterization of the Reduced RUM as a logit model involving any number of attributes and the associated parameter constraints are derived. As a practical illustration, the LCA routine in Mplus is used for fitting the Reduced RUM to two synthetic data sets and to a real-world data set; for comparison, the results obtained by using the MCMC implementation in OpenBUGS are also provided.

Key words: cognitive diagnosis, general cognitive diagnostic models, LCDM, Reduced RUM, EM, MCMC, Mplus.

The Reduced Reparameterized Unified Model (Reduced RUM; Hartz, 2002; Hartz, Roussos, Henson, & Templin, 2005) has received considerable attention among psychometric researchers concerned with cognitive diagnosis models (CDMs) for educational assessment. CDMs are constrained latent class models. For a given ability domain, classes of intellectual proficiency are defined in terms of binary cognitive skills called *attributes*, which an examinee may or may not have mastered (DiBello, Roussos, & Stout, 2007; Haberman & von Davier, 2007; Leighton & Gierl, 2007; Rupp, Templin, & Henson, 2010). Expectation Maximization (EM) or Markov Chain Monte Carlo (MCMC; de la Torre, 2009, 2011; Henson, Templin, & Willse, 2009; von Davier, 2005, 2008, 2011) are used to obtain maximum likelihood estimates of the model parameters that are then used to assign examinees to proficiency classes.

If a researcher wants to use MCMC for fitting the Reduced RUM, then he or she can either choose to write his or her own code, or refer to the MCMC routines implemented, for example, in OpenBUGS (Lunn, Spiegelhalter, Thomas, & Best, 2009) or the Arpeggio Suite (Bolt et al., 2008). Alternatively, the Reduced RUM can be fitted by the EM algorithm. The first option is again that a user writes his or her own code (see, for example, Feng, Habing, & Huebner, 2014). Second, a user can resort to a commercial package that offers an implementation of the EM algorithm for fitting (constrained) latent class models, for example, Latent GOLD (Vermunt & Magidson, 2000) and Mplus (Muthén & Muthén, 1998–2011).

Using a latent class analysis (LCA) routine for fitting the Reduced RUM requires that it be re-expressed as a logit model, with constraints imposed on the parameters of the logistic function.

The parameterization and the associated constraints have been worked out for the Reduced RUM involving two attributes (Henson et al., 2009). However, for more than two attributes, the specific reparameterization and the constraints to be imposed on the parameters of the logistic function are nontrivial and currently unknown.

This article intends to close this gap: The general reparameterization of the Reduced RUM as a logit model involving any number of attributes is presented including the associated parameter constraints. Thus, a researcher can now use the LCA routine, say in Mplus, for fitting educational data by the Reduced RUM.

The presentation is divided into a theoretical and a practical/applied part. The next section briefly reviews definitions and technical key concepts of (general) CDMs and the Reduced RUM. The mathematical derivations and proofs are presented in the subsequent sections. As a practical illustration, the Reduced RUM is fitted to two synthetic data sets and to a real-world data set using the EM algorithm implemented in the constrained LCA routine in Mplus (key parts of the syntax are provided in the appendices). For comparison, all data sets were also fitted with the MCMC routine available in OpenBUGS.

1. Technical Background

1.1. (General) Cognitive Diagnosis Models

Suppose that K binary cognitive skills or attributes constitute a certain ability domain; there are then $M = 2^K$ distinct attribute profiles, each of which characterizes a proficiency class. Let the K-dimensional vector, $\boldsymbol{\alpha}_m = (\alpha_1, \ldots, \alpha_K)^T$, $m = 1, \ldots, M$, denote the binary attribute profile of proficiency class m, where the kth entry indicates whether the respective attribute has been mastered. (Throughout the text, the superscript T denotes the transpose of vectors or matrices; the "prime notation" is reserved for distinguishing between vectors or their scalar entries.) Consider a test of J items for assessing ability in the domain. Each individual item j is associated with a K-dimensional binary vector, $\mathbf{q}_j$, called item-attribute profile, where $q_{jk} = 1$ if a correct answer requires mastery of the kth attribute, and 0 otherwise. Note that item-attribute profiles consisting entirely of zeroes are inadmissible, because they correspond to items that require no skills at all. Hence, given K attributes, there are at most $2^K - 1$ distinct item-attribute profiles. The J item-attribute profiles of a test constitute its Q-matrix, $\mathbf{Q} = \{q_{jk}\}_{(J \times K)}$ (Tatsuoka, 1983, 1985) that summarizes the constraints specifying the associations between items and attributes.

CDMs differ in the way in which mastery and nonmastery of the attributes are believed to affect an examinee's performance on a test item (e.g., compensatory models vs non-compensatory models; conjunctive models vs disjunctive models; for a detailed discussion, see Henson et al., 2009). General CDMs (de la Torre, 2011; Henson et al., 2009; Rupp et al., 2010; von Davier, 2005, 2008, 2011) express the functional relation between attribute mastery and the probability of a correct item response in a unified mathematical form and parameterization that are applicable to "recognizable" CDMs (de la Torre, 2011, p. 181), as discussed previously in the literature, and CDMs "that have not yet been defined" (Henson et al., p. 199), thereby establishing a general standard for model comparison and evaluation.

Define the "kernel" (Rupp et al., 2010, p. 135), of item j, $g(\mathbf{q}_j, \boldsymbol{\alpha})$, as the linear combination of all K attribute main effects, α_k, and their interactions

$$g(\mathbf{q}_j, \boldsymbol{\alpha}) = \gamma_{j0} + \sum_{k=1}^{K} \gamma_{jk} q_{jk} \alpha_k + \sum_{k'=k+1}^{K} \sum_{k=1}^{K-1} \gamma_{jkk'} q_{jk} q_{jk'} \alpha_k \alpha_{k'} + \cdots + \gamma_{j12\ldots K} \prod_{k=1}^{K} q_{jk} \alpha_k, \quad (1)$$

where q_{jk} indicates whether mastery of attribute α_k is required for item j. The attribute interactions are expressed as product terms. For example, the two-way interaction of attributes α_k and $\alpha_{k'}$ is written as $q_{jk}q_{jk'}\alpha_k\alpha_{k'}$. (The order of an interaction corresponds to the number of parenthetical attribute subscripts of the associated coefficient, $\gamma_{j(\ldots)}$.) Attribute main effects and interaction terms can be removed from the kernel by constraining the corresponding entries in the parameter vector, $\boldsymbol{\gamma}_j = (\gamma_{j0}, \gamma_{jk}, \gamma_{j(kk')}, \ldots, \gamma_{j(12\ldots K)})^T$, to zero.

Let Y_j denote the response to the binary test item j. The expression of the item response function (IRF), $P(Y_j = 1 \mid \boldsymbol{\alpha})$, in terms of $g(\mathbf{q}_j, \boldsymbol{\alpha})$ must guarantee that $0 \leq P(Y_j = 1 \mid \boldsymbol{\alpha}) \leq 1$, which, for example, can be achieved by using the logit link, $P(Y_j = 1 \mid \boldsymbol{\alpha}) = e^{g(\mathbf{q}_j, \boldsymbol{\alpha})}/(1 + e^{g(\mathbf{q}_j, \boldsymbol{\alpha})})$. Based on this general logistic function, von Davier, 2005, 2008, 2011 defined the IRF of his General Diagnostic Model (GDM) by constraining all interaction terms in $g(\mathbf{q}_j, \boldsymbol{\alpha})$ to be zero (see Equations 1 and 2; von Davier, 2005, pp. 3–4). Henson et al. (2009) used the unconstrained general logistic function as the IRF of their Log-linear Cognitive Diagnosis Model (LCDM; see Equation 11 in Henson et al., 2009, p. 197). The specific IRFs of recognizable CDMs like the Deterministic Input Noisy Output "AND" gate (DINA) model (Junker & Sijtsma, 2001, Macready & Dayton, 1977), the Deterministic Input Noisy Output "OR" gate (DINO) model (Templin & Henson, 2006), and the Reduced RUM were then derived by Henson et al. (2009) through constraining the coefficients in $g(\mathbf{q}_j, \boldsymbol{\alpha})$. The logit link was also used by de la Torre (2011) for the IRF of a general CDM that he called the Generalized DINA (G-DINA) model (see Equation 2; de la Torre, 2011, p. 181.). In addition, de la Torre (2011) proposed to use the identity link, $P(Y_j = 1 \mid \boldsymbol{\alpha}) = g(\mathbf{q}_j, \boldsymbol{\alpha})$, and the log link, $P(Y_j = 1 \mid \boldsymbol{\alpha}) = e^{g(\mathbf{q}_j, \boldsymbol{\alpha})}$, for expressing the IRF of the G-DINA model (see Equations 1 and 3; de la Torre, 2011, pp. 181–182; these two link functions require further constraints on the parameters to guarantee that $P(Y_j = 1 \mid \boldsymbol{\alpha})$ is bounded by 0 and 1). The IRFs of various recognizable CDMs were derived by de la Torre (2011) based on the G-DINA model.

1.2. The Reduced RUM as a General Cognitive Diagnosis Model Based on the Logit Link

The IRF of the Reduced RUM in its traditional parameterization is

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = \pi_j^* \prod_{k=1}^{K} r_{jk}^{*\, q_{jk}(1-\alpha_k)}, \tag{2}$$

where $0 < \pi_j^* < 1$ denotes the probability of a correct answer for an examinee who has mastered all the attributes required by item j, and $0 < r_{jk}^* < 1$ is a penalty parameter for not mastering the kth attribute.

For $K = 2$ attributes, Henson et al. (2009) obtained the expression of the Reduced RUM as a general CDM based on the logit link by first transforming Equation 2 to the equivalent, but mathematically more tractable Inverse RUM, which then allowed for deriving the expressions of the parameters of the logistic function (see Equations 25–27, Henson et al., 2009, p. 201). The IRF is

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = \frac{e^{\beta_{j0}+\beta_{j1}q_{j1}\alpha_1+\beta_{j2}q_{j2}\alpha_2+\beta_{j12}q_{j1}q_{j2}\alpha_1\alpha_2}}{1 + e^{\beta_{j0}+\beta_{j1}q_{j1}\alpha_1+\beta_{j2}q_{j2}\alpha_2+\beta_{j12}q_{j1}q_{j2}\alpha_1\alpha_2}},$$
$$\text{subject to:} \quad \beta_{jk} > 0 \quad k = 1, 2 \tag{3}$$

where

$$\beta_{j12} = \ln\left(\frac{1 + e^{\beta_{j0}}}{1 + e^{\beta_{j0}+\beta_{j1}} + e^{\beta_{j0}+\beta_{j2}} - e^{\beta_{j0}+\beta_{j1}+\beta_{j2}}}\right) \tag{4}$$

The constraint on the β_{jk} is mathematically not required because $0 < \pi_j^*, r_{jk}^* < 1$. However, $\beta_{jk} > 0$ is necessary to guarantee monotonicity—that is, for each examinee, who masters certain attributes, the probability of a correct response must be equal to or greater than the probability of a correct response when these attributes are not mastered (Henson et al., 2009, p. 198).

The inclusion of the interaction effect, $\alpha_1 \alpha_2$, in the model—in addition to the two main effects—is needed for modeling the probability of a correct response to an item that requires the mastery of two attributes. Specifically, the coefficient β_{j12} quantifies the relation between item j and attribute 2 conditional on the mastery of attribute 1 (and vice versa—see, Henson et al., 2009, p. 198).

1.3. The Coefficient β_{j12}: Further Considerations

The functional expression of the coefficient β_{j12} of the interaction term in Equation 4 implies the constraint—not explicitly mentioned in Henson et al. (2009)—that the argument of the log function be strictly positive (the logarithm of a negative argument is not defined). Because it always holds that $1 + e^{\beta_{j0}} > 0$, the constraint reduces to the requirement that the denominator of the fraction in parentheses be strictly positive:

$$1 + e^{\beta_{j0}+\beta_{j1}} + e^{\beta_{j0}+\beta_{j2}} - e^{\beta_{j0}+\beta_{j1}+\beta_{j2}} > 0 \tag{5}$$

Certain software packages (e.g., `Mplus`) cannot handle this constraint in the form of Equation 5, but require that it be reformulated as an upper bound on one of the parameters, β_{j0}, β_{j1}, or β_{j2}. Thus, in arbitrarily choosing β_{j2} as the parameter with the highest index $k = K = 2$, Equation 5 is re-expressed as

$$e^{\beta_{j2}} < \frac{1 + e^{\beta_{j0}+\beta_{j1}}}{e^{\beta_{j0}+\beta_{j1}} - e^{\beta_{j0}}}$$

Note that $e^{\beta_{j0}+\beta_{j1}} - e^{\beta_{j0}} > 0$ must be true because $e^{\beta_{j0}} > 0$ and $e^{\beta_{j1}} > 1$ due to $\beta_{j1} > 0$. Hence, $\beta_{j2} < \ln\left(1 + e^{\beta_{j0}+\beta_{j1}}\right) - \ln\left(e^{\beta_{j1}} - 1\right) - \beta_{j0}$, and the constraints on Equation 3 are given in explicit form as

(1) $0 < \beta_{j1}$

(2) $0 < \beta_{j2} < \ln\left(1 + e^{\beta_{j0}+\beta_{j1}}\right) - \ln\left(e^{\beta_{j1}} - 1\right) - \beta_{j0}$

2. The Reduced RUM as a General Cognitive Diagnosis Model Based on the Log Link

The Reduced RUM can also be expressed as a general CDM using the log link. Rewrite Equation 2 as

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = \pi_j^* \prod_{k=1}^{K} r_{jk}^{* \, q_{jk}(1-\alpha_k)}$$

$$= e^{\ln(\pi_j^*) + \sum_{k=1}^{K} \ln\left(r_{jk}^{* \, q_{jk}(1-\alpha_k)}\right)}$$

$$= e^{\ln(\pi_j^*) + \sum_{k=1}^{K} \left(q_{jk} \ln(r_{jk}^*) - q_{jk}\alpha_k \ln(r_{jk}^*)\right)}$$

$$= e^{\ln(\pi_j^*) + \sum_{k=1}^{K} \ln(r_{jk}^*) q_{jk} + \sum_{k=1}^{K} - \ln(r_{jk}^*) q_{jk}\alpha_k}$$

Setting

$$\delta_{j0} = \ln(\pi_j^*) + \sum_{k=1}^{K} \ln(r_{jk}^*) q_{jk}$$

$$\delta_{jk} = -\ln(r_{jk}^*)$$

then results in the IRF of the Reduced RUM as a general CDM with log link:

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = e^{\delta_{j0} + \sum_{k=1}^{K} \delta_{jk} q_{jk} \alpha_k} \tag{6}$$

$$\text{subject to: (1) } \delta_{j0} < 0$$

$$\text{(2) } 0 < \sum_{k=1}^{K} \delta_{jk} \alpha_k < |\delta_{j0}|$$

Constraints (1) and (2) are implied by the restrictions on the traditional parameters of the Reduced RUM, $0 < \pi_j^*, r_{jk}^* < 1$. (Note that $r_{jk}^* = e^{-\delta_{jk}}$ due to $\delta_{jk} = -\ln(r_{jk}^*)$; and because $\delta_{j0} = \ln(\pi_j^*) + \sum_{k=1}^{K} \ln(r_{jk}^*) q_{jk}$ it holds that $\pi_j^* = e^{\delta_{j0} + \sum_{k=1}^{K} \delta_{jk} q_{jk}}$.) Observe that the log link-based IRF of the Reduced RUM has a simpler parameter structure than the logit model because no interaction effects are included in the model—the log-link form is a main-effects-only model.

3. The Reduced RUM as a General Cognitive Diagnosis Model: The Connection Between the Log Link and the Logit Link

Henson et al. (2009) derived the reparameterization of the Reduced RUM as a logit model from the equivalent, but mathematically more tractable Inverse RUM. This section demonstrates that the log link provides an alternative transformation for deriving the Reduced RUM in the form of a general CDM using the logit link.

Consider item j requiring the mastery of K attributes (i.e., $\mathbf{q}_j = (11 \ldots 1)^T$, a K-dimensional vector of ones); its IRF based on the log link is given by Equation 6. The parameters of the logit model are derived by inspecting Equation 6 separately for each of the $2^K = M$ attribute profiles of the different proficiency classes.

3.1. The Attribute Profile $\boldsymbol{\alpha} = (00 \cdots 0)^T$

The response probabilities of examinees in the first proficiency class are adequately modeled by setting all parameters in Equation 6 to zero except δ_{j0}:

$$P(Y_j = 1 \mid \boldsymbol{\alpha}_1) = e^{\delta_{j0}} \tag{7}$$

The equivalent IRF using the logit link must result in the same item response probability. Thus, the logit model too must be an intercept-only model:

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = \frac{e^{\beta_{j0}}}{1 + e^{\beta_{j0}}} \tag{8}$$

Then, equating 7 and 8 and solving for δ_{j0} gives

$$\delta_{j0} = \beta_{j0} - \ln \left(1 + e^{\beta_{j0}} \right)$$

3.2. The Attribute Profiles $\boldsymbol{\alpha} = \mathbf{e}_k$

Let $\mathbf{e}_k$ denote a unit vector with the kth element equal to 1, and the remaining entries all 0. Thus, $\boldsymbol{\alpha} = \mathbf{e}_k$ indicates the single-attribute profile of the proficiency class whose examinees master only the kth attribute. Thus, the IRF using the log link is

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = e^{\delta_{j0} + \delta_{jk}}$$

The equivalent logit model is

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = \frac{e^{\beta_{j0} + \beta_{jk}}}{1 + e^{\beta_{j0} + \beta_{jk}}}$$

Equating the two IRFs, substituting $\delta_{j0} = \beta_{j0} - \ln\left(1 + e^{\beta_{j0}}\right)$, and solving for δ_{jk} gives

$$\delta_{jk} = \beta_{jk} - \ln\left(1 + e^{\beta_{j0} + \beta_{jk}}\right) + \ln\left(1 + e^{\beta_{j0}}\right)$$

3.3. The Attribute Profiles $\boldsymbol{\alpha} = (1 \cdots 10 \cdots 0)^T$, where $||\boldsymbol{\alpha}|| = K'$, $K' = 2, \ldots, K$

Without loss of generality, the entries in $\boldsymbol{\alpha}$ are assumed to be ordered such that the first $K' = 2, \ldots, K$ positions in $\boldsymbol{\alpha}$ are occupied by the entries $\alpha_k = 1$. The IRF based on the log link is

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = e^{\delta_{j0} + \sum_{k=1}^{K'} \delta_{jk}}$$

which upon substituting for δ_{j0} and δ_{jk} the expressions obtained earlier is written as

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = e^{\beta_{j0} + \sum_{k=1}^{K'} \beta_{jk} - \sum_{k=1}^{K'} \ln\left(1 + e^{\beta_{j0} + \beta_{jk}}\right) + (K'-1)\ln\left(1 + e^{\beta_{j0}}\right)} \tag{9}$$

It is trivial to verify that the equivalent IRF using the logit link must include all $\sum_{k=2}^{K'} \binom{K'}{k}$ interaction terms:

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = \frac{e^{\beta_{j0} + \sum_{k=1}^{K'} \beta_{jk} + \sum_{k'=k+1}^{K'} \sum_{k=1}^{K'-1} \beta_{jkk'} + \cdots + \beta_{j1\ldots K'}}}{1 + e^{\beta_{j0} + \sum_{k=1}^{K'} \beta_{jk} + \sum_{k'=k+1}^{K'} \sum_{k=1}^{K'-1} \beta_{jkk'} + \cdots + \beta_{j1\ldots K'}}} \tag{10}$$

(Otherwise, the fundamental requirement were violated that the item response probabilities of Equations 9 and 10 must be equal.) Equating the IRFs of Equations 9 and 10 and solving for $\beta_{j1\ldots K'}$ then results in

$$\beta_{j1\ldots K'} = \ln\left(\frac{\left(1 + e^{\beta_{j0}}\right)^{K'-1}}{\prod_{k=1}^{K'}\left(1 + e^{\beta_{j0} + \beta_{jk}}\right) - \left(1 + e^{\beta_{j0}}\right)^{K'-1} e^{\beta_{j0} + \sum_{k=1}^{K} \beta_{jk}}}\right)$$
$$- \sum_{k'=k+1}^{K'} \sum_{k=1}^{K'-1} \beta_{jkk'} - \cdots - \sum_{k_{K'-1}=k_{K'-2}+1}^{K'} \cdots \sum_{k_2=k_1+1}^{3} \sum_{k_1=1}^{2} \beta_{jk_1\ldots k_{K'-1}} \tag{11}$$

From the general expression of $\beta_{j1...K'}$ in Equation 11 the coefficient of any K'-way interaction term is readily obtained provided the indices of the summation and product operators are adjusted accordingly. For example, consider item j requiring the mastery of $K = 3$ attributes. If the coefficients of the three two-way interactions are sought, then $K' = 2$ and Equation 11 provides

$$
\begin{aligned}
\beta_{j1K'} = \beta_{j12} &= \ln\left(\frac{(1 + e^{\beta_{j0}})^{2-1}}{\prod_{k=1}^{2}(1 + e^{\beta_{j0}+\beta_{jk}}) - (1 + e^{\beta_{j0}})^{2-1}e^{\beta_{j0}+\sum_{k=1}^{2}\beta_{jk}}}\right) \\
&= \ln\left(\frac{1 + e^{\beta_{j0}}}{(1 + e^{\beta_{j0}+\beta_{j1}})(1 + e^{\beta_{j0}+\beta_{j2}}) - (1 + e^{\beta_{j0}})e^{\beta_{j0}+\beta_{j1}+\beta_{j2}}}\right)
\end{aligned}
$$

Note, however, that for β_{j13} and β_{j23}, the index k must be adjusted to the values in the sets $\{1, 3\}$ and $\{2, 3\}$, respectively, that refer to the indices of the main effects constituting the corresponding interaction terms:

$$
\begin{aligned}
\beta_{j13} &= \ln\left(\frac{(1 + e^{\beta_{j0}})^{2-1}}{\prod_{k\in\{1,3\}}(1 + e^{\beta_{j0}+\beta_{jk}}) - (1 + e^{\beta_{j0}})^{2-1}e^{\beta_{j0}+\sum_{k\in\{1,3\}}\beta_{jk}}}\right) \\
&= \ln\left(\frac{1 + e^{\beta_{j0}}}{(1 + e^{\beta_{j0}+\beta_{j1}})(1 + e^{\beta_{j0}+\beta_{j3}}) - (1 + e^{\beta_{j0}})e^{\beta_{j0}+\beta_{j1}+\beta_{j3}}}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\beta_{j23} &= \ln\left(\frac{(1 + e^{\beta_{j0}})^{2-1}}{\prod_{k\in\{2,3\}}(1 + e^{\beta_{j0}+\beta_{jk}}) - (1 + e^{\beta_{j0}})^{2-1}e^{\beta_{j0}+\sum_{k\in\{2,3\}}\beta_{jk}}}\right) \\
&= \ln\left(\frac{1 + e^{\beta_{j0}}}{(1 + e^{\beta_{j0}+\beta_{j2}})(1 + e^{\beta_{j0}+\beta_{j3}}) - (1 + e^{\beta_{j0}})e^{\beta_{j0}+\beta_{j2}+\beta_{j3}}}\right)
\end{aligned}
$$

The coefficient of the three-way interaction (i.e., $K' = K = 3$) is obtained from Equation 11 as

$$
\begin{aligned}
\beta_{j123} &= \ln\left(\frac{(1 + e^{\beta_{j0}})^{(3-1)}}{\prod_{k=1}^{3}(1 + e^{\beta_{j0}+\beta_{jk}}) - (1 + e^{\beta_{j0}})^{3-1}e^{\beta_{j0}+\sum_{k=1}^{3}\beta_{jk}}}\right) - \sum_{k'=k+1}^{3}\sum_{k=1}^{2}\beta_{jkk'} \\
&= \ln\left(\frac{(1 + e^{\beta_{j0}})^2}{(1 + e^{\beta_{j0}+\beta_{j1}})(1 + e^{\beta_{j0}+\beta_{j2}})(1 + e^{\beta_{j0}+\beta_{j3}}) - (1 + e^{\beta_{j0}})^2 e^{\beta_{j0}+\beta_{j1}+\beta_{j2}+\beta_{j3}}}\right) \\
&\quad - \beta_{j13} - \beta_{j12} - \beta_{j23}
\end{aligned}
$$

3.4. The Constraint on $\beta_{j1...K'}$

The expression of $\beta_{j1...K'}$ in Equation 11 is defined only if the argument of the log function is strictly positive. This condition is satisfied if

$$
\prod_{k=1}^{K'}\left(1 + e^{\beta_{j0}+\beta_{jk}}\right) - \left(1 + e^{\beta_{j0}}\right)^{K'-1}e^{\beta_{j0}+\sum_{k=1}^{K'}\beta_{jk}} > 0 \tag{12}
$$

Recall that certain software packages (e.g., Mplus) cannot handle the constraint in the form of Equation 12, but require that it be rephrased as an upper bound on one of the $K' + 1$ parameters. The convention adopted here is to choose (arbitrarily) the last parameter, $\beta_{jK'}$. Thus, Equation 12 is rewritten as

$$\left(1 + e^{\beta_{j0}+\beta_{jK'}}\right) \prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right) - \left(1 + e^{\beta_{j0}}\right)^{K'-1} e^{\beta_{j0}+\sum_{k=1}^{K'-1}\beta_{jk}} e^{\beta_{jK'}} > 0$$

Then,

$$\prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right) + e^{\beta_{j0}} e^{\beta_{jK'}} \prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right) - e^{\beta_{j0}} e^{\beta_{jK'}} \left(1 + e^{\beta_{j0}}\right)^{K'-1} e^{\sum_{k=1}^{K'-1}\beta_{jk}} > 0$$

$$\prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right) > e^{\beta_{jK'}} e^{\beta_{j0}} \left(\left(1 + e^{\beta_{j0}}\right)^{K'-1} e^{\sum_{k=1}^{K'-1}\beta_{jk}} - \prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right) \right)$$

$$= e^{\beta_{jK'}} e^{\beta_{j0}} \left(\prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}}\right) e^{\beta_{jk}} - \prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right) \right)$$

$$= e^{\beta_{jK'}} e^{\beta_{j0}} \left(\prod_{k=1}^{K'-1} \left(e^{\beta_{jk}} + e^{\beta_{j0}+\beta_{jk}}\right) - \prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right) \right) \quad (13)$$

Because $e^{\beta_{jk}} > 1$ for all k (due to $\beta_{jk} > 0$)

$$\prod_{k=1}^{K'-1} \left(e^{\beta_{jk}} + e^{\beta_{j0}+\beta_{jk}}\right) - \prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right) > 0$$

Thus,

$$e^{\beta_{j0}} \left(\prod_{k=1}^{K'-1} \left(e^{\beta_{jk}} + e^{\beta_{j0}+\beta_{jk}}\right) - \prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right) \right) > 0$$

must be true, and Equation 13 can be written as

$$e^{\beta_{jK'}} < \frac{\prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right)}{e^{\beta_{j0}} \left(\prod_{k=1}^{K'-1} \left(e^{\beta_{jk}} + e^{\beta_{j0}+\beta_{jk}}\right) - \prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right) \right)}$$

which implies the general expression of the upper bound, $U_{jK'}^{(1...K')}$:

$$\beta_{jK'} < U_{jK'}^{(1...K')}$$

$$= \sum_{k=1}^{K'-1} \ln\left(1 + e^{\beta_{j0}+\beta_{jk}}\right) - \ln\left(\prod_{k=1}^{K'-1} \left(e^{\beta_{jk}} + e^{\beta_{j0}+\beta_{jk}}\right) - \prod_{k=1}^{K'-1} \left(1 + e^{\beta_{j0}+\beta_{jk}}\right) \right) - \beta_{j0}$$

$$(14)$$

Based on Equation 14, the upper bound on any main-effect coefficient can be identified provided the indices of the summation and the product operator are adequately adjusted.

3.5. The Least Upper Bound on the Coefficient of the Last Main Effect

In the previous section, it was shown that the constraint on the coefficient of any interaction term can be rephrased as an upper bound on the coefficient of any main effect that constitutes the interaction term (recall that the intercept would do as well). By convention, this upper bound is usually imposed on the coefficient of the last main effect, $\beta_{jK'}$ (i.e., the coefficient of the main effect with the highest index $k = K'$). In this section, it is demonstrated that the upper bound derived from the constraint on the coefficient of the highest-order interaction term always guarantees the least upper bound—that is, all upper bounds derivable from the constraints on the coefficients of any lower-order interaction terms are larger.

Consider an item that requires the mastery of K attributes. Hence, the coefficient of the last main effect is β_{jK}. There are $\sum_{k=1}^{K-1} \binom{K-1}{k}$ interaction terms that involve the last main effect, α_K. Thus, $\sum_{k=1}^{K-1} \binom{K-1}{k}$ different upper bounds on β_{jK} can be derived. (For example, let $K = 4$; thus, there are $\sum_{k=1}^{3} \binom{3}{k} = 7$ upper bounds on β_{j4}.) It was demonstrated earlier that the least upper bound on β_{jK} can be identified by exploring the relations among the candidate upper bounds. Consider the upper bound on β_{jK}, $U_{\beta_{jK}}^{(1...K)}$, obtained from the constraint on the coefficient of the K-way interaction, $\beta_{j1...K}$,

$$
\beta_{jK} < U_{\beta_{jK}}^{(1...K)}
$$

$$
= \sum_{k=1}^{K-1} \ln\left(1 + e^{\beta_{j0}+\beta_{jk}}\right) - \ln\left(\prod_{k=1}^{K-1} \left(e^{\beta_{jk}} + e^{\beta_{j0}+\beta_{jk}}\right) - \prod_{k=1}^{K-1}\left(1 + e^{\beta_{j0}+\beta_{jk}}\right)\right) - \beta_{j0}
$$

Without loss of generality, $U_{\beta_{jK}}^{(1...K)}$ is compared to the upper bounds on β_{jK} obtainable from the constraints on the coefficients of any K'-way interaction, $\beta_{j((K-K'+1)...K)}$, where $K' = 2, \ldots, K - 1$:

$$
\beta_{jK} < U_{\beta_{jK}}^{((K-K'+1)...K)}
$$

$$
= \sum_{k=K-K'+1}^{K-1} \ln\left(1 + e^{\beta_{j0}+\beta_{jk}}\right)
$$

$$
- \ln\left(\prod_{k=K-K'+1}^{K-1} \left(e^{\beta_{jk}} + e^{\beta_{j0}+\beta_{jk}}\right) - \prod_{k=K-K'+1}^{K-1}\left(1 + e^{\beta_{j0}+\beta_{jk}}\right)\right) - \beta_{j0}
$$

The difference of any pair of upper bounds is always strictly positive:

$$
U_{\beta_{jK}}^{((K-K'+1)...K)} - U_{\beta_{jK}}^{(1...K)}
$$

$$
= \sum_{k=K-K'+1}^{K-1} \ln\left(1 + e^{\beta_{j0}+\beta_{jk}}\right)
$$

$$
- \ln\left(\prod_{k=K-K'+1}^{K-1} \left(e^{\beta_{jk}} + e^{\beta_{j0}+\beta_{jk}}\right) - \prod_{k=K-K'+1}^{K-1}\left(1 + e^{\beta_{j0}+\beta_{jk}}\right)\right) - \beta_{j0}
$$

$$
- \sum_{k=1}^{K-1} \ln\left(1 + e^{\beta_{j0}+\beta_{jk}}\right) + \ln\left(\prod_{k=1}^{K-1} \left(e^{\beta_{jk}} + e^{\beta_{j0}+\beta_{jk}}\right) - \prod_{k=1}^{K-1}\left(1 + e^{\beta_{j0}+\beta_{jk}}\right)\right) + \beta_{j0}
$$

$$= \ln\left(\prod_{k=1}^{K-1}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)-\prod_{k=1}^{K-1}\left(1+e^{\beta_{j0}+\beta_{jk}}\right)\right)$$

$$-\ln\left(\prod_{k=K-K'+1}^{K-1}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)-\prod_{k=K-K'+1}^{K-1}\left(1+e^{\beta_{j0}+\beta_{jk}}\right)\right)-\sum_{k=1}^{K-K'}\ln\left(1+e^{\beta_{j0}+\beta_{jk}}\right)$$

$$= \ln\left(\prod_{k=1}^{K-K'}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)\prod_{k=K-K'+1}^{K-1}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)\right.$$

$$\left.-\prod_{k=1}^{K-K'}\left(1+e^{\beta_{j0}+\beta_{jk}}\right)\prod_{k=K-K'+1}^{K-1}\left(1+e^{\beta_{j0}+\beta_{jk}}\right)\right)$$

$$-\ln\left(\prod_{k=K-K'+1}^{K-1}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)-\prod_{k=K-K'+1}^{K-1}\left(1+e^{\beta_{j0}+\beta_{jk}}\right)\right)-\sum_{k=1}^{K-K'}\ln\left(1+e^{\beta_{j0}+\beta_{jk}}\right)$$

$$> \ln\left(\prod_{k=1}^{K-K'}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)\prod_{k=K-K'+1}^{K-1}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)\right.$$

$$\left.-\prod_{k=1}^{K-K'}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)\prod_{k=K-K'+1}^{K-1}\left(1+e^{\beta_{j0}+\beta_{jk}}\right)\right)$$

$$-\ln\left(\prod_{k=K-K'+1}^{K-1}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)-\prod_{k=K-K'+1}^{K-1}\left(1+e^{\beta_{j0}+\beta_{jk}}\right)\right)-\sum_{k=1}^{K-K'}\ln\left(1+e^{\beta_{j0}+\beta_{jk}}\right)$$

$$= \ln\left(\prod_{k=1}^{K-K'}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)\left(\prod_{k=K-K'+1}^{K-1}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)-\prod_{k=K-K'+1}^{K-1}\left(1+e^{\beta_{j0}+\beta_{jk}}\right)\right)\right)$$

$$-\ln\left(\prod_{k=K-K'+1}^{K-1}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)-\prod_{k=K-K'+1}^{K-1}\left(1+e^{\beta_{j0}+\beta_{jk}}\right)\right)-\sum_{k=1}^{K-K'}\ln\left(1+e^{\beta_{j0}+\beta_{jk}}\right)$$

$$= \ln\left(\prod_{k=1}^{K-K'}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)\right)+\ln\left(\prod_{k=K-K'+1}^{K-1}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)\right.$$

$$\left.-\prod_{k=K-K'+1}^{K-1}\left(1+e^{\beta_{j0}+\beta_{jk}}\right)\right)$$

$$-\ln\left(\prod_{k=K-K'+1}^{K-1}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)-\prod_{k=K-K'+1}^{K-1}\left(1+e^{\beta_{j0}+\beta_{jk}}\right)\right)-\sum_{k=1}^{K-K'}\ln\left(1+e^{\beta_{j0}+\beta_{jk}}\right)$$

$$= \ln\left(\prod_{k=1}^{K-K'}\left(e^{\beta_{jk}}+e^{\beta_{j0}+\beta_{jk}}\right)\right)-\sum_{k=1}^{K-K'}\ln\left(1+e^{\beta_{j0}+\beta_{jk}}\right)$$

$$> 0 \quad \text{due to } e^{\beta_{jk}}>1$$

(After minor adjustments of the indices of the product and the summation operators, this result generalizes to any other K'-way interaction that involves the main effect α_K.) Among all upper

bounds on β_{jK} that can be derived from the constraint on the coefficient of any interaction term involving the main effect α_K the upper bound obtainable from the coefficient of the highest-order interaction term is always the least upper bound.

4. Practical Illustration: Fitting the Reduced RUM Using the LCA Routine in `Mplus`

To illustrate the practical application of the theoretical results on the reparameterization of the Reduced RUM as a logit model, two synthetic data sets and a real-world data set on the "Examination for the Certificate of Proficiency in English" (ECPE) were analyzed. The model was estimated using the EM algorithm implemented in the LCA routine in `Mplus`—to provide a benchmark, all data sets were also fitted with the MCMC routine available in `OpenBUGS`. Parts of the `Mplus` syntax for implementing the parameterization of the logistic function and the associated parameter constraints can be found in the appendices (as general references on how to write `Mplus` command files for fitting CDMs consult Rupp et al., 2010; Templin & Hoffman, 2013).

4.1. The Synthetic Data Sets

Two data sets were simulated each containing the responses of $N = 3000$ examinees to $J = 30$ items conforming to the Reduced RUM involving $K = 3$ and $K = 4$ attributes, respectively. The examinees' attribute profiles were generated based on the multivariate normal threshold model (Chiu, Douglas, & Li, 2009). Each attribute profile was linked to a latent continuous ability vector, $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iK})^T \sim \mathcal{N}_K (\mathbf{0}, \boldsymbol{\Sigma})$, with values along the main diagonal of $\boldsymbol{\Sigma}$ equal to 1.00 and off-diagonal entries set to 0.50. The $\boldsymbol{\theta}_i$ were randomly sampled and their components dichotomized according to

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

resulting in the attribute profile $\boldsymbol{\alpha}_i$.

For all items, the baseline parameter was fixed at $\pi_j^* = 0.90$; the penalty parameters, r_{jk}^*, were all set to 0.60. For $K = 3$ attributes, the Q-matrix replicated each of the single and two-attribute item profiles four times; the three-attribute item profiles were replicated six times. For $K = 4$ attributes, the Q-matrix replicated each of the feasible item-attribute profiles twice.

4.2. Results: The Synthetic Data Set Involving $K = 3$ Attributes

On a machine with 8 GB RAM, 2.30 GHz Intel Core processor, and 64-bit OS, the EM algorithm (`Mplus`) used 2,054 seconds of CPU time; MCMC (`OpenBUGS`) required 16,109 seconds of CPU time. (For the simulated data, to secure high accuracy of the MCMC parameter estimates, 8000 and 10000 were chosen as the length of the burn-in and the number of updates, respectively.) None of the estimates of the (traditional) parameters of the Reduced RUM obtained from the EM algorithm deviated from the known true model parameters by more than 0.080; for the MCMC estimates, the maximum deviation from the known true model parameters was 0.082.

The parameter estimates of the Reduced RUM in traditional form, $\hat{\pi}_j^*, \hat{r}_{jk}^*, k = 1, 2, 3$, were retrieved from the EM-based estimates of the parameters of the logit model according to (see p. 9)

$$\pi_j^* = e^{\delta_{j0} + \sum_{k=1}^K \delta_{jk} q_{jk}} = e^{\beta_{j0} - \ln(1 + e^{\beta_{j0}}) + \sum_{k=1}^K \left(\beta_{jk} q_{jk} - \ln\left(1 + e^{\beta_{j0} + \beta_{jk}}\right) q_{jk} + \ln\left(1 + e^{\beta_{j0}}\right) q_{jk}\right)}$$

$$= \frac{e^{\beta_{j0} + \sum_{k=1}^K \beta_{jk} q_{jk}} \prod_{k=1}^K \left(1 + e^{\beta_{j0}}\right)^{q_{jk}}}{\left(1 + e^{\beta_{j0}}\right) \prod_{k=1}^K \left(1 + e^{\beta_{j0} + \beta_{jk}}\right)^{q_{jk}}}$$

and

$$r_{jk}^* = e^{-\delta_{jk}} = e^{-\beta_{jk}+\ln\left(1+e^{\beta_{j0}+\beta_{jk}}\right)-\ln\left(1+e^{\beta_{j0}}\right)} = \frac{1+e^{\beta_{j0}+\beta_{jk}}}{e^{\beta_{jk}}\left(1+e^{\beta_{j0}}\right)}$$

As an illustration of these conversions, consider item 7. The coefficient estimates $\hat{\beta}_{70} = -1.397$, $\hat{\beta}_{71} = 0.625$, $\hat{\beta}_{72} = 0.741$, $\hat{\beta}_{73} = 0.668$, $\hat{\beta}_{712} = 0.209$, $\hat{\beta}_{713} = 0.179$, $\hat{\beta}_{723} = 0.232$, and $\hat{\beta}_{7123} = 0.868$ were given in the Mplus output (not reported here). The estimate $\hat{\pi}_7^*$ of item 7 was then computed as

$$\hat{\pi}_7^* = \frac{e^{-1.397+0.625+0.741+0.668}(1+e^{-1.397})^2}{(1+e^{-1.397+0.625})(1+e^{-1.397+0.741})(1+e^{-1.397+0.668})} = 0.8935971$$

The estimates of r_{71}^*, r_{72}^*, and r_{73}^* were

$$\hat{r}_{71}^* = \frac{1+e^{-1.397+0.625}}{e^{0.625}(1+e^{-1.397})} = 0.6274156$$

$$\hat{r}_{72}^* = \frac{1+e^{-1.397+0.741}}{e^{0.741}(1+e^{-1.397})} = 0.5804160$$

$$\hat{r}_{73}^* = \frac{1+e^{-1.397+0.668}}{e^{0.668}(1+e^{-1.397})} = 0.6093545$$

The conversion of the parameter estimates can be carried out directly in Mplus; the syntax for item 7 is provided in Appendix 1. As an aside, the estimates of the β-coefficients can also be used to verify that the coefficients of the interaction terms are functions of the parameters of the constituting main effects—as an example, consider β_{712}:

$$\hat{\beta}_{712} = \ln\left(\frac{1+e^{-1.397}}{(1+e^{-1.397+0.625})(1+e^{-1.397+0.741})-(1+e^{-1.397})e^{-1.397+0.625+0.741}}\right)$$
$$= 0.2095423$$

Fitting the data also involved modeling the relation between the latent ability dimensions underlying the attribute profiles, often referred to as the higher-order structure among the attributes. Rupp et al. (2010) discussed several models that can be used for this purpose. In the simulation studies reported here, the saturated log-linear model was used. (For technical details, consult Rupp et al., 2010, Chaps. 8 and 9; the Mplus syntax is provided in Appendix 1—when using MCMC (OpenBUGS), the relation between the latent traits was analyzed based on the multivariate normal threshold model. It can be shown that the two approaches are mathematically equivalent.) The parameter estimates of the saturated log-linear model were G_0 = 0.007, G_11 = −1.014, G_12 = −1.052, G_13 = −1.130, G_212 = 1.093, G_213 = 1.032, G_223 = 1.012, and G_3123 = 0.053. Based on these estimates, the proportions of examinees in each proficiency class were computed that were then used to estimate the tetrachoric correlations between pairs of attributes (the known true tetrachoric correlations are given in parentheses): $(\alpha_1, \alpha_2) = 0.498$ (0.488); $(\alpha_1, \alpha_3) = 0.484$ (0.496); and $(\alpha_2, \alpha_3) = 0.480$ (0.489).

4.3. Results: The Synthetic Data Set Involving $K = 4$ Attributes

The CPU times for the EM algorithm (`Mplus`) were 7430 and 23,762 seconds for MCMC (`OpenBUGS`) observed on a machine with 8 GB RAM, 2.30 GHz Intel Core processor, and 64-bit OS. (18,000 and 10,000 were chosen as the length of the burn-in and the number of updates, respectively.) None of the estimates of the (traditional) parameters of the Reduced RUM obtained from the EM algorithm deviated from the known true model parameters by more than 0.111; for the MCMC estimates, the maximum deviation from the known true model parameters was 0.186. Appendix 2 provides the `Mplus` syntax for the main and interaction effects of the most complex test item 15, $\mathbf{q}_{15} = (1111)^T$, and the associated constraints.

Modeling the higher-order structure among the $K = 4$ attributes resulted in the following parameter estimates of the saturated log-linear model: `G_0` $= 0.066$, `G_11` $= -1.389$, `G_12` $= -1.539$, `G_13` $= -1.542$, `G_14` $= -1.567$, `G_212` $= 1.181$, `G_213` $= 1.279$, `G_214` $= 1.216$, `G_223` $= 1.030$, `G_224` $= 1.198$, `G_234` $= 1.268$, `G_3123` $= -0.338$, `G_3124` $= -0.649$, `G_3134` $= -0.836$, `G_3234` $= -0.124$, and `G_41234` $= 0.745$. The estimated tetrachoric correlations between the attribute pairs (α_1, α_2), (α_1, α_3), (α_1, α_4), (α_2, α_3), (α_2, α_4), and (α_3, α_4) were thus 0.497 (0.523), 0.498 (0.498), 0.449 (0.462), 0.538 (0.515), 0.530 (0.537), and 0.524 (0.500), respectively (the known true tetrachoric correlations are given in parentheses).

4.4. The "Examination for the Certificate of Proficiency in English (ECPE)" Data

The "Examination for the Certificate of Proficiency in English (ECPE)" for non-native speakers was developed by the English Language Institute at the University of Michigan. The test is administered annually to examinees from Africa, Asia, Europe, and Latin America. The data used here are a subset from the ECPE grammar section from a single year and have been previously analyzed by Buck and Tatsuoka (1998), Feng et al. (2014), Henson and Templin (2007), Templin and Bradshaw (2014), and Templin and Hoffman (2013). Responses to $J = 28$ items were collected from $N = 2922$ examinees. The test involved $K = 3$ attributes ($\alpha_1 =$ lexical skills; $\alpha_2 =$ morphosyntactic skills; and $\alpha_3 =$ cohesive skills); but none of the items required the mastery of all $K = 3$ attributes (the complete Q-matrix is given in Templin & Hoffman, 2013).

4.5. Results: The "Examination for the Certificate of Proficiency in English (ECPE)" Data

On a machine with 8 GB RAM, 2.30 GHz Intel Core processor, and 64-bit OS, the CPU times for the EM algorithm (`Mplus`) were 568 seconds and 8,319 seconds for MCMC (`OpenBUGS`). The data were first fitted without including the higher-order attribute structure using the default setting of 0.50 for the item parameter starting values. `Mplus` did not converge and terminated prematurely with an error message recommending to change the starting values. Inspection of the (premature) parameter estimates revealed several problems: (a) the estimate of the main effect parameter for item 8, $\hat{\beta}_{83}$, was negative, which is a violation of the positivity constraint; (b) all parameter estimates of item 22 were unusually large in magnitude. After increasing the `Mplus` starting value of the main effect of item 8 from the default setting of 0.50–0.80, and decreasing the `Mplus` starting value of the intercept parameter of item 22 from the default value of 0.50–0.10, the estimation process converged properly. Next, the estimation of the measurement model was augmented by fitting the higher-order attribute structure with the saturated log-linear model. The algorithm again terminated prematurely suggesting that the `Mplus` starting values be changed. Indeed, the estimates of the coefficients of the main effects in the saturated log-linear model were unusually large. After decreasing their starting values from the default setting of 0.50 to -1.00, the algorithm converged properly. The `Mplus` item parameter estimates of the logit model were converted to the traditional Reduced RUM parameter estimates. They are presented in Table 1 together with the MCMC estimates. With minor exceptions, the EM and MCMC parameter estimates are more or less identical.

TABLE 1.
The ECPE data: estimates of the traditional item parameters of the Reduced RUM obtained from EM (`Mplus`) and MCMC (`OpenBUGS`).

Item	Mplus				MCMC			
	π^*	r_1^*	r_2^*	r_3^*	π^*	r_1^*	r_2^*	r_3^*
1	0.929	0.876	0.852	–	0.920	0.885	0.792	–
2	0.907	–	0.815	–	0.893	–	0.784	–
3	0.784	0.640	–	0.824	0.781	0.633	–	0.857
4	0.825	–	–	0.562	0.828	–	–	0.570
5	0.957	–	–	0.779	0.957	–	–	0.784
6	0.927	–	–	0.760	0.928	–	–	0.765
7	0.943	0.738	–	0.704	0.940	0.750	–	0.696
8	0.968	–	0.843	–	0.956	–	0.814	–
9	0.788	–	–	0.672	0.790	–	–	0.673
10	0.892	0.576	–	–	0.892	0.574	–	–
11	0.926	0.769	–	0.695	0.925	0.771	–	0.702
12	0.733	0.527	–	0.362	0.731	0.539	–	0.366
13	0.908	0.727	–	–	0.907	0.728	–	–
14	0.826	0.658	–	–	0.824	0.659	–	–
15	0.958	–	–	0.761	0.958	–	–	0.766
16	0.909	0.753	–	0.713	0.906	0.761	–	0.714
17	0.943	–	0.921	0.918	0.938	–	0.896	0.928
18	0.910	–	–	0.785	0.910	–	–	0.793
19	0.839	–	–	0.537	0.842	–	–	0.544
20	0.760	0.501	–	0.510	0.759	0.507	–	0.516
21	0.918	0.853	–	0.699	0.916	0.862	–	0.700
22	0.796	–	–	0.370	0.800	–	–	0.380
23	0.939	–	0.704	–	0.916	–	0.655	–
24	0.701	–	0.481	–	0.659	–	0.430	–
25	0.775	0.674	–	–	0.775	0.674	–	–
26	0.783	–	–	0.689	0.784	–	–	0.695
27	0.695	0.421	–	–	0.694	0.424	–	–
28	0.909	–	–	0.701	0.910	–	–	0.708

The `Mplus` estimates of the tetrachoric correlations between the attribute pairs (lexical, morphosyntactic), (lexical, cohesive), and (morphosyntactic, cohesive) were 0.865, 0.786, and 0.911, respectively, which are lower than the estimates obtained from MCMC: 0.915, 0.887, and 0.913.

5. Discussion

The Reduced RUM has been frequently studied in simulations and applications to real-world data sets (e.g., Feng et al., 2014; Henson & Douglas, 2005; Henson & Templin, 2007; Henson, Roussos, Douglas, & He, 2008; Henson, Templin, & Douglas, 2007; Kim, 2011; Liu, Douglas, & Henson, 2009; Templin, Henson, Templin, & Roussos, 2008). Researchers have appreciated the flexibility of the Reduced RUM in modeling the probability of correct item responses for different attribute profiles. However, this flexibility comes at the cost of a "significant degree of complexity" of the estimation process (Feng et al., 2014, p. 138). In fact, with the exception of Feng et al. (2014), the estimation method of choice in the studies referenced above was MCMC. But MCMC requires advanced technical skills so that its usefulness is likely restricted to researchers with a solid background in statistics.

To date, the options for educational researchers, who wish to use the Reduced RUM in their testing programs and empirical research, but do not feel comfortable writing their own code, are still rather limited. Feng et al. (2014) recently reported the implementation of the EM algorithm for estimating the Reduced RUM as a routine in R. However, this routine is not (yet) publicly available. Alternatively, educational researchers can use a commercial package that offers an implementation of the EM algorithm for fitting (constrained) latent class models. However, using an LCA routine for fitting the Reduced RUM requires that it be re-expressed as a logit model, with constraints imposed on the parameters of the logistic function—but these were only known for models involving at most $K = 2$ attributes. In this article, the general parameterization of the Reduced RUM as a logit model involving any number of attributes and the associated parameter constraints were derived. Thus, educational researchers and practitioners can now use the EM algorithm in the LCA routines of commercial software packages like Latent GOLD and Mplus for fitting the Reduced RUM to data sets with a realistic number of attributes. Several aspects concerning the theoretical and practical implications of this result remain to be addressed.

In response to the "lack of available software for researchers and practitioners" (Templin & Hoffman, 2013, p. 63) of advanced, complex CDMs, Templin and Hoffman (2013) prepared a tutorial for fitting the (saturated) LCDM with Mplus. (At least equally important, and an immense aid to the practitioner, Templin and Hoffman (2013) direct the reader to a SAS macro written by Templin for generating the complex Mplus syntax of the parameter constraints of the LCDM automatically.) The saturated LCDM subsumes many of the familiar, recognizable CDMs that can be derived by imposing appropriate constraints on the model parameters of the saturated LCDM—among them the Reduced RUM (Henson et al., 2009; Templin & Bradshaw, 2014). So, why is a separate treatment of the technical aspects of the Reduced RUM and their software implementation needed in the first place? Casual inspection may easily deceive one to believe that the saturated LCDM is the Reduced RUM because both models contain identical main and interaction terms. However, the constraints on the parameters of the Reduced RUM are different from those of the saturated LCDM; they represent a distinct view on how the mastery of attributes and the probability of a correct response are related. The technical level of these constraints is nontrivial requiring tailored syntax to implement them in Mplus.

In comparison with the parameterization of the Reduced RUM as a logit model, the traditional parameterization of the Reduced RUM has two immediate practical advantages. First, the parameters, π_j^* and r_{jk}^*, are bounded by 0 and 1. They are well-defined and have a direct and meaningful interpretation as probabilities that "modulate" the overall probability of a correct item response (which itself is the result of the complex interplay between attributes required for an item and those mastered by an examinee). In contrast, the parameters of the Reduced RUM as a logit model are not readily interpretable as probabilities; although they quantify the contribution of the attributes (and their interactions) to a correct item response—but the logit scale is not as intuitively accessible as the probability scale. Second, the well-defined meaning of the traditional parameters of the Reduced RUM allows a researcher to detect ill-devised items immediately. Recall that all parameters must be bounded by 0 and 1. High-quality items usually have a large π_j^* and r_{jk}^* that take on only moderate values (Henson & Templin, 2007). This information cannot be gleaned from the parameters of the Reduced RUM as a logit model.

Finally, what recommendations can be made to educational practitioners? First, the CPU times observed with the few examples reported here seem to suggest to prefer the EM algorithm in the Mplus LCA routine over MCMC (OpenBUGS). Although, neither the EM nor the MCMC run times make conducting large-scale simulations currently a viable option. On the other hand, having access to an operational EM algorithm for analyzing a given data set with the Reduced RUM involving any number of attributes—although potentially time consuming—is certainly a big improvement.

Second, as might be gathered from the computational applications reported here, the specification of the starting values for the parameter estimates in Mplus appears to be more an art than a skill. The recommendation is to try first the default settings (i.e., 0.50 for all parameters). If Mplus encounters difficulties, then it will very likely terminate with a nondescript error message recommending a change of the starting values. The few applications run here were relatively easy to fix by inspecting the (premature) parameter estimates provided by Mplus and adjusting the setting of the starting values accordingly. However, it is not unlikely that this might not work for other applications. In such a situation, the general recommendation is to contact the Mplus technical support.

Third, composing and debugging Mplus syntax for fitting CDMs is a tedious and time-consuming undertaking. Therefore, a function written in R has been made available to researchers to aid them in writing their own Mplus code. The only input required to the R function are the data and the Q-matrix underlying the test; the complete Mplus syntax for several CDMs—including the Reduced RUM—is then generated automatically. This R function is available from the authors upon request.

Appendix 1: Mplus Syntax: The Synthetic Data Set Involving $K = 3$ Attributes

Model Definition and Constraints

The notation follows Rupp et al. (2010) that may serve as general reference on writing Mplus code for fitting CDMs. The preamble of the Mplus command file specifies the number of $J = 30$ test items. In the subsequent MODEL declaration, for each of the 8 proficiency classes separately, all $J = 30$ test items must be specified (for details, consult Templin & Hoffman's, 2013, comprehensive tutorial). The MODEL declaration must also contain the syntax for computing the means of the proficiency classes that are needed later on for modeling the higher-order structure among the attributes (more on this below):

```
%OVERALL%
[c#1] (m1); ! attribute profile alpha = (000)
[c#2] (m2); ! attribute profile alpha = (100)
[c#3] (m3); ! attribute profile alpha = (010)

! The code for proficiency classes 4, alpha=(001),
! to 8, alpha=(111), has been omitted.
```

In the MODEL CONSTRAINT section of the syntax, the constraints on the model parameters must be defined for each item individually. An item parameter is denoted by "Lj_ab" in the Mplus command language, where "j" is the item number; "a" specifies whether the parameter refers to a main effect ($a = 1$) or an interaction term of order $a > 1$; and "b" indicates the attribute(s) involved. As an example, consider item 1 that requires the mastery of only the first attribute: $\mathbf{q}_1 = (100)^T$. The IRF of item 1 is

$$P(Y_1 = 1 \mid \boldsymbol{\alpha}) = \frac{e^{\beta_{10}+\beta_{11}\alpha_1}}{1 + e^{\beta_{10}+\beta_{11}\alpha_1}}$$

The intercept parameter, β_{10}, is written as L1_0; the coefficient of the main effect, β_{11}, is written as L1_11. Recall that β_{10} is unconstrained; hence, the only parameter constraint is $\beta_{11} > 0$, which is written in the Mplus command language as L1_11 > 0. As a more complex example, consider item 7 that requires the mastery of all $K = 3$ attributes; thus, $\mathbf{q}_7 = (111)^T$. The IRF of item 7 is

$$P(Y_7 = 1 \mid \boldsymbol{\alpha}) = \frac{e^{\beta_{70}+\beta_{71}\alpha_1+\beta_{72}\alpha_2+\beta_{73}\alpha_3+\beta_{712}\alpha_1\alpha_2+\beta_{713}\alpha_1\alpha_3+\beta_{723}\alpha_2\alpha_3+\beta_{7123}\alpha_1\alpha_2\alpha_3}}{1 + e^{\beta_{70}+\beta_{71}\alpha_1+\beta_{72}\alpha_2+\beta_{73}\alpha_3+\beta_{712}\alpha_1\alpha_2+\beta_{713}\alpha_1\alpha_3+\beta_{723}\alpha_2\alpha_3+\beta_{7123}\alpha_1\alpha_2\alpha_3}}$$

$$\text{subject to:} (1)\ 0 < \beta_{7k} \qquad k = 1, 2$$

$$(2)\ 0 < \beta_{73} < \ln\left(\left(1 + e^{\beta_{70}+\beta_{71}}\right)\left(1 + e^{\beta_{70}+\beta_{72}}\right)\right)$$

$$- \ln\left(\left(1 + e^{\beta_{70}}\right)^2 e^{\beta_{71}+\beta_{72}} - \left(1 + e^{\beta_{70}+\beta_{71}}\right)\left(1 + e^{\beta_{70}+\beta_{72}}\right)\right) - \beta_{70}$$

where

$$\beta_{712} = \ln\left(\frac{1 + e^{\beta_{70}}}{\left(1 + e^{\beta_{70}+\beta_{71}}\right)\left(1 + e^{\beta_{70}+\beta_{72}}\right) - \left(1 + e^{\beta_{70}}\right)e^{\beta_{70}+\beta_{71}+\beta_{72}}}\right)$$

$$\beta_{713} = \ln\left(\frac{1 + e^{\beta_{70}}}{\left(1 + e^{\beta_{70}+\beta_{71}}\right)\left(1 + e^{\beta_{70}+\beta_{73}}\right) - \left(1 + e^{\beta_{70}}\right)e^{\beta_{70}+\beta_{71}+\beta_{73}}}\right)$$

$$\beta_{723} = \ln\left(\frac{1 + e^{\beta_{70}}}{\left(1 + e^{\beta_{70}+\beta_{72}}\right)\left(1 + e^{\beta_{70}+\beta_{73}}\right) - \left(1 + e^{\beta_{70}}\right)e^{\beta_{70}+\beta_{72}+\beta_{73}}}\right)$$

$$\beta_{7123} = \ln\left(\frac{\left(1 + e^{\beta_{70}}\right)^2}{\left(1 + e^{\beta_{70}+\beta_{71}}\right)\left(1 + e^{\beta_{70}+\beta_{72}}\right)\left(1 + e^{\beta_{70}+\beta_{73}}\right) - \left(1 + e^{\beta_{70}}\right)^2 e^{\beta_{70}+\beta_{71}+\beta_{72}+\beta_{73}}}\right)$$

$$- \left(\beta_{713} + \beta_{712} + \beta_{723}\right)$$

All model parameters in the IRF of item 7 are defined in the first line of the MODEL CONSTRAINT section, followed by the specification of the model parameters that characterize the different proficiency classes. Third, the constraints on the three main effects, β_{71} (L7_11 in Mplus syntax), β_{72} (L7_12), and β_{73} (L7_13), that guarantee monotonicity are defined. (In addition, note the complex syntax of Constraint 2 that defines the upper bound on the coefficient of the last main effect, β_{73} (L7_13).) Finally, the equations of the interaction terms of item 7 must be implemented. (Recall that the parameters of all interaction terms are functions of the parameters of the constituting main effects.)

```
MODEL CONSTRAINT:

! The code for items 1 to 6 has been omitted.

NEW(L7_0 L7_11 L7_12 L7_13 L7_212 L7_213 L7_223 L7_3123);
T7_1 = -(L7_0);
T7_2 = -(L7_0+L7_11);
T7_3 = -(L7_0+L7_12);
T7_4 = -(L7_0+L7_13);
T7_5 = -(L7_0+L7_11+L7_12+L7_212);
T7_6 = -(L7_0+L7_11+L7_13+L7_213);
T7_7 = -(L7_0+L7_12+L7_13+L7_223);
T7_8 = -(L7_0+L7_11+L7_12+L7_13+L7_212+L7_213+L7_223+L7_3123);
L7_11>0;
L7_12>0;
L7_13>0;
L7_13<log((1+exp(L7_0+L7_11)))+log((1+exp(L7_0+L7_12)))
    -log(((1+exp(L7_0))^2)*exp(L7_11+L7_12)
    -(1+exp(L7_0+L7_11))*(1+exp(L7_0+L7_12)))-L7_0;
```

```
L7_212=log(((1+exp(L7_0))/((1+exp(L7_0+L7_11))*(1+exp(L7_0+L7_12))
        -(1+exp(L7_0))*exp(L7_0+L7_11+L7_12)));
L7_213=log(((1+exp(L7_0))/((1+exp(L7_0+L7_11))*(1+exp(L7_0+L7_13))
        -(1+exp(L7_0))*exp(L7_0+L7_11+L7_13)));
L7_223=log(((1+exp(L7_0))/((1+exp(L7_0+L7_12))*(1+exp(L7_0+L7_13))
        -(1+exp(L7_0))*exp(L7_0+L7_12+L7_13)));
L7_3123=log(((1+exp(L7_0))^2)
        /((1+exp(L7_0+L7_11))*(1+exp(L7_0+L7_12))*(1+exp(L7_0+L7_13))
        -((1+exp(L7_0))^2)*exp(L7_0+L7_11+L7_12+L7_13)))
        -(L7_212+L7_213+L7_223);

! The code for items 8 to 30 has been omitted.
```

Converting the Parameter Estimates from the Logit to the Traditional Form of the Reduced RUM

The parameter estimates of the Reduced RUM in traditional form are retrieved from the EM-based parameter estimates obtained for the logit model. As an example of how to perform the parameter conversion, consider item 7 (see also p. 9):

$$\pi_7^* = \frac{e^{\beta_{70}+\beta_{71}+\beta_{72}+\beta_{73}}(1+e^{\beta_{70}})^2}{(1+e^{\beta_{70}+\beta_{71}})(1+e^{\beta_{70}+\beta_{72}})(1+e^{\beta_{70}+\beta_{73}})}$$

$$r_{71}^* = \frac{1+e^{\beta_{70}+\beta_{71}}}{e^{\beta_{71}}(1+e^{\beta_{70}})}$$

$$r_{72}^* = \frac{1+e^{\beta_{70}+\beta_{72}}}{e^{\beta_{72}}(1+e^{\beta_{70}})}$$

$$r_{73}^* = \frac{1+e^{\beta_{70}+\beta_{73}}}{e^{\beta_{73}}(1+e^{\beta_{70}})}$$

The conversion of the parameter estimates can be carried out directly in `Mplus`:

```
! The code for items 1 to 6 has been omitted.

NEW(pi7 r7_1 r7_2 r7_3);
pi7=exp(L7_0+L7_11+L7_12+L7_13)*(1+exp(L7_0))^2/
((1+exp(L7_0+L7_11))*(1+exp(L7_0+L7_12))*(1+exp(L7_0+L7_13)));
r7_1=(1+exp(L7_0+L7_11))/(exp(L7_11)*(1+exp(L7_0)));
r7_2=(1+exp(L7_0+L7_12))/(exp(L7_12)*(1+exp(L7_0)));
r7_3=(1+exp(L7_0+L7_13))/(exp(L7_13)*(1+exp(L7_0)));

! The code for items 8 to 30 has been omitted.
```

Modeling the Higher-Order Structure Among Attributes

As discussed in Rupp et al. (2010), several models can be used for the higher-order structure among attributes. In the simulation studies here, the saturated log-linear model is used. Following the notation in Rupp et al. (2010), the parameters are denoted as `G_ab`, where "a" indicates if the parameter refers to a main effect ($a = 1$) or an interaction term of order $a > 1$; "b" specifies the attribute(s) involved.

```
NEW(G_0 G_11 G_12 G_13 G_212 G_213 G_223 G_3123);
m1 = -(G_11+G_12+G_13+G_212+G_213+G_223+G_3123);
```

```
m2 = -(G_12+G_13+G_212+G_213+G_223+G_3123);
m3 = -(G_11+G_13+G_212+G_213+G_223+G_3123);
m4 = -(G_11+G_12+G_212+G_213+G_223+G_3123);
m5 = -(G_13+G_213+G_223+G_3123);
m6 = -(G_12+G_212+G_223+G_3123);
m7 = -(G_11+G_212+G_213+G_3123);
G_0 = -(G_11+G_12+G_13+G_212+G_213+G_223+G_3123);
```

Appendix 2: `Mplus` Syntax: The Synthetic Data Set Involving $K = 4$ Attributes

The most complex test item 15 requiring the mastery of $K = 4$ attributes is used as an example here for the implementation of the parameters of the logistic function and their associated constraints. First, the constraints on the main effect parameters, β_{151}, β_{152}, β_{153}, and β_{154}, that guarantee monotonicity must be defined. The least upper bound on the parameter associated with the last main effect, β_{154} (`L15_14` in `Mplus` syntax), is derived from Equation 14 as

$$
\begin{aligned}
\beta_{154} &< \sum_{k=1}^{4-1} \ln(1 + e^{\beta_{150}+\beta_{15k}}) - \ln\left(\prod_{k=1}^{4-1}(e^{\beta_{15k}} + e^{\beta_{150}+\beta_{15k}}) - \prod_{k=1}^{4-1}\left(1 + e^{\beta_{150}+\beta_{jk}}\right)\right) - \beta_{150} \\
&= \ln\left(1 + e^{\beta_{150}+\beta_{151}}\right) + \ln\left(1 + e^{\beta_{150}+\beta_{152}}\right) + \ln\left(1 + e^{\beta_{150}+\beta_{153}}\right) \\
&\quad - \ln\left(\left(1+\beta_{150}\right)^3 e^{\beta_{151}+\beta_{152}+\beta_{153}} - \left(1+e^{\beta_{150}+\beta_{151}}\right)\left(1+e^{\beta_{150}+\beta_{152}}\right)\left(1+e^{\beta_{150}+\beta_{153}}\right)\right) - \beta_{150}
\end{aligned}
$$

As an example of a rather complex model parameter, the coefficient of the four-way interaction term of item 15, β_{151234}, is given (see Equation 11):

$$
\begin{aligned}
\beta_{151234} &= \ln\left(\frac{(1 + e^{\beta_{150}})^{4-1}}{\prod_{k=1}^{4}(1 + e^{\beta_{150}+\beta_{15k}}) - (1 + e^{\beta_{150}})^{4-1}e^{\beta_{150}+\sum_{k=1}^{4}\beta_{15k}}}\right) \\
&\quad - \left(\sum_{k'=k+1}^{4}\sum_{k=1}^{3}\beta_{15kk'} + \sum_{k_3=k_2+1}^{4}\sum_{k_2=k_1+1}^{3}\sum_{k_1=1}^{2}\beta_{15k_1k_2k_3}\right) \\
&= \ln\left(\frac{(1 + e^{\beta_{150}})^3}{(1 + e^{\beta_{150}+\beta_{151}})\cdots(1 + e^{\beta_{150}+\beta_{154}}) - (1 + e^{\beta_{150}})^3 e^{\beta_{150}+\beta_{151}+\beta_{152}+\beta_{153}+\beta_{154}}}\right) \\
&\quad - (\beta_{1512}+\beta_{1513}+\beta_{1514}+\beta_{1523}+\beta_{1524}+\beta_{1534}+\beta_{15123}+\beta_{15124}+\beta_{15134}+\beta_{15234})
\end{aligned}
$$

Here is the complete `Mplus` syntax for item 15:

```
NEW(L15_0 L15_11 L15_12 L15_13 L15_14 L15_212 L15_213 L15_214 L15_223
L15_224 L15_234 L15_3123 L15_3124 L15_3134 L15_3234 L15_41234);
T15_1 = -(L15_0);
T15_2 = -(L15_0+L15_11);
T15_3 = -(L15_0+L15_12);
T15_4 = -(L15_0+L15_13);
T15_5 = -(L15_0+L15_14);
T15_6 = -(L15_0+L15_11+L15_12+L15_212);
T15_7 = -(L15_0+L15_11+L15_13+L15_213);
T15_8 = -(L15_0+L15_11+L15_14+L15_214);
T15_9 = -(L15_0+L15_12+L15_13+L15_223);
T15_10 = -(L15_0+L15_12+L15_14+L15_224);
```

```
T15_11 = -(L15_0+L15_13+L15_14+L15_234);
T15_12 = -(L15_0+L15_11+L15_12+L15_13+L15_212+L15_213+L15_223+L15_3123);
T15_13 = -(L15_0+L15_11+L15_12+L15_14+L15_212+L15_214+L15_224+L15_3124);
T15_14 = -(L15_0+L15_11+L15_13+L15_14+L15_213+L15_214+L15_234+L15_3134);
T15_15 = -(L15_0+L15_12+L15_13+L15_14+L15_223+L15_224+L15_234+L15_3234);
T15_16 = -(L15_0+L15_11+L15_12+L15_13+L15_14+L15_212+L15_213+L15_214
+L15_223+L15_224+L15_234+L15_3123+L15_3124+L15_3134+L15_3234+L15_41234);
L15_11>0;
L15_12>0;
L15_13>0;
L15_14>0;
L15_14<log((1+exp(L15_0+L15_11)))+log((1+exp(L15_0+L15_12)))
+log((1+exp(L15_0+L15_13)))
    -log(((1+exp(L15_0))^3)*exp(L15_11+L15_12+L15_13)
    -(1+exp(L15_0+L15_11))*(1+exp(L15_0+L15_12))
    *(1+exp(L15_0+L15_13)))-L15_0;
L15_212=log((1+exp(L15_0))/((1+exp(L15_0+L15_11))*(1+exp(L15_0+L15_12))
       -(1+exp(L15_0))*exp(L15_0+L15_11+L15_12)));
L15_213=log((1+exp(L15_0))/((1+exp(L15_0+L15_11))*(1+exp(L15_0+L15_13))
       -(1+exp(L15_0))*exp(L15_0+L15_11+L15_13)));
L15_214=log((1+exp(L15_0))/((1+exp(L15_0+L15_11))*(1+exp(L15_0+L15_14))
       -(1+exp(L15_0))*exp(L15_0+L15_11+L15_14)));
L15_223=log((1+exp(L15_0))/((1+exp(L15_0+L15_12))*(1+exp(L15_0+L15_13))
       -(1+exp(L15_0))*exp(L15_0+L15_12+L15_13)));
L15_224=log((1+exp(L15_0))/((1+exp(L15_0+L15_12))*(1+exp(L15_0+L15_14))
       -(1+exp(L15_0))*exp(L15_0+L15_12+L15_14)));
L15_234=log((1+exp(L15_0))/((1+exp(L15_0+L15_13))*(1+exp(L15_0+L15_14))
       -(1+exp(L15_0))*exp(L15_0+L15_13+L15_14)));
L15_3123=log(((1+exp(L15_0))^2)
       /((1+exp(L15_0+L15_11))*(1+exp(L15_0+L15_12))*(1+exp(L15_0+L15_13))
       -((1+exp(L15_0))^2)*exp(L15_0+L15_11+L15_12+L15_13)))
       -(L15_212+L15_213+L15_223);
L15_3124=log(((1+exp(L15_0))^2)
       /((1+exp(L15_0+L15_11))*(1+exp(L15_0+L15_12))*(1+exp(L15_0+L15_14))
       -((1+exp(L15_0))^2)*exp(L15_0+L15_11+L15_12+L15_14)))
       -(L15_212+L15_214+L15_224);
L15_3134=log(((1+exp(L15_0))^2)
       /((1+exp(L15_0+L15_11))*(1+exp(L15_0+L15_13))*(1+exp(L15_0+L15_14))
       -((1+exp(L15_0))^2)*exp(L15_0+L15_11+L15_13+L15_14)))
       -(L15_213+L15_214+L15_234);
L15_3234=log(((1+exp(L15_0))^2)
       /((1+exp(L15_0+L15_12))*(1+exp(L15_0+L15_13))*(1+exp(L15_0+L15_14))
       -((1+exp(L15_0))^2)*exp(L15_0+L15_12+L15_13+L15_14)))
       -(L15_223+L15_224+L15_234);
L15_41234=log(((1+exp(L15_0))^3)
       /((1+exp(L15_0+L15_11))*(1+exp(L15_0+L15_12))*(1+exp(L15_0+L15_13))
       *(1+exp(L15_0+L15_14))
       -((1+exp(L15_0))^3)*exp(L15_0+L15_11+L15_12+L15_13+L15_14)))
       -(L15_212+L15_213+L15_214+L15_223+L15_224+L15_234+L15_3123+L15_3124
       +L15_3134+L15_3234);
```

References

Bolt, D., Chen, H., DiBello, L., Hartz, S., Henson, R., Roussos, L., et al. (2008). *The Arpeggio Suite: Software for cognitive skills diagnostic assessment (Computer software)*. St. Paul, MN: Assessment Systems.

Buck, G., & Tatsuoka, K. K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, *15*, 119–157.

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

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

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*, 179–199.

DiBello, L. V., Roussos, L. A., & Stout, W. F. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Volume 26. Psychometrics* (pp. 979–1030). Amsterdam: Elsevier.

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

Haberman, S. J., & von Davier, M. (2007). Some notes on models for cognitively based skill diagnosis. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Volume 26. Psychometrics* (pp. 1031–1038). Amsterdam: Elsevier.

Hartz, S. M. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practicality* (Doctoral dissertation). Available from ProQuest Dissertations and Theses Database (UMI No. 3044108).

Hartz, S. M., Roussos, L. A., Henson, R. A., & Templin, J. L. (2005). *The fusion model for skill diagnosis: Blending theory with practicality*, Unpublished manuscript.

Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement*, *29*, 262–277.

Henson, R., Roussos, L. A., Douglas, J., & He, X. (2008). Cognitive diagnostic attribute-level discrimination indices. *Applied Psychological Measurement*, *32*, 275–288.

Henson, R., Templin, J. L., & Douglas, J. (2007). Using efficient model based sum-scores for conducting skills diagnoses. *Journal of Educational Measurement*, *44*, 361–376.

Henson, R. A., & Templin, J. (2007, April). *Large-scale language assessment using cognitive diagnosis models*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

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

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.

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

Liu, Y., Douglas, J. A., & Henson, R. A. (2009). Testing person fit in cognitive diagnosis. *Applied Psychological Measurement*, *33*, 579–598.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique, and future directions. *Statistics in Medicine*, *28*, 3049–3067.

Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, *33*, 379–416.

Muthén, L. K., & Muthén, B. O. (1998–2011). Mplus *user's guide* (Version 6.1) (Computer software and manual). Los Angeles: Muthén & Muthén.

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

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

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconception in the pattern classification approach. *Journal of Educational Statistics*, *12*, 55–73.

Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, *79*, 317–339.

Templin, J., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus. *Educational Measurement: Issues and Practice*, *32*, 37–50.

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

Templin, J. L., Henson, R. A., Templin, S. E., & Roussos, L. A. (2008). Robustness of hierarchical modeling of skill association in cognitive diagnosis models. *Applied Psychological Measurement*, *32*, 559–574.

Vermunt, J. K., & Magidson, J. (2000). Latent GOLD's users's guide. Boston: Statistical Innovations Inc.

von Davier, M. (2005, September). *A general diagnostic model applied to language testing data* (Research report No. RR-05-16). Princeton, NJ: Educational Testing Service.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*, 287–301.

von Davier, M. (2011, September). *Equivalency of the DINA model and a constrained general diagnostic model* (Research report No. RR-11-37). Princeton, NJ: Educational Testing Service.