Assessing the Completeness of the Q-Matrix in Cognitively Diagnostic Modeling

Hans-Friedrich Köhn* Chia-Yi Chiu†

Abstract

The Q-matrix of a cognitively diagnostic test is said to be complete if it allows for the identification of all possible proficiency classes among examinees. Completeness of the Q-matrix is therefore a key requirement for any cognitively diagnostic test. However, completeness of the Q-matrix is often difficult to establish, especially, for tests with a large number of items involving multiple skills. As an additional complication, completeness is not an intrinsic property of the Q-matrix, but can only be assessed in reference to a specific cognitive diagnosis model (CDM) supposed to underly the data—that is, the Q-matrix of a given test can be complete for one model but incomplete for another. For different types of CDMs, conditions of Q-completeness are studied. Rules are derived to determine the completeness of a given Q-matrix.

Key Words: cognitive diagnosis, diagnostic classification, Q-completeness

1. Introduction

Cognitively diagnostic (CD) modeling in educational assessment (DiBello, Roussos, and Stout, 2007; Haberman and von Davier, 2007; Leighton and Gierl, 2007; Rupp, Templin, and Henson, 2010) describes an examinee's ability as a composite of specific discrete (cognitive) skills, each of which an examinee may or may not have mastered. Distinct skill profiles define classes of proficiency. Fitting educational testing data within a CD-framework seeks to estimate the parameters of the underlying model and to assign examinees to proficiency classes (i.e., estimate their individual skill profiles).

If the material constituting a knowledge domain requires K skills, then there are $M = 2^K$ distinct proficiency classes, each of which is defined by a K-dimensional binary skill profile $\boldsymbol{\alpha}_m = (\alpha_1, \alpha_2, \ldots, \alpha_k \ldots \alpha_K)'$, with $m = 1, 2, \ldots, M$. Model parameters and skill profiles are estimated from examinees' observed responses Y_j, $j = 1, 2, \ldots, J$, to the J items in the test. Each individual item itself is associated with a K-dimensional binary vector, $\mathbf{q}_j$, called item-skill profile, where $q_{jk} = 1$ if a correct answer requires mastery of the k^{th} skill, and 0 otherwise. Note that item-skill profiles consisting entirely of zeroes are inadmissible, because they correspond to items that require no skills at all. Hence, given K skills, there are at most $2^K - 1$ distinct item-skill profiles. The J item-skill profiles of a test constitute its Q-matrix, $\mathbf{Q} = \{q_{jk}\}_{(J \times K)}$, (Tatsuoka, 1985) that summarizes the constraints specifying the associations between items and skills.

The Q-matrix is an integral component of any test that is based on the CD framework. The Q-matrix must fulfill the requirement that it be complete—formally, $\boldsymbol{S}(\boldsymbol{\alpha}) = \boldsymbol{S}(\boldsymbol{\alpha}^*) \Rightarrow \boldsymbol{\alpha} = \boldsymbol{\alpha}^*$, where $\boldsymbol{S}(\boldsymbol{\alpha}) = E(\boldsymbol{Y} \mid \boldsymbol{\alpha})$ is the (conditional) expectation of item response vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_J)'$, given skill profile $\boldsymbol{\alpha}$. Verbally stated, a Q-matrix is said to be complete if it allows for the identification of all M possible proficiency classes among examinees (Chiu, Douglas, and Li, 2009). Said differently,

*University of Illinois at Urbana-Champaign, Champaign, IL

†Rutgers, The State University of New Jersey, New Brunswick, NJ

an incomplete Q-matrix does not allow for the identification of all M proficiency classes thus, risking that examinees are assigned to proficiency classes to which they do not belong. Completeness of the Q-matrix is therefore a key requirement for any CD test.

However, completeness of the Q-matrix is often difficult to establish, especially, for tests with a large number of items involving multiple skills. As an additional complication, completeness is not an intrinsic property of the Q-matrix, but can only be assessed in reference to a specific cognitive diagnosis model (CDM) supposed to underly the data—that is, the Q-matrix of a given test can be complete for one model, but incomplete for another.

This article studies aspects of Q-completeness for different CDMs and how these can be used for assessing whether a given Q-matrix is complete. The approach developed here relies on the theoretical framework of general CDMs (de la Torre, 2011; Henson, Templin, and Willse, 2009; Rupp et al., 2010; von Davier 2005, 2008).

2. Technical Background: General Cognitive Diagnosis Models

CDMs model the functional relation between skill mastery and the probability of a correct item response. The distinct parameterizations of specific CDMs reflects differences in the underlying theories on how (non-)mastery of skills affects an examinee's test performance. General CDMs allow for expressing these distinct functional relations in a unified mathematical form and parameterization. The archetypal general CDM is von Davier's (2005, 2008) General Diagnostic Model (GDM). Von Davier defined $h(\mathbf{q}_j, \boldsymbol{\alpha}_i)$ as a general function of the skill profile of item j and the skill profile $\boldsymbol{\alpha}_i$ of examinee i to allow for the flexible modeling of examinees' responses to item j. The item response function (IRF) of presumably the most popular version of von Davier's GDM is formed by the logistic function of the linear combination of all K skill main effects

$$P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i) = \frac{\exp(\beta_{j0} + \boldsymbol{\beta}_j' h(\mathbf{q}_j, \boldsymbol{\alpha}_i))}{1 + \exp(\beta_{j0} + \boldsymbol{\beta}_j' h(\mathbf{q}_j, \boldsymbol{\alpha}_i))} = \frac{\exp(\beta_{j0} + \sum_{k=1}^{K} \beta_{jk} q_{jk} \alpha_{ik})}{1 + \exp(\beta_{j0} + \sum_{k=1}^{K} \beta_{jk} q_{jk} \alpha_{ik})}$$

where q_{jk} indicates whether mastery of skill α_{ik} is required for item j. Henson et al. (2009) specified v_j as the linear combination of the K skill main effects, α_k, and all their two-way, three-way, ..., K-way interactions

$$v_j = \beta_{j0} + \sum_{k=1}^{K} \beta_{jk} q_{jk} \alpha_{ik} + \sum_{k'=k+1}^{K} \sum_{k=1}^{K-1} \beta_{j(kk')} q_{jk} q_{jk'} \alpha_{ik} \alpha_{ik'} + \cdots + \beta_{j12\ldots K} \prod_{k=1}^{K} q_{jk} \alpha_{ik}$$

and defined the IRF of a general CDM termed the Log-Linear Cognitive Diagnosis Model (LCDM) as

$$P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i) = \frac{\exp(v_j)}{1 + \exp(v_j)}$$

By imposing appropriate constraints on the β-coefficients in v_j, the IRFs of specific CDMs can be expressed as submodels of the LCDM. In addition to the logit link, de la Torre (2011) proposed the identity link, $P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i) = v_j$, and the log link, $P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i) = \exp\{v_j\}$, for constructing the IRF of a general CDM called the Generalized DINA (G-DINA) model. Note that the identity and the log link require additional constraints on the coefficients to guarantee $0 \leq P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i) \leq 1$. (For brevity, the examinee index i is henceforth omitted if the context permits.)

3. Conditions of Q-Completeness

Recall that completeness of the Q-matrix must be determined in reference to a particular CDM; a Q-matrix can be complete for one model but incomplete for another. The parameterization of general CDMs based on the kernel v_j suggests a distinction of CDMs that is used to structure the subsequent investigation of the conditions of completeness of the Q-matrix: (a) Two special CDMs: The Deterministic Input Noisy Output "AND" gate (DINA) model (Junker and Sijtsma, 2001; Macready and Dayton, 1977) and the Deterministic Input Noisy Output "OR" gate (DINO) model (Templin and Henson, 2006); (b) CDMs with main effects only; (c) CDMs with main effects and interaction effects; (d) CDMs with only interaction effects.

3.1 Two Special CDMs: The DINA Model and the DINO Model

3.1.1 The Deterministic Input Noisy Output "AND" Gate Model

The IRF of the DINA model in specific parameterization is $P(Y_j = 1 \mid \boldsymbol{\alpha}) = (1-s_j)^{\eta_j} g_j^{(1-\eta_j)}$, subject to $0 < g_j < 1 - s_j < 1 \ \forall j$. The conjunction parameter η_j is defined as $\eta_j = \prod_{k=1}^{K} \alpha_k^{q_{jk}}$; η_j indicates whether examinee i has mastered all the skills needed to answer item j correctly. The item-related parameters $s_j = P(Y_j = 0 \mid \eta_j = 1)$ and $g_j = P(Y_j = 1 \mid \eta_j = 0)$ formalize the probabilities of a slip (failing to answer item j correctly despite having the skills required to do so) and guessing (answering item j correctly despite lacking the skills required to do so), respectively. For item j, define the set $\mathcal{L}_j = \{k \mid q_{jk} = 1\}$ that contains the non-zero elements in the item skill vector $\mathbf{q}_j$ (i.e., the indices of all required skills α_k). Thus, the IRF of the DINA model as a general CDM using the logit link is

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = \frac{\exp\left(\beta_{j0} + \beta_{j(\forall k \in \mathcal{L}_j)} \prod_{k \in \mathcal{L}_j} \alpha_k\right)}{1 + \exp\left(\beta_{j0} + \beta_{j(\forall k \in \mathcal{L}_j)} \prod_{k \in \mathcal{L}_j} \alpha_k\right)}$$

subject to $\beta_{j(\forall k \in \mathcal{L}_j)} > 0$. (If $k \in \mathcal{L}_j = \{k \mid q_{jk} = 1\}$, then $q_{jk} = 1$ is always true; hence, q_{jk} has been dropped from the IRF.)

3.1.2 The Deterministic Input Noisy Output "OR" Gate Model

The DINO model (Templin and Henson, 2006) is a disjunctive CDM (i.e., mastery of a subset of the required skills is a sufficient condition for maximizing the probability of a correct item response). Define the disjunction parameter $\omega_j = 1 - \prod_{k=1}^{K}(1-\alpha_k)^{q_{jk}}$ that indicates whether at least one of the skills required for item j has been mastered. The IRF of the DINO model in specific parameterization is $P(Y_j = 1 \mid \boldsymbol{\alpha}) = (1-s_j)^{\omega_j} g_j^{(1-\omega_j)}$. The condition that mastery of just one skill of those required for item j already maximizes the probability of a correct response translates into the constraint that all coefficients—except β_{j0}—in Equation 1 be equal; only their signs oscillate depending on the order a of the terms in the equation: $(-1)^{a+1}$; for main effects, $a = 1$; for two-way interactions $a = 2$, and so on. Some algebra then leads to the compact form of the IRF of the DINO model using the logit link

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = \frac{\exp\left(\beta_{j0} + \beta_{jk}\left(1 - \prod_{l \in \mathcal{L}_j}(1-\alpha_l)\right)\right)}{1 + \exp\left(\beta_{j0} + \beta_{jk}\left(1 - \prod_{l \in \mathcal{L}_j}(1-\alpha_l)\right)\right)} \quad \text{for some } k \in \mathcal{L}_j$$

subject to $\beta_{jk} > 0$.

3.1.3 Q-Completeness

Chiu and collaborators proved for the DINA model (Chiu et al., 2009) and the DINO model (Chiu and Köhn, 2015) that $\mathbf{Q}$ is complete if and only if each skill is represented by at least one single-skill item—that is, $\mathbf{Q}$ has rows, $\mathbf{e}_1, \ldots, \mathbf{e}_K$, among its J rows, where $\mathbf{e}_k$ represents a $1 \times K$ vector, with the k^{th} element, e_k, equal to 1, and all other entries equal to 0. As an illustration for the DINA model, consider the two Q-matrices, $\mathbf{Q}_{1:3}$ and $\mathbf{Q}_{4:6}$, each with $K = 3$ skills and $J = 3$ items

$$\mathbf{Q}_{1:3} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \qquad \mathbf{Q}_{4:6} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

with the matrix subscripts referring to the item indices $j = 1, \ldots, 6$. $\mathbf{Q}_{1:3}$ is not complete, whereas $\mathbf{Q}_{4:6}$ is complete, as the computation of the expected item-response profiles $\boldsymbol{S}(\boldsymbol{\alpha})$ demonstrates. For the DINA model, the entries in $\boldsymbol{S}(\boldsymbol{\alpha})$ are defined as

$$S_j(\boldsymbol{\alpha}) = E(Y_j \mid \boldsymbol{\alpha}) = P(Y_j \mid \boldsymbol{\alpha}) = \frac{\exp\left(\beta_{j0} + \beta_{j(\forall k \in \mathcal{L}_j)} \prod_{k \in \mathcal{L}_j} \alpha_k\right)}{1 + \exp\left(\beta_{j0} + \beta_{j(\forall k \in \mathcal{L}_j)} \prod_{k \in \mathcal{L}_j} \alpha_k\right)}$$

The subsequent table only reports the coefficients that are retained in $S_j(\boldsymbol{\alpha})$, but not the expression of the entire logistic function.

	$\mathbf{Q}_{1:3}$			$\mathbf{Q}_{4:6}$		
$\boldsymbol{\alpha}$	$\mathbf{q}_1 = (011)$	$\mathbf{q}_2 = (101)$	$\mathbf{q}_3 = (110)$	$\mathbf{q}_4 = (100)$	$\mathbf{q}_5 = (010)$	$\mathbf{q}_6 = (001)$
	$S_1(\boldsymbol{\alpha})$	$S_2(\boldsymbol{\alpha})$	$S_3(\boldsymbol{\alpha})$	$S_4(\boldsymbol{\alpha})$	$S_5(\boldsymbol{\alpha})$	$S_6(\boldsymbol{\alpha})$
(000)	β_{10}	β_{20}	β_{30}	β_{40}	β_{50}	β_{60}
(100)	β_{10}	β_{20}	β_{30}	$\beta_{40} + \beta_{41}$	β_{50}	β_{60}
(010)	β_{10}	β_{20}	β_{30}	β_{40}	$\beta_{50} + \beta_{52}$	β_{60}
(001)	β_{10}	β_{20}	β_{30}	β_{40}	β_{50}	$\beta_{60} + \beta_{63}$
(110)	β_{10}	β_{20}	$\beta_{30} + \beta_{3(12)}$	$\beta_{40} + \beta_{41}$	$\beta_{50} + \beta_{52}$	β_{60}
(101)	β_{10}	$\beta_{20} + \beta_{2(13)}$	β_{30}	$\beta_{40} + \beta_{41}$	β_{50}	$\beta_{60} + \beta_{63}$
(011)	$\beta_{10} + \beta_{1(23)}$	β_{20}	β_{30}	β_{40}	$\beta_{50} + \beta_{52}$	$\beta_{60} + \beta_{63}$
(111)	$\beta_{10} + \beta_{1(23)}$	$\beta_{20} + \beta_{2(13)}$	$\beta_{30} + \beta_{3(12)}$	$\beta_{40} + \beta_{41}$	$\beta_{50} + \beta_{52}$	$\beta_{60} + \beta_{63}$

Clearly, $\mathbf{Q}_{1:3}$ is not complete because, for example, $\boldsymbol{\alpha}_1 = (000)^T \neq \boldsymbol{\alpha}_2 = (100)^T$, but $\boldsymbol{S}(\boldsymbol{\alpha}_1) = \boldsymbol{S}(\boldsymbol{\alpha}_2) = (\frac{e^{\beta_{10}}}{1+e^{\beta_{10}}}, \frac{e^{\beta_{20}}}{1+e^{\beta_{20}}}, \frac{e^{\beta_{30}}}{1+e^{\beta_{30}}})$. Thus, $\mathbf{Q}_{1:3}$ does not allow to distinguish between all $\boldsymbol{\alpha}$ (i.e., all the $M = 2^K$ proficiency classes). However, if items 4–6 of $\mathbf{Q}_{4:6}$ are included, then $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^* \Rightarrow \boldsymbol{S}(\boldsymbol{\alpha}) \neq \boldsymbol{S}(\boldsymbol{\alpha}^*)$. In case of these single skill items $\mathbf{q}_j = \mathbf{e}_k$; hence, the term $\beta_{j(\forall k \in \mathcal{L}_j)} \prod_{k \in \mathcal{L}_j} \alpha_k$ is reduced to a skill "main effect"—β_{jk}—that then allows also to discriminate between $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\alpha}_3$, and $\boldsymbol{\alpha}_4$.)

In summary, for the DINA model and the DINO model, the inclusion of all K single-skill items in the Q-matrix is a necessary condition for its completeness. For other CDMs, however, this is a sufficient, but not a necessary condition—that is, alternative compositions of the Q-matrix that do not include all single-skill items also guarantee completeness.

3.2 CDMs With Main Effects Only

As an example for a CDM with main effects only, consider the GDM that has IRF and expected item response $S_j(\boldsymbol{\alpha})$

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = \frac{\exp\left(\beta_{j0} + \sum_{k=1}^{K} \beta_{jk} q_{jk} \alpha_k\right)}{1 + \exp\left(\beta_{j0} + \sum_{k=1}^{K} \beta_{jk} q_{jk} \alpha_k\right)} = S_j(\boldsymbol{\alpha})$$

For the GDM, $\mathbf{Q}_{4:6}$ is guaranteed to be complete due to the sufficiency condition. However, $\mathbf{Q}_{1:3}$ is also complete for the GDM despite the removal of all interaction effects $\beta_{j(kk')}$—that is, $\mathbf{S}(\boldsymbol{\alpha}) = \mathbf{S}(\boldsymbol{\alpha}^*) \Rightarrow \boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ still holds:

	$\mathbf{Q}_{1:3}$		
$\boldsymbol{\alpha}$	$\mathbf{q}_1 = (011)$	$\mathbf{q}_2 = (101)$	$\mathbf{q}_3 = (110)$
	$S_1(\boldsymbol{\alpha})$	$S_2(\boldsymbol{\alpha})$	$S_3(\boldsymbol{\alpha})$
(000)	β_{10}	β_{20}	β_{30}
(100)	β_{10}	$\beta_{20} + \beta_{21}$	$\beta_{30} + \beta_{31}$
(010)	$\beta_{10} + \beta_{12}$	β_{20}	$\beta_{30} + \beta_{32}$
(001)	$\beta_{10} + \beta_{13}$	$\beta_{20} + \beta_{23}$	β_{30}
(110)	$\beta_{10} + \beta_{12}$	$\beta_{20} + \beta_{21}$	$\beta_{30} + \beta_{31} + \beta_{32}$
(101)	$\beta_{10} + \beta_{13}$	$\beta_{20} + \beta_{21} + \beta_{23}$	$\beta_{30} + \beta_{31}$
(011)	$\beta_{10} + \beta_{12} + \beta_{13}$	$\beta_{20} + \beta_{23}$	$\beta_{30} + \beta_{32}$
(111)	$\beta_{10} + \beta_{12} + \beta_{13}$	$\beta_{20} + \beta_{21} + \beta_{23}$	$\beta_{30} + \beta_{31} + \beta_{32}$

3.3 CDMs With Main Effects and Interaction Effects

Take the (saturated) LCDM as an example for a model containing all main effects and all interaction effects. For $K = 3$ skills, the IRF is

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = \frac{\exp\left(\beta_{j0} + \sum_{k=1}^{K} \beta_{jk} q_{jk}\alpha_k + \sum_{k'=k+1}^{3}\sum_{k=1}^{2} \beta_{j(kk')} q_{jk} q_{jk'} \alpha_k \alpha_{k'} + \beta_{j(123)} \prod_{k=1}^{3} q_{jk}\alpha_k\right)}{1 + \exp\left(\beta_{j0} + \sum_{k=1}^{3} \beta_{jk} q_{jk}\alpha_k + \sum_{k'=k+1}^{3}\sum_{k=1}^{2} \beta_{j(kk')} q_{jk} q_{jk'} \alpha_k \alpha_{k'} + \beta_{j(123)} \prod_{k=1}^{3} q_{jk}\alpha_k\right)}$$

Note that the expression of the expected response $S_j(\boldsymbol{\alpha})$ is equal to the IRF of item j. For the saturated LCDM, $\mathbf{Q}_{4:6}$ is complete due to the sufficiency condition of Q-matrices containing all K single skill items. $\mathbf{Q}_{1:3}$, on the other hand, does not contain any single-skill item, but is also complete for the saturated LCDM, as the calculation of the $\boldsymbol{S}(\boldsymbol{\alpha})$ demonstrates:

	$\mathbf{Q}_{1:3}$		
$\boldsymbol{\alpha}$	$\mathbf{q}_1 = (011)$	$\mathbf{q}_2 = (101)$	$\mathbf{q}_3 = (110)$
	$S_1(\boldsymbol{\alpha})$	$S_2(\boldsymbol{\alpha})$	$S_3(\boldsymbol{\alpha})$
(000)	β_{10}	β_{20}	β_{30}
(100)	β_{10}	$\beta_{20} + \beta_{21}$	$\beta_{30} + \beta_{31}$
(010)	$\beta_{10} + \beta_{12}$	β_{20}	$\beta_{30} + \beta_{32}$
(001)	$\beta_{10} + \beta_{13}$	$\beta_{20} + \beta_{23}$	β_{30}
(110)	$\beta_{10} + \beta_{12}$	$\beta_{20} + \beta_{21}$	$\beta_{30} + \beta_{31} + \beta_{32} + \beta_{3(12)}$
(101)	$\beta_{10} + \beta_{13}$	$\beta_{20} + \beta_{21} + \beta_{23} + \beta_{2(13)}$	$\beta_{30} + \beta_{31}$
(011)	$\beta_{10} + \beta_{12} + \beta_{13} + \beta_{1(23)}$	$\beta_{20} + \beta_{23}$	$\beta_{30} + \beta_{32}$
(111)	$\beta_{10} + \beta_{12} + \beta_{13} + \beta_{1(23)}$	$\beta_{20} + \beta_{21} + \beta_{23} + \beta_{2(13)}$	$\beta_{30} + \beta_{31} + \beta_{32} + \beta_{3(12)}$

3.4 CDMs With No Main Effects, But Only Interaction Effects

What are the consequences if all main effects are removed from, say the saturated LCDM? For example, consider

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = \frac{\exp\left(\beta_{j0} + \sum_{k'=k+1}^{K}\sum_{k=1}^{K-1} \beta_{j(kk')} q_{jk} q_{jk'} \alpha_k \alpha_{k'} + \cdots + \beta_{j(12\ldots K)} \prod_{k=1}^{K} q_{jk}\alpha_k\right)}{1 + \exp\left(\beta_{j0} + \sum_{k'=k+1}^{K}\sum_{k=1}^{K-1} \beta_{j(kk')} q_{jk} q_{jk'} \alpha_k \alpha_{k'} + \cdots + \beta_{j(12\ldots K)} \prod_{k=1}^{K} q_{jk}\alpha_k\right)}$$

Then, as the inspection of the $\boldsymbol{S}(\boldsymbol{\alpha})$ immediately shows, matrix $\mathbf{Q}_{1:3}$ is no longer complete because some $\boldsymbol{S}(\boldsymbol{\alpha}) = \boldsymbol{S}(\boldsymbol{\alpha}^*)$ despite $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$: Four of the proficiency classes are not identifiable. Note that, different from the DINA model, using $\mathbf{Q}_{4:6}$ as Q-matrix instead of $\mathbf{Q}_{1:3}$ does not resolve the completeness issue but rather seems to worsen it because then, none of the proficiency classes is identifiable:

	$\mathbf{Q}_{1:3}$			$\mathbf{Q}_{4:6}$		
$\boldsymbol{\alpha}$	$\mathbf{q}_1 = (011)$	$\mathbf{q}_2 = (101)$	$\mathbf{q}_3 = (110)$	$\mathbf{q}_4 = (100)$	$\mathbf{q}_5 = (010)$	$\mathbf{q}_6 = (001)$
	$S_1(\boldsymbol{\alpha})$	$S_2(\boldsymbol{\alpha})$	$S_3(\boldsymbol{\alpha})$	$S_4(\boldsymbol{\alpha})$	$S_5(\boldsymbol{\alpha})$	$S_6(\boldsymbol{\alpha})$
(000)	β_{10}	β_{20}	β_{30}	β_{40}	β_{50}	β_{60}
(100)	β_{10}	β_{20}	β_{30}	β_{40}	β_{50}	β_{60}
(010)	β_{10}	β_{20}	β_{30}	β_{40}	β_{50}	β_{60}
(001)	β_{10}	β_{20}	β_{30}	β_{40}	β_{50}	β_{60}
(110)	β_{10}	β_{20}	$\beta_{30} + \beta_{3(12)}$	β_{40}	β_{50}	β_{60}
(101)	β_{10}	$\beta_{20} + \beta_{2(13)}$	β_{30}	β_{40}	β_{50}	β_{60}
(011)	$\beta_{10} + \beta_{1(23)}$	β_{20}	β_{30}	β_{40}	β_{50}	β_{60}
(111)	$\beta_{10} + \beta_{1(23)}$	$\beta_{20} + \beta_{2(13)}$	$\beta_{30} + \beta_{3(12)}$	β_{40}	β_{50}	β_{60}

4. Synthesis and Conclusions: Q-Completeness

In light of the last result, it comes as no surprise that models without main effects, but only containing interaction terms—at least to our knowledge—have never been proposed in the literature: These models cannot discriminate between the M proficiency classes. Said differently, for models without the k^{th} main effect, any Q-matrix is incomplete.

The DINA model and the DINO model form a category of their own: A Q-matrix to be used with either of the two models is complete if and only if it contains among its J items all K single skill items, with item skill vectors $\mathbf{q}_1 = \mathbf{e}_1$, $\mathbf{q}_1 = \mathbf{e}_1$, …, $\mathbf{q}_K = \mathbf{e}_K$, where $\mathbf{e}_k$ was defined earlier as a unit vector with all elements equal 0 except the k^{th} entry (for proofs of this claim, consult Chiu et al., 2009; Chiu and Köhn, 2015).

For CDMs containing only main effects, the following claim about the completeness of $\mathbf{Q}$ can be made. Consider two skill profiles $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$. Then there exists at least one k such that $\alpha_k = 1$ and $\alpha_k^* = 0$. In addition, assume that q_{jk} in $\mathbf{Q}$ is 1 for some j. For models that contain only main effects, a $J \times K$ matrix $\mathbf{Q}$ is complete if and only if it contains K linearly independent q-vectors and $\sum_{k'=1,k'\neq k}^{K} \beta_{jk'} q_{jk'}(\alpha_{k'} - \alpha_{k'}^*) \neq \beta_{jk}$ for some k. For an illustration, look at

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

that consists of three linearly independent q-vectors. But the constraint $\sum_{k'=1,k'\neq k}^{K} \beta_{jk'} q_{jk'}(\alpha_{k'} - \alpha_{k'}^*) \neq \beta_{jk}$ is possibly violated, as can be readily verified:

	$\mathbf{Q}_{1:3}$		
$\boldsymbol{\alpha}$	$\mathbf{q}_1 = (101)$	$\mathbf{q}_2 = (011)$	$\mathbf{q}_3 = (111)$
	$S_1(\boldsymbol{\alpha})$	$S_2(\boldsymbol{\alpha})$	$S_3(\boldsymbol{\alpha})$
(001)	$\beta_{10} \quad + \beta_{13}$	$\beta_{20} \quad + \beta_{23}$	$\beta_{30} \quad + \beta_{33}$
(110)	$\beta_{10} + \beta_{11}$	$\beta_{20} + \beta_{22}$	$\beta_{30} + \beta_{31} + \beta_{32}$

If $\beta_{13} = \beta_{11}$, $\beta_{23} = \beta_{22}$, and $\beta_{33} = \beta_{31} + \beta_{32}$, then the two proficiency classes with skill profiles (001) and (110) cannot be distinguished. However, this particular constellation is presumably pretty rare; it can only occur if the expected responses for distinct $\boldsymbol{\alpha}$ are not nested within each other.

For CDMs containing main effects and interaction terms, the general condition for Q-completeness can be stated as follows. Consider two skill profiles $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$. Then there exists at least one k such that $\alpha_k = 1$ and $\alpha_k^* = 0$. In addition, assume that q_{jk} in $\mathbf{Q}$ is 1 for some j. For models that contain main effects and interaction terms, a $J \times K$ matrix $\mathbf{Q}$ is complete if and only if it contains K linearly independent q-vectors and $\sum_{k'=1,k'\neq k}^{K} \beta_{jk'} q_{jk'}(\alpha_{k'} - \alpha_{k'}^*) + \cdots + \beta_{j(12\ldots K)} \prod_{k=1}^{K} q_{jk} \Big(\prod_{k=1}^{K} \alpha_k - \prod_{k=1}^{K} \alpha_k^* \Big) \neq -\beta_{jk}$ for some k.

REFERENCES

Chiu, C.-Y., and Köhn, H.-F. (2015), "Consistency of Cluster Analysis for Cognitive Diagnosis: The DINO Model and the DINA Model Revisited," *Applied Psychological Measurement, 39*, 465–479.

Chiu, C.-Y., Douglas, J. A., and Li, X. (2009), "Cluster Analysis for Cognitive Diagnosis: Theory and Applications," *Psychometrika, 74*, 633–665.

de la Torre, J. (2011), "The Generalized DINA Model Framework," *Psychometrika, 76*, 179–199.

DiBello, L. V., Roussos, L. A., and Stout, W. F. (2007), "Review of Cognitively Diagnostic Assessment and a Summary of Psychometric Models," in *Handbook of Statistics: Vol. 26. Psychometrics*, eds. C. R. Rao and S. Sinharay, Amsterdam: Elsevier, pp. 979–1030.

Haberman, S. J., and von Davier, M. (2007), "Some Notes on Models for Cognitively Based Skill Diagnosis," In C. R. Rao and S. Sinharay (Eds.), *Handbook of Statistics: Vol. 26. Psychometrics* in *Handbook of Statistics: Vol. 26. Psychometrics*, eds. C. R. Rao and S. Sinharay, Amsterdam: Elsevier, pp. 1031–1038.

Henson, R. A., Templin, J. L., and 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., and Sijtsma, K. (2001), "Cognitive Assessment Models With Few Assumptions, and Connections With Nonparametric Item Response Theory," *Applied Psychological Measurement, 25*, 258–272.

Leighton, J., and Gierl, M. (2007), *Cognitive Diagnostic Assessment for Education: Theory and Applications*, Cambridge, UK: Cambridge University Press.

Macready, G. B., and Dayton, C. M. (1977), "The Use of Probabilistic Models in the Assessment of Mastery," *Journal of Educational Statistics, 33*, 379–416.

Rupp, A. A., Templin, J. L., and Henson, R. A. (2010), *Diagnostic Measurement: Theory, Methods, and Applications*, New York: Guilford.

Tatsuoka, K. (1985), "A Probabilistic Model for Diagnosing Misconception in the Pattern Classification Approach," *Journal of Educational Statistics, 12*, 55–73.

Templin, J. L., and Henson, R. A. (2006), "Measurement of Psychological Disorders Using Cognitive Diagnosis Models," *Psychological Methods, 11*, 287–305.

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.