PSYCHOMETRIKA—VOL. 81, NO. 3, 625–649
SEPTEMBER 2016
DOI: 10.1007/s11336-015-9471-z

IDENTIFIABILITY OF DIAGNOSTIC CLASSIFICATION MODELS

GONGJUN XU

UNIVERSITY OF MINNESOTA

STEPHANIE ZHANG

GOOGLE INC.

Diagnostic classification models (DCMs) are important statistical tools in cognitive diagnosis. In this paper, we consider the issue of their identifiability. In particular, we focus on one basic and popular model, the DINA model. We propose sufficient and necessary conditions under which the model parameters are identifiable from the data. The consequences, in terms of the consistency of parameter estimates, of fulfilling or failing to fulfill these conditions are illustrated via simulation. The results can be easily extended to the DINO model through the duality of the DINA and DINO models. Moreover, the proposed theoretical framework could be applied to study the identifiability issue of other DCMs.

Key words: diagnostic classification models, the DINA model, model identifiability, Q-matrix.

1. Introduction

Diagnostic measurement is the process of arriving at a classification-based decision about an individual's latent traits, based on his or her observed responses. The idea of conducting diagnostic measurement has been gaining increasing traction, thanks in part to demand from the fields of education and psychology. Both are traditional psychometric fields, with strong interest in measurement techniques for psychological constructs such as skills, knowledge, personality traits, or psychological disorders. While diagnosis has strong medical roots, educational interest in the field is more recent. A diagnostic, skills-based focus is a key part of recent government initiatives such as the United States Department of Education's Race to the Top, which included "building data systems that measure student growth and success, and inform teachers and principals about how they can improve instruction" as one of its four main goals (U.S. Department of Education, 2009). Measuring students' growth and success means obtaining diagnostic information about their skill set; this is very important for constructing efficient, focused remedial strategies for improving student and teacher results.

Diagnostic classification models (DCMs) are important statistical tools in cognitive diagnosis and have gained increasing interest in the recent years. The aim is to detect the presence or absence of multiple fine-grained skills or attributes. This provides more informative feedback on, for example, student skill-sets, and allows for the design of more effective intervention strategies (Rupp, Templin, & Henson, 2010). These models provide specific attribute profiles for each subject, which allows for effective intervention for personal improvement. A partial list of the DCMs studied in the literature includes the conjunctive DINA and NIDA models (Junker & Sijtsma, 2001; Rupp, Templin, & Henson, 2010), the reparameterized unified/fusion model (RUM) (DiBello, Stout, & Roussos, 1995), the compensatory DINO and NIDO models (Templin & Henson, 2006), the rule space method (Tatsuoka, 1985, 2009), the attribute hierarchy method

Correspondence should be made to Gongjun Xu, School of Statistics, University of Minnesota, 313 Ford Hall, 224 Church St SE, Minneapolis, MN 55455 USA. Email: xuxxx360@umn.edu

(Leighton, Gierl, & Hunka, 2004), the Generalized DINA models (de la Torre, 2011), and the general diagnostic model (von Davier, 2005, 2008; von Davier & Yamamoto, 2004); see also de la Torre (2011), de la Torre and Douglas (2004), Henson, Templin and Willse (2009) for more developments and approaches to cognitive diagnosis.

Although DCMs have many attractive traits for practitioners looking to perform diagnostic classification and are a fertile area of active research, several persistent statistical issues make their application problematic. Key among them is statistical identifiability, i.e., the feasibility of recovering model parameters based on the observed data. It is a prerequisite for most common statistical inferences, especially parameter estimation, and its study dates back to Koopmans (1950) and Koopmans and Reiersøl (1950); further developments can be found in Gabrielsen (1978), Paulino and de Bragançca Pereira (1994), and San Martín and Quintana (2002). In particular, for DCMs such as DINA model, we need that distinct sets of model parameters yield distinct distributions of the observed responses, and thus, distinct likelihood functions. When the data does not differentiate between different parameterizations, we lose statistically desirable properties such as the consistency of parameter estimates.

Researchers have long been aware of the fact that Q-matrix based DCMs are generally not identifiable (DeCarlo, 2011; DiBello, Stout, & Roussos, 1995; Maris & Bechger, 2009; Tatsuoka, 2009, 1991), though there is a tendency to gloss over the issue in practice due to a lack of theoretical development on the topic (de la Torre and Douglas, 2004). With Q-matrix based DCMs, identifiability also affects the classification of respondents according to their latent traits, which is dependent on the accuracy of the parameter estimates. Unprincipled use of standard DCMs may lead to misleading conclusions about the respondents' latent traits (Maris & Bechger, 2009; Tatsuoka, 2009). Consider the DINA model as an example. Estimation of the parameters has been studied extensively in the literature and different estimation procedures have been proposed. de la Torre (2009) uses the EM algorithm and the MCMC method to estimate the slipping and guessing parameters in the DINA model. Identifiability is important no matter what estimation method is used; it is necessary for the consistency of the EM algorithm and for the interpretability and convergence of the estimates generated by the MCMC method.

Under the DINA model, when both the slipping and guessing parameters are zero, Chiu (2009) proposes the completeness assumption of the Q-matrix, which turns out to be the sufficient and necessary condition for the identifiability in this ideal case. See DeCarlo (2011) for further discussion. Recently, in Chen, Liu, Xu, and Ying (2014), the authors provide sufficient and necessary conditions for the model identifiability when the guessing parameters are known. Yet in the more commonly seen and practical case that neither the slipping nor the guessing parameters are known, the corresponding theoretical justification of the estimates' consistency is still lacking. This paper focuses on the identifiability of the parameters in the DINA model. In particular, we propose sufficient and necessary conditions under which the slipping, guessing, and population parameters are estimable from the data. The analysis method developed in this paper is based on the theoretical framework in Liu, Xu, and Ying (2012, 2013). This paper generalizes the result in Chen et al. (2014) and the derivation is technically more subtle. We emphasize that this method is generic in the sense that it can be employed for the analysis of other diagnostic classification or latent class models, which is an interesting future research topic.

The remainder of this paper is organized as follows. Section 2 contains useful background on the DINA model some central concepts in diagnostic classification modeling. Section 3 introduces the issue of identifiability and the relevant aspects of the theoretical framework developed in Liu et al. (2012, 2013). This is followed by our main results, the sufficient and necessary conditions for the identifiability of the DINA model, in Section 4. These theoretical results are illustrated through several simulations in Section 5. The corresponding proofs are given in Appendix.

2. Notation and model specification

2.1. Problem Setting

Most DCMs begin from the same basic setting, in which subjects (respondents) provide observed responses to the items (questions) which make up the assessment, and these responses depend in some way on unobserved latent attributes (traits). We consider tests consisting of a pre-specified number of items J depending on a known number of attributes K, given to N subjects.

A complete set of K latent traits is known as an attribute profile; these profiles are denoted by column vectors $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^\top$, where $\alpha_k \in \{0, 1\}$ indicate the absence or presence, respectively, of the kth attribute and the superscript $\top$ denotes the transpose. In addition, let $\mathbf{R} = (R_1, \ldots, R_J)^\top$ denote a random vector of responses to the J test items. Both $\boldsymbol{\alpha}$ and $\mathbf{R}$ are subject-specific; a particular subject i's attribute and response vectors may be denoted by $\boldsymbol{\alpha}^i$ and $\mathbf{R}^i$, respectively, for $i = 1, \ldots, N$. Note that when unspecified, vectors indexed by $\boldsymbol{\alpha}$ or $\mathbf{r}$ will presume the natural lexicographic (alphabetic) ordering on the indices. For example, in the case of $\{0, 1\}^2$, $(0, 0) < (0, 1) < (1, 0) < (1, 1)$.

The underlying cognitive structure, i.e., the relationship between the items and the attributes, is generally described by the Q-matrix, initially proposed by Tatsuoka (1983). A Q-matrix Q is a $J \times K$ binary matrix with entries $q_{jk} \in \{0, 1\}$ indicating the absence or presence, respectively, of a link between the jth item and the kth attribute; this link is often referred to as an "attribute requirement." The row vectors, $\mathbf{q}_j$ of Q correspond to the full attribute requirements of each item. For instance, we consider the following 3×2 Q-matrix, which gives the corresponding item and attribute relationships.

$Q =$ (1)

	addition	multiplication
$2 + 1$	1	0
3×2	0	1
$(2 + 1) \times 2$	1	1

2.2. The DINA Model

This paper focuses on the DINA model, one of the most widely used DCMs. The DINA model assumes a conjunctive relationship among attributes, that is, it is necessary to possess all the attributes indicated by the Q-matrix to be capable of providing a positive response to an item. In addition, having additional unnecessary attributes does not compensate for the lack of the necessary attributes. The DINA model is particularly popular in the context of educational testing.

Under the DINA model, given an attribute profile $\boldsymbol{\alpha}$ and a Q-matrix Q, we can define the quantity

$$\xi_j(Q, \boldsymbol{\alpha}) = \prod_{k=1}^{K} (\alpha_k)^{q_{jk}} = I(\alpha_k \geq q_{jk} : k = 1, \ldots, K), \tag{2}$$

which indicates whether a respondent with attribute profile $\boldsymbol{\alpha}$ possesses all the attributes required for item j. If we suppose no uncertainty in the response, then a respondent i with attribute profile $\boldsymbol{\alpha}^i$ will have responses $R^i_j = \xi_j(Q, \boldsymbol{\alpha}^i)$ for each $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, J\}$. Thus, the vectors $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_J)^\top$ for each $\boldsymbol{\alpha} \in \{0, 1\}^K$ are known as *ideal response vectors*.

In the DINA model, uncertainty is incorporated at the item level, using the slipping and guessing parameters $\mathbf{s}$ and $\mathbf{g}$; the names "slipping" and "guessing" arise from the educational applications. For each item $j = 1, \ldots, J$, the slipping parameter $s_j = P(R_j = 0|\xi_j = 1)$

denotes the probability of the respondent making an incorrect response despite mastering all necessary skills and having a correct ideal response; similarly, the guessing parameter $g_j = P(R_j = 1|\xi_j = 0)$ denotes the probability of a correct response despite an incorrect ideal response. In the technical development, it is more convenient to work with the complement of the slipping parameter, referred to as $\mathbf{c} = 1 - \mathbf{s}$; this alternative parameterization is used throughout the paper.

Conditional on $Q, \boldsymbol{\alpha}, \mathbf{c}, \mathbf{g}$, an individual's responses R_j are jointly independent Bernoullis with success probabilities

$$P(R_j = 1|Q, \boldsymbol{\alpha}, \mathbf{c}, \mathbf{g}) = c_j^{\xi_j(Q,\boldsymbol{\alpha})} g_j^{1-\xi_j(Q,\boldsymbol{\alpha})}. \tag{3}$$

Thus, the probability of a particular response vector $\mathbf{r} \in \{0, 1\}^J$ given $Q, \boldsymbol{\alpha}, \mathbf{c}, \mathbf{g}$ is

$$P(\mathbf{r}|Q, \boldsymbol{\alpha}, \mathbf{c}, \mathbf{g}) = \prod_{j=1}^{J} c_j^{\xi_j r_j} g_j^{(1-\xi_j) r_j} (1 - c_j)^{\xi_j (1-r_j)} (1 - g_j)^{(1-\xi_j)(1-r_j)}, \tag{4}$$

where ξ_j is shorthand for $\xi_j(Q, \boldsymbol{\alpha})$.

In addition to $\mathbf{c}$ and $\mathbf{g}$, the response distribution also depends on the distribution of attribute profiles. We assume that the respondents are a random sample of size N from a designated population, so that their attribute profiles $\boldsymbol{\alpha}^i$, $i = 1, \dots, N$ are i.i.d. random variables following a multinomial distribution with probabilities

$$P(\boldsymbol{\alpha}^i = \boldsymbol{\alpha}) = p_{\boldsymbol{\alpha}},$$

where $p_{\boldsymbol{\alpha}} \in [0, 1]\ \forall\ \boldsymbol{\alpha} \in \{0, 1\}^K$ and $\sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} = 1$. The distribution is thus characterized by the column vector $\mathbf{p} = (p_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \{0, 1\}^K)$. Given $\mathbf{p}$, it is possible to calculate the marginal probabilities of each response, rather than the conditional probabilities given the latent variables $\boldsymbol{\alpha}$. Specifically, conditional on $Q, \mathbf{c}, \mathbf{g}, \mathbf{p}$, the response vectors are i.i.d. random variables following a multinomial distribution with probabilities

$$P(\mathbf{r}|Q, \mathbf{c}, \mathbf{g}, \mathbf{p}) = \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} P(\mathbf{r}|Q, \mathbf{c}, \mathbf{g}, \boldsymbol{\alpha}) p_{\boldsymbol{\alpha}} \tag{5}$$

for each $\mathbf{r} \in \{0, 1\}^J$. The conditional probabilities $P(\mathbf{r}|Q, \mathbf{c}, \mathbf{g}, \boldsymbol{\alpha})$ are calculated as in (4).

3. Important Concepts

The following list summarizes some general notation used throughout the rest of the discussion.

- For a matrix M, let M_i denote the ith row of M. This may be extended to $M_{1:n}$, which denotes the submatrix containing the first n rows of M.
- Let $\mathcal{R}_M = \{M_i^\top : i = 1, \dots, d_1\}$ denote the set of row vectors of a $d_1 \times d_2$ matrix M.
- The matrix $\mathcal{I}_d$ is the $d \times d$ identity matrix.
- The column vector $\mathbf{e}_i$ is a standard basis vector; its ith element is one and the rest are zero.

- The symbols $\mathbf{0}$ and $\mathbf{1}$ denote the zero and one column vectors, i.e., $(0, \ldots, 0)^\top$ and $(1, \ldots, 1)^\top$, respectively.
- Given d-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$, let $\mathbf{u} \succ \mathbf{v}$ if the entries $u_i > v_i$ for all $i \in \{1, \ldots, d\}$. Similarly define the operations $\prec$, $\succeq$, and $\preceq$.

3.1. *Identifiability*

This paper is concerned with the general statistical concept of identifiability, as applied to cognitive diagnosis, and the DINA model in particular. This is the issue of primary concern when examining the consistency of estimates in diagnostic classification models.

We say that a set of parameters θ for a family of distributions $\{f(x|\theta) : \theta \in \Theta\}$ is identifiable if distinct values of θ correspond to distinct probability density functions, i.e., for any θ there is no $\tilde{\theta} \in \Theta \backslash \{\theta\}$ such that $f(x|\theta) = f(x|\tilde{\theta})$. In addition, we say that a set of parameters θ is locally identifiable if there exists a neighborhood of θ, $\mathcal{N}_\theta \in \Theta$, such that there is no $\tilde{\theta} \in \mathcal{N}_\theta \backslash \{\theta\}$ such that $f(x|\theta) = f(x|\tilde{\theta})$.

Both identifiability and local identifiability of latent class models are well-established concepts in latent class analysis (e.g., Goodman, 1974; McHugh, 1956). Developments in the IRT models can be found in Bechger,Verstralen and Verhelst (2002), Maris and Bechger (2004), and San Martín, Rolin and Castro (2013). Identifiability is an important prerequisite for many types of statistical inference, such as parameter estimation and hypothesis testing. Local identifiability is a weaker form of identifiability, which ensures that the model parameters are identifiable in a neighborhood of the true parameter values. This is necessary for the model parameters to be estimable. Here we say the parameter is estimable if we can construct a consistent estimator for the parameter. That is, for parameter θ, there exists $\hat{\theta}_N$ such that $\hat{\theta}_N - \theta \to 0$ in probability as the sample size $N \to \infty$. For the DINA model, when the parameters are identifiable, the corresponding maximum likelihood estimator is consistent. On the other hand, the local identifiability ensures that the likelihood function has a consistent local maximizer $\hat{\theta}_N$ in the neighborhood of the true parameter θ.

3.2. *Completeness*

Identifiability has long been a concern for diagnostic classification models. The earliest work on identifiability concerns the identifiability of ideal response vectors, which depends on the Q-matrix. Chiu (2009) calls Q-matrices under which $\boldsymbol{\xi}(Q, \boldsymbol{\alpha}) \neq \boldsymbol{\xi}(Q, \boldsymbol{\alpha}')$ for all $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}'$ *complete*. The mathematical requirements on the Q-matrix for completeness are well known (Chiu, 2009; DeCarlo, 2011; DiBello, Stout, & Roussos, 1995; Tatsuoka, 1991); we use these requirements to create a mathematical definition of completeness:

Definition 1. A Q-matrix is said to be *complete* if $\{\mathbf{e}_j^\top : j = 1, \ldots, K\} \subset \mathcal{R}_Q$; otherwise, we say that Q is *incomplete*.

To interpret, for each attribute there must exist an item requiring that and only that attribute. The Q-matrix is complete if there exist K rows of Q that can be ordered to form the K-dimensional identity matrix $\mathcal{I}_K$. A simple (and minimal) example of a complete Q-matrix is the $K \times K$ identity matrix $\mathcal{I}_K$.

3.3. *The T-Matrix*

The T-matrix is a central construct in proving results about identifiability; in particular, it sets up a linear dependence between the attribute distribution and the response distribution (See Liu et al., 2012, 2013).

The T-matrix specifies the probability that an individual with attribute profile $\boldsymbol{\alpha}$ will answer all items in some subset of the items $S \subseteq \{1, \ldots, J\}$ correctly. The subsets S may be indexed by response vectors $\mathbf{r} \in \{0, 1\}^J$ with exactly the items in S correct, i.e., $r_j = 1$ iff $j \in S$. Then the T-matrix contains the probabilities that, given $Q, \mathbf{c}, \mathbf{g}, \boldsymbol{\alpha}$, the random response vectors $\mathbf{R} \succeq \mathbf{r}$ for each $\mathbf{r} \in \{0, 1\}^J$. Thus the entries of the T-matrix $T(Q, \mathbf{c}, \mathbf{g})$, indexed by $\mathbf{r}$ and $\boldsymbol{\alpha}$, are

$$\begin{aligned} t_{\mathbf{r},\boldsymbol{\alpha}}(Q, \mathbf{c}, \mathbf{g}) &= P(\mathbf{R} \succeq \mathbf{r} | Q, \mathbf{c}, \mathbf{g}, \boldsymbol{\alpha}) \\ &= \sum_{\mathbf{r}' \succeq \mathbf{r}} P(\mathbf{r}' | Q, \mathbf{c}, \mathbf{g}, \boldsymbol{\alpha}) = \prod_{j: r_j = 1} c_j^{\xi_j(Q,\boldsymbol{\alpha})} g_j^{1-\xi_j(Q,\boldsymbol{\alpha})} \end{aligned}$$

for all $\mathbf{r} \neq \mathbf{0}$. Note that $t_{\mathbf{0},\boldsymbol{\alpha}}(Q, \mathbf{c}, \mathbf{g}) = P(\mathbf{R} \succeq \mathbf{0}) = 1$ for all $Q, \mathbf{c}, \mathbf{g}, \boldsymbol{\alpha}$.

We use the following example to illustrate the introduced T-matrix.

Example 1. Suppose that we are interested in testing two attributes with the Q-matrix as introduced in Equation (1). The population is naturally divided into four strata with the corresponding distribution vector $\mathbf{p} = (p_{00}, p_{01}, p_{10}, p_{11})^\top$. Let the first to fourth columns of the T-matrix be indexed by attribute profiles (0, 0), (0, 1), (1, 0), and (1, 1), respectively. Consider the first four rows of T: the first three rows of T correspond to items one, two and three, and the fourth row corresponds to the combination of items one and two. Then under the Q-matrix in Equation (1), the first four rows of the T-matrix is

$$T(Q, \mathbf{c}, \mathbf{g}) = \begin{pmatrix} g_1 & g_1 & 1 - s_1 & 1 - s_1 \\ g_2 & 1 - s_2 & g_2 & 1 - s_2 \\ g_3 & g_3 & g_3 & 1 - s_3 \\ g_1 g_2 & g_1(1 - s_2) & (1 - s_1) g_2 & (1 - s_1)(1 - s_2) \\ & \cdots & \cdots & \end{pmatrix},$$

where g_j and s_j are the corresponding guessing and slipping parameters for the j th item.

By definition, multiplying the T-matrix by the the distribution of attribute profiles $\mathbf{p}$ results in a vector containing the marginal probabilities of successfully answering each subset of items correctly. The $\mathbf{r}$th entry of this vector is

$$\begin{aligned} T_{\mathbf{r}}(Q, \mathbf{c}, \mathbf{g})^\top \mathbf{p} &= \sum_{\boldsymbol{\alpha}} t_{\mathbf{r},\boldsymbol{\alpha}}(Q, \mathbf{c}, \mathbf{g}) p_{\boldsymbol{\alpha}} \\ &= \sum_{\mathbf{r}' \succeq \mathbf{r}} P(\mathbf{r}' | Q, \mathbf{c}, \mathbf{g}, \mathbf{p}) = P(\mathbf{R} \succeq \mathbf{r} | Q, \mathbf{c}, \mathbf{g}, \mathbf{p}). \end{aligned}$$

Thus, $T(Q, \mathbf{c}, \mathbf{g})$ describes the linear dependence between the distribution of attribute profiles $\mathbf{p}$ and the response distribution:

$$T(Q, \mathbf{c}, \mathbf{g})\mathbf{p} = \begin{pmatrix} 1 \\ P(R_1 = 1 | Q, \mathbf{c}, \mathbf{g}, \mathbf{p}) \\ \vdots \\ P(R_J = 1 | Q, \mathbf{c}, \mathbf{g}, \mathbf{p}) \\ P(R_1 = 1, R_2 = 1 | Q, \mathbf{c}, \mathbf{g}, \mathbf{p}) \\ \vdots \\ P(R_j = 1 \text{ for } j = 1, \ldots, J | Q, \mathbf{c}, \mathbf{g}, \mathbf{p}) \end{pmatrix}.$$

When referring to the ideal situation, where $\mathbf{s} = \mathbf{g} = \mathbf{0}$, we may omit the second and third arguments of $T(\cdot)$ to write $T(Q) = T(Q, \mathbf{1}, \mathbf{0})$.

4. Main Results

4.1. Identifiability Conditions

Through this paper, we assume that $\mathbf{c} \succ \mathbf{g}$, that $\mathbf{p} \succ \mathbf{0}$, and that the Q-matrix is pre-specified and correct. In addition, we list below five conditions that will be used in the upcoming identifiability theorems. It will be shown under various model assumptions that certain specific combinations of these conditions are either necessary and/or sufficient for the identifiability of the unknown parameters.

(C1) Q is complete. When this holds, we assume without loss of generality (WLOG) that the Q-matrix takes the following form:

$$Q = \begin{pmatrix} \mathcal{I}_K \\ Q' \end{pmatrix}. \tag{6}$$

(C2) Each attribute is required by at least two items.

(C3) Each attribute is required by at least three items.

(C4) Suppose Q has the structure defined in (6). For each $k \in \{1, \dots, K\}$, there must exist subsets S_k^+, S_k^- of the items in Q' such that S_k^+ and S_k^- have attribute requirements that are identical except in the kth attribute, which is required by an item in S_k^+ but not by any item in S_k^-.

Remark 1. In understanding condition C4, note that the attribute requirement of a set of items S is simply the set of attributes required to respond correctly to all items in S in the ideal case. When $S = \emptyset$, then no item must be answered correctly, so no attributes are required. Condition C4 may be satisfied if, for example, the Q-matrix contains two copies of the identity matrix i.e., Q can be written as (after some row permutation)

$$Q = \begin{pmatrix} \mathcal{I}_K \\ \mathcal{I}_K \\ Q'' \end{pmatrix},$$

where Q'' may be any structure. In this case

$$Q' = \begin{pmatrix} \mathcal{I}_K \\ Q'' \end{pmatrix}.$$

Then for every $k \in \{1, \dots, K\}$, we can take item set S_k^+ as $\{K + k\}$ and $S_k^- = \emptyset$. That is, the item set S_k^+ contains only the $K + k$th item in Q (equivalently, the kth item in Q'), which requires solely the kth attribute while the item set S_k^- requires no attributes. This implies that C4 is satisfied.

The Q-matrix would also fulfill condition C4 if there exists an item in Q' which requires $1 - \mathcal{I}_K$, i.e., the Q can be written as (after some row permutation)

$$Q = \begin{pmatrix} \mathcal{I}_K \\ 1 - \mathcal{I}_K \\ Q'' \end{pmatrix}.$$

Then for every $k \in \{1, \ldots, K\}$, we can take $S_k^+ = \{K+1, \ldots, 2K\}$ containing items $K+1, \ldots, 2K$ in the above Q-matrix which have structure $1 - \mathcal{I}_K$. We know S_k^+ requires all attributes. Furthermore, we take $S_k^- = \{K+k\}$, i.e., the $K+k$th item which requires all attributes except the kth one. Thus, S_k^+ and S_k^- have attribute requirements that are identical except in the kth attribute. This implies condition C4.

Remark 2. The following is an equivalent method of writing condition C4, which can be used directly to check whether a Q-matrix fulfills it or not. For each $k \in \{1, \ldots, K\}$, let $S_k^- = \{\ell : \exists\, j > K \text{ s.t. } q_{jk=0}, q_{j\ell} = 1\}$ be the set of attributes required by some item j in Q' not requiring the kth attribute. If there exists some item $j > K$ in Q' requiring the kth attribute and no attributes not in $S_k^- \cup \{k\}$ for every $k \in \{1, \ldots, K\}$, then condition C4 is fulfilled.

4.2. *Theorems*

We start with the simplest case, in which both the slipping and the guessing parameters are known.

Theorem 1. *Population proportion parameters* $\mathbf{p}$ *are identifiable only if condition C1 is satisfied. Moreover, condition C1 is sufficient when both the slipping and the guessing parameters are known.*

Theorem 1 states that when $\mathbf{s}$ and $\mathbf{g}$ are known, the completeness of the Q-matrix is a sufficient and necessary condition for the identifiability of $\mathbf{p}$. Completeness ensures that there is enough information in the response data for each attribute profile to have its own distinct ideal response vector. When a Q-matrix is incomplete, we can easily construct a non-identifiable example. For instance, consider the incomplete Q-matrix

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

The population parameter $\mathbf{p}$ is non-identifiable in this case even when $\mathbf{s}$ and $\mathbf{g}$ are known. Subjects with attribute profiles $\boldsymbol{\alpha}^1 = (1, 0)^\top$ and $\boldsymbol{\alpha}^2 = (0, 0)^\top$ have the same conditional response probabilities $P(\mathbf{r}|Q, \mathbf{c}, \mathbf{g}, \boldsymbol{\alpha})$, so weight can be transferred between $p_{\boldsymbol{\alpha}^1}$ and $p_{\boldsymbol{\alpha}^2}$ with no effect on the marginal probabilities $P(\mathbf{r}|Q, \mathbf{c}, \mathbf{g}, \mathbf{p})$, and thus no effect on the likelihood.

We now weaken our assumptions by taking only the guessing parameter $\mathbf{g}$ as known. Then stronger conditions are needed for identifiability; the necessary and sufficient conditions are given in Theorem 2 below.

Theorem 2. (Chen et al., 2014, Theorem 2) *Under the DINA model with known guessing parameter* $\mathbf{g}$*, the slipping parameter* $\mathbf{s}$ *and the population proportion parameter* $\mathbf{p}$ *are identifiable if and only if conditions C1 and C2 hold.*

Consider the Q-matrices

$$Q^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Q^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}. \tag{7}$$

From Theorem 2, we can see that when the guessing parameter $\mathbf{g}$ is known Q^1 describes a non-identifiable model while Q^2 describes an identifiable one.

Theorem 2 assumes $\mathbf{g}$ is known, which is not satisfied in many applications. In the most difficult setting, neither the slipping nor the guessing parameters are known. Then we have the following two results.

Theorem 3. (Necessary Conditions) *Under the DINA model, s, g and* $\mathbf{p}$ *are locally identifiable only if conditions C1 and C3 hold.*

Theorem 4. (Sufficient Conditions) *Suppose conditions C1 and C3 hold. Then* $\mathbf{s} = (s_1, \dots, s_J)$ *and* $\mathbf{g}^* = (g_{K+1}, \dots, g_J)$ *are identifiable. Moreover, if condition C4 also holds, then the model is fully identifiable.*

According to Theorem 3, neither Q^1 nor Q^2 from (7) describe identifiable DINA models when $\mathbf{s}$, $\mathbf{g}$, and $\mathbf{p}$ are all unknown. In order for the model to be identifiable, at the very least conditions C1 and C3 must hold. Consider the following four Q-matrices where both conditions hold:

$$Q^3 = \begin{pmatrix} \mathcal{I}_2 \\ \mathcal{I}_2 \\ \mathbf{1}^\top \end{pmatrix}, \quad Q^4 = \begin{pmatrix} \mathcal{I}_3 \\ 1 - \mathcal{I}_3 \end{pmatrix},$$

$$Q^5 = \begin{pmatrix} & \mathcal{I}_4 & & \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad Q^6 = \begin{pmatrix} \mathcal{I}_6 \\ \mathbf{1}^\top \\ \mathbf{1}^\top \end{pmatrix}.$$

The first two Q-matrices, Q^3 and Q^4, fulfill condition C4 in addition to condition C1 and C3 and describe identifiable models. Note that there is a gap between the sufficient and necessary conditions since condition C4 is not necessary, but conditions C1 and C3 are not sufficient. The Q-matrix Q^5, which describes an identifiable model but does not fulfill condition C4, is an example of the former; the Q-matrix Q^6, which fulfills both conditions C1 and C3 but represents a non-identifiable model, is an example of the latter. More specifically in regards to Q^6, the first part of Theorem 4 applies and $\mathbf{s}$, g_7, and g_8 are identifiable, while $(g_1, \dots, g_6)$ and $\mathbf{p}$ are not.

The main results (Theorems 3, 4) give relatively simple conditions to check the identifiability of the DINA model. Matlab code is available from the authors upon request. In practice, when researches find that the proposed Q-matrix does not satisfy the identifiability conditions, it is recommended to design new items such that the identifiability conditions are satisfied. Otherwise, the estimated parameters may be inconsistent as illustrated in the simulation study. Consider the fraction subtraction data (de la Torre & Douglas, 2004; Tatsuok, 1990) for an example. The data consist of responses to 20 items involving subtraction of fractions by 536 examinees. There are eight attributes: (1) convert a whole number to a fraction; (2) separate a whole number from a fraction; (3) simplify before subtracting; (4) find a common denominator; (5) borrow from whole number part; (6) column borrow to subtract the second numerator from the first; (7) subtract numerators; and (8) reduce answers to simplest form. The Q-matrix is given in Table 1. We can see that the Q-matrix is incomplete, which makes the model parameters unidentifiable. In particular, it is easy to see that attribute profiles $\mathbf{0}$ and $\mathbf{e}_1$ have the same conditional response probabilities and therefore $p_{\mathbf{0}}$ and $p_{\mathbf{e}_1}$ are unidentifiable; more discussions on the nonidentifiability of this data set can be found in DeCarlo (2011) and Zhang, DeCarlo, and Ying (2013).

The above discussion of the identifiability is from a frequentist perspective. Similarly, if Bayesian estimation method is employed, such as MCMC methods, the proposed identifiability conditions ensure the convergence of the estimates and the interpretability. When the identifiability conditions are not satisfied, the Bayesian estimators may heavily depend on the prior information.

TABLE 1.
The Q-matrix specified in de la Torre and Douglas (2004).

Item ID	Content	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
1	$\frac{5}{3}-\frac{3}{4}$	0	0	0	1	0	1	1	0
2	$\frac{3}{4}-\frac{3}{8}$	0	0	0	1	0	0	1	0
3	$\frac{5}{6}-\frac{1}{9}$	0	0	0	1	0	0	1	0
4	$3\frac{1}{2}-2\frac{3}{2}$	0	1	1	0	1	0	1	0
5	$4\frac{3}{5}-3\frac{4}{10}$	0	1	0	1	0	0	1	1
6	$\frac{6}{7}-\frac{4}{7}$	0	0	0	0	0	0	1	0
7	$3-2\frac{1}{5}$	1	1	0	0	0	0	1	0
8	$\frac{2}{3}-\frac{2}{3}$	0	0	0	0	0	0	1	0
9	$3\frac{7}{8}-2$	0	1	0	0	0	0	0	0
10	$4\frac{4}{12}-2\frac{7}{12}$	0	1	0	0	1	0	1	1
11	$4\frac{1}{3}-2\frac{4}{3}$	0	1	0	0	1	0	1	0
12	$\frac{11}{8}-\frac{1}{8}$	0	0	0	0	0	0	1	1
13	$3\frac{3}{8}-2\frac{5}{6}$	0	1	0	1	1	0	1	0
14	$3\frac{4}{5}-3\frac{2}{5}$	0	1	0	0	0	0	1	0
15	$2-\frac{1}{3}$	1	0	0	0	0	0	1	0
16	$4\frac{5}{7}-1\frac{4}{7}$	0	1	0	0	0	0	1	0
17	$7\frac{3}{5}-\frac{4}{5}$	0	1	0	0	1	0	1	0
18	$4\frac{1}{10}-2\frac{8}{10}$	0	1	0	0	1	1	1	0
19	$4-1\frac{4}{3}$	1	1	1	0	1	0	1	0
20	$4\frac{1}{3}-1\frac{5}{3}$	0	1	1	0	1	0	1	0

5. Simulations

In this section, we conduct simulation studies to illustrate the results in the last section. We generate data from the DINA model under different Q-matrices and check the parameter estimators $\hat{\mathbf{s}}$, $\hat{\mathbf{g}}$, and $\hat{\mathbf{p}}$. All Q-matrices are designs for $K = 5$ attributes. In total, there are three different simulation settings, each of which is detailed below:

Setting A We begin with the complete Q-matrix $Q^A = \mathcal{I}_5$. This is the minimal Q-matrix necessary for completeness. The item parameters are randomly selected under the condition that $\mathbf{c} \succ \mathbf{g}$ and $\mathbf{p} \succ \mathbf{0}$:

$$\mathbf{s} = \big(0.08,\ 0.15,\ 0.23,\ 0.25,\ 0.20\big)^\top$$

and

$$\mathbf{g} = \big(0.08,\ 0.22,\ 0.16,\ 0.19,\ 0.14\big)^\top .$$

The population parameter is

$$\begin{aligned}\mathbf{p}\ = (&0.018,\ 0.028,\ 0.025,\ 0.033,\ 0.039,\ 0.025,\ 0.072,\ \ 0.026,\\ &0.037,\ \ \ 0.019,\ 0.042,\ 0.021,\ 0.025,\ 0.042,\ 0.067,\ \ 0.042,\\ &0.025,\ \ \ 0.059,\ 0.024,\ 0.034,\ 0.043,\ 0.013,\ 0.024,\ \ 0.029,\\ &0.019,\ \ \ 0.028,\ 0.016,\ 0.019,\ 0.062,\ 0.020,\ 0.015,\ 0.012)^\top\end{aligned}.$$

Setting B The second Q matrix is

$$Q^B = \begin{pmatrix}\mathcal{I}_5\\ \mathbf{1}_{1\times 5}\end{pmatrix}.$$

It fulfills both condition C1 and condition C2. The population parameter is the same as in Setting A, as are the item parameters for the first five items. The sixth item has item parameters $s_6 = 0.18$ and $g_6 = 0.11$.

Setting C The third Q-matrix is

$$Q^C = \begin{pmatrix}\mathcal{I}_5\\ 1-\mathcal{I}_5\end{pmatrix}.$$

It fulfills conditions C1, C3, and C4. The population parameter is the same as in Setting A, as are the item parameters for the first five items. The last five items have slipping and guessing parameters

$$(s_6, \ldots, s_{10}) = (0.21, 0.09, 0.27, 0.24, 0.14)$$

and

$$(g_6, \ldots, g_{10}) = (0.23, 0.13, 0.18, 0.13, 0.11).$$

The consistency of the estimator $\hat{\mathbf{p}}$ is an important consequence of the identifiability of the model. By Theorem 1, the model in Setting A is identifiable if $\mathbf{s}$ and $\mathbf{g}$ are known. For Setting A, we generate data sets with different sample sizes (N from 1 to 500). For each data set, we use 1000 estimation trials with different initial values for the EM algorithm to estimate the model parameter $\mathbf{p}$. Figure 1 plots the L^2 error of the maximum likelihood estimates of $\mathbf{p}$ as the sample size N grows. The L^2 error is the Euclidean distance between the estimate and the true value. The mean curve is the average L^2 error over 1000 estimation trials for the same data set. We can see that it converges to 0 as the sample size increases, which provides numerical evidence of the identifiability of $\mathbf{p}$. To further show the consistency, we present one randomly selected trial ("trial 1" in Figure 1) and it can be seen that the estimation curve is consistent with the mean curve. All other estimation curves show similar consistency and are not presented here.

As stated in Theorem 2, Q^A is associated with a model that is non-identifiable when only $\mathbf{g}$ is known. Then, even for very large sample sizes parameter estimates will not converge to the

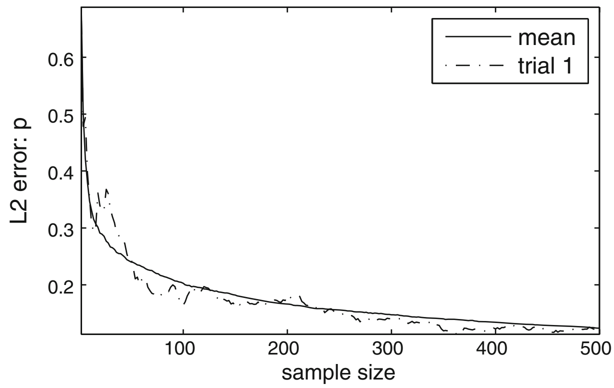

FIGURE 1.
L^2 error for estimates of **p** under Setting A, when **s** and **g** are known.

truth. This occurs because multiple sets of parameters maximize the same marginal likelihood. Shown below are two distinct sets of parameters that both maximize the likelihood for a random response matrix generated under Setting A for 500,000 individuals. They were obtained by running multiple repetitions of the EM algorithm in order to obtain true global maximizers. For the first set of estimates,

$$\begin{aligned}\hat{\mathbf{p}} = (&0.013, & 0.014, 0.045, 0.044, 0.011, 0.010, 0.095, & 0.030\\ &0.021, & 0.014, 0.080, 0.034, 0.004, 0.021, 0.100, & 0.051\\ &0.019, & 0.037, 0.039, 0.040, 0.025, 0.000, 0.034, & 0.028\\ &0.021, & 0.018, 0.029, 0.025, 0.052, 0.012, 0.020, & 0.015)^\top\end{aligned}$$

and

$$\hat{\mathbf{s}} = \big(0.03,\ 0.19,\ 0.17,\ 0.42,\ 0.11\big)^\top.$$

In the second set, the estimates are

$$\begin{aligned}\hat{\mathbf{p}} = (&0.009, & 0.007, 0.005, 0.016, 0.024, 0.018, 0.030, & 0.007\\ &0.039, & 0.027, 0.034, 0.019, 0.002, 0.055, 0.037, & 0.035\\ &0.015, & 0.122, 0.015, 0.046, 0.041, 0.024, 0.013, & 0.031\\ &0.040, & 0.079, 0.015, 0.035, 0.085, 0.048, 0.011, & 0.018)^\top\end{aligned}$$

and

$$\hat{\mathbf{s}} = \big(0.34,\ 0.26,\ 0.13,\ 0.05,\ 0.35\big)^\top.$$

Neither estimate is close to the true parameters. Both sets of parameters give the same loglikelihood of -1.7181×10^6 and produce identical response probabilities. In fact, the proof of Theorem 2 shows that there are an infinite number of such parameter sets. Moreover, no matter how large the sample size may grow, the estimates will never converge to the true value.

The DINA model is identifiable when only **g** is known only if condition C2 is fulfilled, in addition to condition C1. The Q-matrix Q^B from Setting B fulfills both conditions, and when

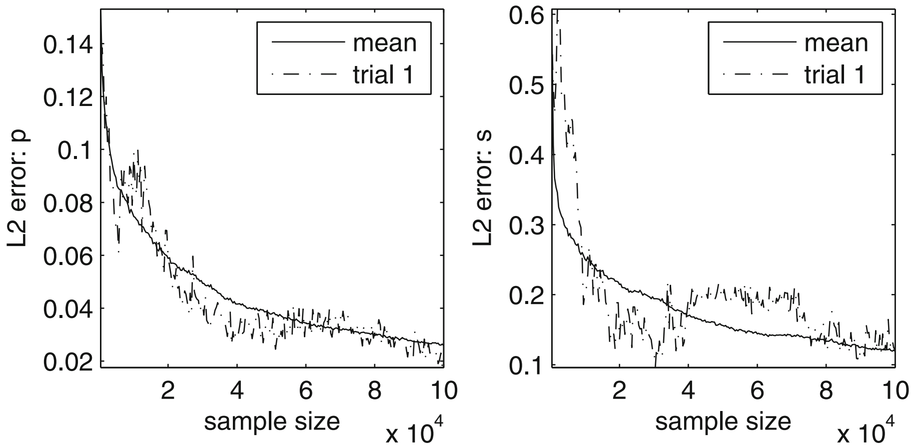

FIGURE 2.
L^2 error for estimates of $\mathbf{p}$ (*left*) and $\mathbf{s}$ (*right*) under Setting B, with $\mathbf{g}$ known.

$\mathbf{g}$ is known, the maximum likelihood estimates of both $\mathbf{s}$ and $\mathbf{p}$ will be consistent. The L^2 error of both estimates is plotted in Figure 2, where the mean L^2 error over 1000 simulations and one randomly selected trial are presented. We can see that the estimates are consistent.

When $\mathbf{g}$ is unknown then non-identifiability occurs and the maximum likelihood estimates are not guaranteed to converge. Two distinct sets of parameters that both maximize the likelihood for a random set of responses generated for $N = 500,000$ individuals are displayed below (assuming $\mathbf{g}$ is unknown). For the first set of estimates,

$$\begin{aligned}\hat{\mathbf{p}} &= (0.017,\ 0.019,\ 0.017,\ 0.015,\ 0.048,\ 0.019,\ 0.068,\ \ 0.006,\\ &\quad 0.060,\quad 0.014,\ 0.045,\ 0.015,\ 0.067,\ 0.057,\ 0.124,\ \ 0.019,\\ &\quad 0.022,\quad 0.039,\ 0.012,\ 0.010,\ 0.048,\ 0.006,\ 0.011,\ \ 0.018,\\ &\quad 0.025,\quad 0.037,\ 0.016,\ 0.011,\ 0.114,\ 0.008,\ 0.001,\ 0.013)^\top\\ \hat{\mathbf{s}} &= \left(0.15,\ 0.01,\ 0.08,\ 0.28,\ 0.28,\ 0.11\right)^\top.\end{aligned}$$

and

$$\hat{\mathbf{g}} = \left(0.09,\ 0.17,\ 0.25,\ 0.26,\ 0.20,\ 0.22\right)^\top.$$

In the second set,

$$\begin{aligned}\hat{\mathbf{p}} &= (0.016,\ 0.034,\ 0.032,\ 0.061,\ 0.026,\ 0.020,\ 0.074,\ \ 0.035,\\ &\quad 0.029,\ 0.014,\ 0.045,\ 0.028,\ 0.004,\ 0.028,\ 0.041,\ \ 0.043,\\ &\quad 0.026,\ 0.079,\ 0.038,\ 0.083,\ 0.032,\ 0.000,\ 0.032,\ \ 0.028,\\ &\quad 0.012,\ 0.017,\ 0.019,\ 0.023,\ 0.037,\ 0.019,\ 0.018,\ 0.010)^\top\\ \hat{\mathbf{s}} &= \left(0.04,\ 0.33,\ 0.30,\ 0.04,\ 0.04,\ 0.11\right)^\top.\end{aligned}$$

and

$$\hat{\mathbf{g}} = \left(0.09,\ 0.17,\ 0.26,\ 0.26,\ 0.21,\ 0.00\right)^\top.$$

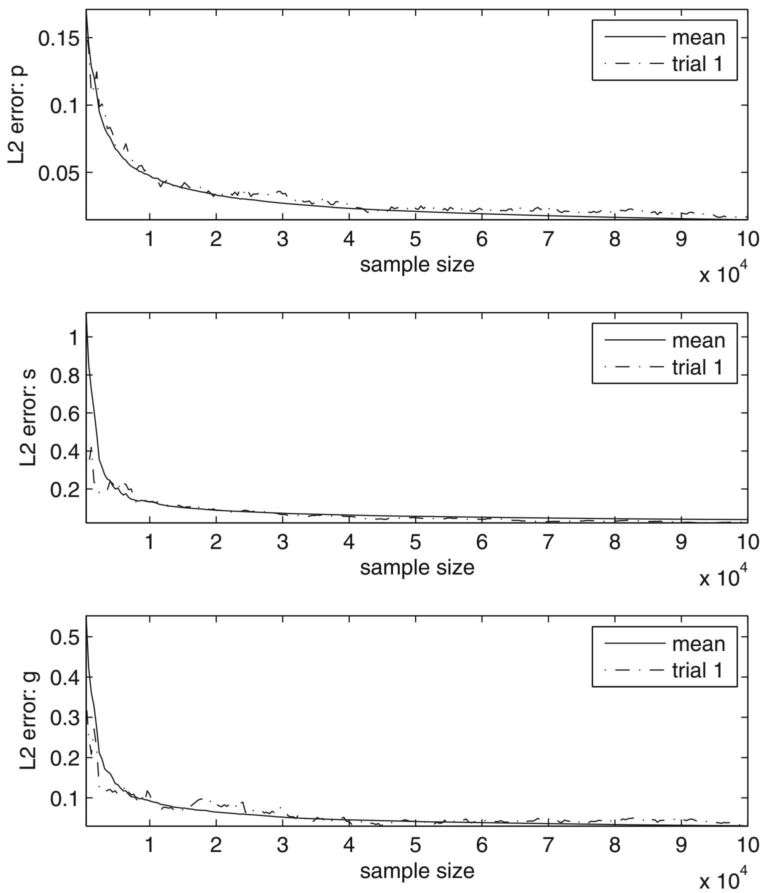

FIGURE 3.
L^2 error for estimates of **p** (*top*), **s** (*middle*), and **g** (*bottom*) under Setting C, when all parameters are unknown.

Neither estimate is close to the true parameters. Note that since there are only two items measuring each attribute, Condition C3 does not hold. The simulation does not fulfill even the first set of sufficiency conditions in Theorem 4, and neither **s** nor g_6 are estimated consistently. Both sets of parameters give the same loglikelihood of -1.8961×10^6 and produce identical response probabilities.

When the Q-matrix fulfills Condition C4, parameters can be estimated consistently, as shown in the last set of simulations. The Q-matrix Q^C from Setting C fulfills Condition C4, in addition to Conditions C1 and C3. The plots in Figure 3 show the convergence of the maximum likelihood estimates of **p**, **s**, and **g** under Setting C.

6. Discussion

Identification is a serious problem in diagnostic classification modeling. The creation of assessments where all parameters are identified is essential in ensuring proper statistical inference and interpretable parameters. The work here lays out some relatively simple conditions to check the identifiability of the DINA model. In particular, the final sufficiency conditions are less restrictive than those suggested by the conventional wisdom that requires, for each attribute, three items devoted solely to measuring that attribute.

In this paper, the attribute profile is modeled using a saturated model with $2^K - 1$ attribute profile parameters. It would be interesting to consider the identifiability conditions under the unsaturated models. For instance, consider the independent model case where mastering each attribute is assumed to be independent and follow a Bernoulli distribution. Specifically, let p_k be the probability of mastering the kth attribute; then the probability of having attribute profile $\boldsymbol{\alpha}$ is modeled as $p_{\boldsymbol{\alpha}} = \prod_{k=1}^{K}(p_k)^{\alpha_k}(1-p_k)^{(1-\alpha_k)}$. Note that the independent model has K attribute profile parameters. For this case, weaker conditions are expected for the identifiability of the model parameters. In particular, completeness of the Q-matrix may not be needed. We believe a similar technique as in the proof of the main results can be applied for such unsaturated models, and we would like to pursue this in the future.

The results can be easily extended to the DINO (deterministic input; noisy "or" gate) model (Templin & Henson, 2006) through the duality of the DINA and DINO models (Chen, 2014; Zhang et al., 2013). All the theorems apply directly, except for Theorem 4, which a requires slight modification. Since a DINO model with Q-matrix Q and item parameters $\mathbf{s}$ and $\mathbf{g}$ corresponds to a DINA model with the same Q-matrix but slipping parameter $1 - \mathbf{g}$ and guessing parameter $1 - \mathbf{s}$, in the DINO model $\mathbf{g}$ and $\mathbf{s}^* = (s_{K+1}, \ldots, s_J)$ are identifiable when Conditions C1 and C3 are fulfilled.

As far as other DCMs are concerned, identifiability remains an important issue. This is especially true with respect to broader models such as the General Diagnostic Model (von Davier, 2005), the log-linear diagnostic classification model (Henson, Templin & Willse, 2009), or the Generalized DINA Model (de la Torre, 2011). The sets of models which generate identically distributed data grows with the number of parameters, making such general models particularly difficult to work with. The exact identifiability requirements of these models remain a topic of study, though they are likely quite onerous in the most general models. In such cases, a useful alternative to simultaneous parameter estimation and respondent classification would involve separate rounds of item calibration and diagnostic testing.

The Q-matrix in this paper is assumed to be correctly specified. In practice, the Q-matrix is usually constructed by the users and may not be accurate. A misspecified Q-matrix could lead to substantial lack of fit and, consequently, erroneous classification of subjects (de la Torre, 2008; Rupp & Templin, 2008). Thus it is recommended to apply the proposed identifiability results after validating the constructed Q-matrix. To detect the misspecification of the Q-matrix, recently Liu et al. (2013) and Chen et al. (2014) provide the theoretical foundation by proving the identifiability of the Q-matrix from the response data under the DINA model. See also de la Torre (2008), DeCarlo (2012), and Chiu (2013) for the recent developments of Q-matrix misspecification detection methods. We would like to point out that the misspecification issue is closely related to the equivalent models studied in the literature (Bechger et al., 2002; Maris & Bechger, 2004; Maris & Bechger, 2009). In particular, statistical tests as in Bechger et al. (2002) may be used to test the specification of the Q-matrix. We leave this interesting topic for further study.

Appendix: Proof of Theorems

We begin with two important propositions necessary to prove the main results; their own proofs are postponed to the end of this section.

Recall that identifiability and local identifiability depend on the probability density function $f(x;\theta)$, which, when written as a function of the parameters θ becomes the likelihood $L(\theta)$.

Under the model specified in Section 2, given the full set of observations $R = \{\mathbf{R}^i : i = 1, \ldots, N\}$ and a Q-matrix Q, the likelihood of any set of parameters $\mathbf{c}, \mathbf{g}, \mathbf{p}$ can be written as

$$L(\mathbf{c}, \mathbf{g}, \mathbf{p}; R) = \prod_{i=1}^{n} P(\mathbf{R}^i | Q, \mathbf{c}, \mathbf{g}, \mathbf{p}) = \prod_{\mathbf{r} \in \{0,1\}^J} \pi_{\mathbf{r}}^{N_{\mathbf{r}}} \tag{8}$$

where $N_{\mathbf{r}} = |\{i \in \{1, \ldots, N\} : \mathbf{R}^i = \mathbf{r}\}|$ is the number of observations $\mathbf{R}^i$ equal to a particular response vector $\mathbf{r}$ and

$$\pi_{\mathbf{r}} = P(\mathbf{R} = \mathbf{r} | Q, \mathbf{c}, \mathbf{g}, \mathbf{p}) = \sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} \prod_{j=1}^{J} P(R_j = r_j | Q, c_j, g_j, \boldsymbol{\alpha}) \tag{9}$$

is the probability of observing $\mathbf{r}$ given $Q, \mathbf{c}, \mathbf{g}, \mathbf{p}$. The conditional probability $P(R_j = r_j | Q, c_j, g_j, \boldsymbol{\alpha})$ may be expressed as

$$c_j^{r_j \xi_j(Q,\boldsymbol{\alpha})} g_j^{r_j(1-\xi_j(Q,\boldsymbol{\alpha}))} (1 - c_j)^{(1-r_j)\xi_j(Q,\boldsymbol{\alpha})} (1 - g_j)^{(1-r_j)(1-\xi_j(Q,\boldsymbol{\alpha}))}.$$

Defining the likelihood leads to the first of the two propositions, which ties the T-matrix to the likelihood:

Proposition 1. *For two sets of parameters* $(\hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}})$ *and* $(\bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}})$*,*

$$L(\hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}}; R) = L(\bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}}; R)$$

for all observation matrices R if and only if the following equation holds:

$$T(Q, \hat{\mathbf{c}}, \hat{\mathbf{g}})\hat{\mathbf{p}} = T(Q, \bar{\mathbf{c}}, \bar{\mathbf{g}})\bar{\mathbf{p}}. \tag{10}$$

The second proposition describes the linear relationship between certain pairs of T-matrices.

Proposition 2. *There exists a matrix $D(\mathbf{g}^*)$ depending solely on $\mathbf{g}^* = (g_1^*, \ldots, g_J^*)$, such that for any $\mathbf{g}^* \in \mathbb{R}^J$,*

$$T(Q, \mathbf{c} - \mathbf{g}^*, \mathbf{g} - \mathbf{g}^*) = D(\mathbf{g}^*) T(Q, \mathbf{c}, \mathbf{g}).$$

The matrix $D(\mathbf{g}^*)$ is always lower triangular with diagonal $diag(D(\mathbf{g}^*)) = \mathbf{1}$, and thus invertible.

The main idea of the proofs is based on the result of Proposition 1. In particular, to show the identifiability, we only need to show that for two sets of parameters $(\hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}})$ and $(\bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}})$ satisfying (10), $(\hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}}) = (\bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}})$. On the other hand, by the definition, if there exist $(\hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}}) \neq (\bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}})$ and (10) holds, then the model is unidentifiable. We now prove our main theorems and propositions.

Proof of Theorem 1. The case where $\mathbf{s} = \mathbf{g} = \mathbf{0}$ was shown by (Chiu, 2009). For general known $\mathbf{s}$ and $\mathbf{g}$, by Proposition 1, $\mathbf{p}$ is unidentifiable when $Q, \mathbf{c}, \mathbf{g}$ are known iff there exists $\hat{\mathbf{p}}, \bar{\mathbf{p}} \in \mathbb{R}_+^{2^K}$ such that

$$T(Q, \mathbf{c}, \mathbf{g})\hat{\mathbf{p}} = T(Q, \mathbf{c}, \mathbf{g})\bar{\mathbf{p}}.$$

This occurs iff $T(Q, \mathbf{c}, \mathbf{g})$ is not a full-rank matrix.
Suppose that the Q-matrix is not complete. WLOG, we assume that the row vector corresponding to the first attribute is missing, i.e., $\mathbf{e}_1^\top \notin \mathcal{R}_Q$. Then, in the T-matrix, the columns corresponding to attribute profiles $\mathbf{0}$ and $\mathbf{e}_1$ are both equal to

$$\left(1, g_1, \ldots, g_J, g_1 g_2, \ldots, g_1 \cdots g_j\right)^\top,$$

and $\operatorname{rank}(T(Q, \mathbf{c}, \mathbf{g})) < 2^K$.
When Q is complete, assume WLOG that $Q_{1:K} = I_K$. The matrix $T(Q, \mathbf{c}, \mathbf{g})$ is full-rank iff $T(Q, \mathbf{c} - \mathbf{g}, \mathbf{0})$ is full-rank, since, by Proposition 2, $T(Q, \mathbf{c} - \mathbf{g}, \mathbf{0}) = D(\mathbf{g})T(Q, \mathbf{c}, \mathbf{g})$ and $D(\mathbf{g})$ is invertible. Consider the rows of $T(Q, \mathbf{c} - \mathbf{g}, \mathbf{0})$ corresponding to combinations of the first K items, i.e., $\mathbf{r} \in \{0, 1\}^J$ s.t. $r_j = 0$ for all $j > K$. This constitutes an upper-triangular submatrix of size $2^K \times 2^K$ with diagonal entries $\prod_{j:r_j=1}(c_j - g_j) \neq 0$. Thus, $T(Q, \mathbf{c} - \mathbf{g}, \mathbf{0})$ is full-rank, and $\mathbf{p}$ is identifiable. □

Proof of Theorem 2. Theorem 2 has been recently proved in Chen et al. (2014). For completeness, we include a proof under the setting of this paper. When $\mathbf{g}$ is known, we may combine Propositions 1 and 2 to show that two sets of parameters $(\hat{\mathbf{c}}, \mathbf{g}, \hat{\mathbf{p}})$ and $(\bar{\mathbf{c}}, \mathbf{g}, \bar{\mathbf{p}})$ produce equal likelihoods iff

$$\begin{aligned} T(Q, \hat{\mathbf{c}} - \mathbf{g}, \mathbf{0})\hat{\mathbf{p}} &= D(\mathbf{g})T(Q, \hat{\mathbf{c}}, \mathbf{g})\hat{\mathbf{p}} \\ &= D(\mathbf{g})T(Q, \bar{\mathbf{c}}, \mathbf{g})\bar{\mathbf{p}} = T(Q, \bar{\mathbf{c}} - \mathbf{g}, \mathbf{0})\bar{\mathbf{p}}. \end{aligned}$$

Note that $c_j \in (g_j, 1] \Leftrightarrow c_j - g_j \in (0, 1 - g_j]$.

Sufficiency For each item $j \in \{1, \ldots, J\}$, condition C2 implies that there exists some set of items $S^j \subset \{1, \ldots, J\}$ s.t. $j \notin S^j$ and the attributes required by item j are a subset of the attributes required by the items in S^j; then the sets of attributes required by items in S^j and by items in $S^j \cup \{j\}$ are identical. Mathematically, there exists $\mathbf{r}^j \in \{0, 1\}^J$ s.t. $r_j^j = 0$ and

$$T_{\mathbf{r}^j}(Q, \mathbf{1}, \mathbf{0}) = T_{\mathbf{r}^j + \mathbf{e}_j}(Q, \mathbf{1}, \mathbf{0}).$$

To find $\mathbf{r}^j$ for each item j, first suppose that $j \in \{1, \ldots, K\}$. Then $Q_j = \mathbf{e}_j^\top$ and there is some $j' \in \{K + 1, \ldots, J\}$ s.t. $q_{j'j} = 1$. Let $\mathbf{r}^j = \mathbf{e}_{j'}$. Otherwise, when $j \in \{K + 1, \ldots, J\}$, let $\mathbf{r}^j = \sum_{\{\ell: q_{j\ell}=1\}} \mathbf{e}_\ell$.
Then given any two sets of parameters $(\hat{\mathbf{c}}, \mathbf{0}, \hat{\mathbf{p}})$ and $(\bar{\mathbf{c}}, \mathbf{0}, \bar{\mathbf{p}})$ s.t. $T(Q, \hat{\mathbf{c}}, \mathbf{0})\hat{\mathbf{p}} = T(Q, \bar{\mathbf{c}}, \mathbf{0})\bar{\mathbf{p}}$,

$$\hat{c}_j = \frac{T_{\mathbf{e}_j + \mathbf{r}^j}(Q, \hat{\mathbf{c}}, \mathbf{0})\hat{\mathbf{p}}}{T_{\mathbf{r}^j}(Q, \hat{\mathbf{c}}, \mathbf{0})\hat{\mathbf{p}}} = \frac{T_{\mathbf{e}_j + \mathbf{r}^j}(Q, \bar{\mathbf{c}}, \mathbf{0})\bar{\mathbf{p}}}{T_{\mathbf{r}^j}(Q, \bar{\mathbf{c}}, \mathbf{0})\bar{\mathbf{p}}} = \bar{c}_j.$$

Thus, $\hat{\mathbf{c}} = \bar{\mathbf{c}}$; then, by Theorem 1, $\hat{\mathbf{p}} = \bar{\mathbf{p}}$.

Necessity By Theorem 1, condition C1 is necessary. Suppose condition C2 fails to hold. WLOG, it fails to hold for the first attribute and $q_{j1} = 0$ for all $j \neq 1$. Consider any set of parameters $(\hat{\mathbf{c}}, \hat{\mathbf{p}})$ s.t. $\hat{c}_j \in (g_j, 1]$ for all $j \in \{1, \ldots, J\}$ and $\hat{\mathbf{p}} \in (0, 1)^{2^K}$, $\sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} = 1$. There exists $\bar{c}_1$ close enough to $\hat{c}_1$ so that $\bar{c}_1 \in (g_1, 1]$ and $\bar{p}_{\boldsymbol{\alpha}} \in (0, 1)$ for all $\boldsymbol{\alpha} \in \{0, 1\}^K$, where

$$\bar{p}_{\boldsymbol{\alpha}} = \begin{cases} (\hat{c}_1/\bar{c}_1)\hat{p}_{\boldsymbol{\alpha}} & \alpha_1 = 1 \\ \hat{p}_{\boldsymbol{\alpha}} + \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1}(1 - \hat{c}_1/\bar{c}_1) & \alpha_1 = 0 \end{cases}.$$

Then, for any $\mathbf{r} \in \{0, 1\}^J$ s.t. $r_1 = 0$, $T_{\mathbf{r}}(Q, \hat{\mathbf{c}}, \mathbf{0}) = T_{\mathbf{r}}(Q, \bar{\mathbf{c}}, \mathbf{0})$ and

$$\begin{aligned} T_{\mathbf{r}}(Q, \hat{\mathbf{c}}, \mathbf{0})\hat{\mathbf{p}} &= \sum_{\{\boldsymbol{\alpha}:\alpha_1=0\}} t_{\mathbf{r},\boldsymbol{\alpha}}(Q, \hat{\mathbf{c}}, \mathbf{0})(\hat{p}_{\boldsymbol{\alpha}} + \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1}) \\ &= \sum_{\{\boldsymbol{\alpha}:\alpha_1=0\}} t_{\mathbf{r},\boldsymbol{\alpha}}(Q, \bar{\mathbf{c}}, \mathbf{0}))(\bar{p}_{\boldsymbol{\alpha}} + \bar{p}_{\boldsymbol{\alpha}+\mathbf{e}_1}) = T_{\mathbf{r}}(Q, \bar{\mathbf{c}}, \mathbf{0})\bar{\mathbf{p}}. \end{aligned}$$

Otherwise, $r_1 = 1$ and

$$\begin{aligned} T_{\mathbf{r}}(Q, \hat{\mathbf{c}}, \mathbf{0})\hat{\mathbf{p}} &= \sum_{\boldsymbol{\alpha}:\alpha_1=1} t_{\mathbf{r}-\mathbf{e}_1,\boldsymbol{\alpha}}(Q, \hat{\mathbf{c}}, \mathbf{0})\hat{c}_1 \hat{p}_{\boldsymbol{\alpha}} \\ &= \sum_{\boldsymbol{\alpha}:\alpha_1=1} t_{\mathbf{r}-\mathbf{e}_1,\boldsymbol{\alpha}}(Q, \bar{\mathbf{c}}, \mathbf{0})\bar{c}_1 \bar{p}_{\boldsymbol{\alpha}} = T_{\mathbf{r}}(Q, \bar{\mathbf{c}}, \mathbf{0})\bar{\mathbf{p}}. \end{aligned}$$

Thus we have found distinct sets of parameters satisfying (10), and shown that condition C2 is necessary. □

Proof of Theorem 3. Thanks to Theorems 1 and 2, conditions C1 and C2 are necessary for identifiability. We now show the necessity of condition C3. Suppose C3 does not hold, but C1 and C2 do. Then all attributes are required by at least two items and there exists an attribute such that it is only required by two items. WLOG, this is the first attribute.

When both items requiring the first attribute require only the first attribute, the Q-matrix can be written WLOG as

$$Q = \begin{pmatrix} 1 & \mathbf{0}^\top \\ 1 & \mathbf{0}^\top \\ \mathbf{0} & Q' \end{pmatrix}.$$

As was done for r_1 in the proof of necessity for Theorem 2, consider each possible value of $(r_1, r_2) \in \{0, 1\}^2$ to conclude that, for any distinct sets of parameters $(\hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}})$ and $(\bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}})$, $T(Q, \hat{\mathbf{c}}, \hat{\mathbf{g}},)\hat{\mathbf{p}} = T(Q, \bar{\mathbf{c}}, \bar{\mathbf{g}})\bar{\mathbf{p}}$ if for every $\boldsymbol{\alpha} \in \{0, 1\}^K$ s.t. $\alpha_1 = 0$,

$$\begin{cases} \hat{p}_{\boldsymbol{\alpha}} + \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} = \bar{p}_{\boldsymbol{\alpha}} + \bar{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} & (r_1, r_2) = (0, 0) \\ \hat{c}_1 \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \hat{g}_1 \hat{p}_{\boldsymbol{\alpha}} = \bar{c}_1 \bar{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \bar{g}_1 \bar{p}_{\boldsymbol{\alpha}} & (r_1, r_2) = (1, 0) \\ \hat{c}_2 \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \hat{g}_2 \hat{p}_{\boldsymbol{\alpha}} = \bar{c}_2 \bar{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \bar{g}_2 \bar{p}_{\boldsymbol{\alpha}} & (r_1, r_2) = (0, 1) \\ \hat{c}_1 \hat{c}_2 \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \hat{g}_1 \hat{g}_2 \hat{p}_{\boldsymbol{\alpha}} = \bar{c}_1 \bar{c}_2 \bar{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \bar{g}_1 \bar{g}_2 \bar{p}_{\boldsymbol{\alpha}} & (r_1, r_2) = (1, 1) \end{cases}. \tag{11}$$

Otherwise, the Q-matrix can be written WLOG as

$$Q = \begin{pmatrix} 1 & \mathbf{0}^\top \\ 1 & \mathbf{v}^\top \\ \mathbf{0} & Q' \end{pmatrix},$$

where $\mathbf{v}$ is a $(K-1)$-dimensional nonzero vector. Then $T(Q, \hat{\mathbf{c}}, \hat{\mathbf{g}},)\hat{\mathbf{p}} = T(Q, \bar{\mathbf{c}}, \bar{\mathbf{g}})\bar{\mathbf{p}}$ if

$$\begin{cases} \hat{p}_{\boldsymbol{\alpha}} + \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} = \bar{p}_{\boldsymbol{\alpha}} + \bar{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} & \forall\, \boldsymbol{\alpha} : \alpha_1 = 0 \\ \hat{c}_1 \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \hat{g}_1 \hat{p}_{\boldsymbol{\alpha}} = \bar{c}_1 \bar{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \bar{g}_1 \bar{p}_{\boldsymbol{\alpha}} & \forall\, \boldsymbol{\alpha} : \alpha_1 = 0 \\ \hat{c}_2 \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \hat{g}_2 \hat{p}_{\boldsymbol{\alpha}} = \bar{c}_2 \bar{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \bar{g}_2 \bar{p}_{\boldsymbol{\alpha}} & \forall\, \boldsymbol{\alpha} : \alpha_1 = 0, \boldsymbol{\alpha} \succeq (0\ \mathbf{v}^\top) \\ \hat{c}_1 \hat{c}_2 \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \hat{g}_1 \hat{g}_2 \hat{p}_{\boldsymbol{\alpha}} = \bar{c}_1 \bar{c}_2 \bar{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} + \bar{g}_1 \bar{g}_2 \bar{p}_{\boldsymbol{\alpha}} & \forall\, \boldsymbol{\alpha} : \alpha_1 = 0, \boldsymbol{\alpha} \succeq (0\ \mathbf{v}^\top) \end{cases}. \tag{12}$$

Since the equations in (12) are a subset of the equations in (11), finding sets of parameters $(\hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}})$ and $(\bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}})$ fulfilling (11) completes the proof for both types of Q-matrices.
Choose a valid set of parameters $(\hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}})$ s.t. $\hat{p}_{\boldsymbol{\alpha}}/\hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} = \rho$ is constant over all $\boldsymbol{\alpha} \in \{0,1\}^K$ s.t. $\alpha_1 = 0$. Then, for any $\bar{\mathbf{g}} \in \mathbb{R}^J$, setting

$$\bar{c}_j = \begin{cases} \bar{g}_1 + \frac{(\hat{c}_1 - \bar{g}_1)(\hat{c}_2 - \bar{g}_2) + \rho(\hat{g}_1 - \bar{g}_1)(\hat{g}_2 - \bar{g}_2)}{(\hat{c}_2 - \bar{g}_2) + \rho(\hat{g}_2 - \bar{g}_2)}, & j = 1 \\ \bar{g}_2 + \frac{(\hat{c}_1 - \bar{g}_1)(\hat{c}_2 - \bar{g}_2) + \rho(\hat{g}_1 - \bar{g}_1)(\hat{g}_2 - \bar{g}_2)}{(\hat{c}_1 - \bar{g}_1) + \rho(\hat{g}_1 - \bar{g}_1)}, & j = 2 \\ \hat{c}_j, & j = 3, \ldots, J \end{cases}$$

and setting

$$\begin{aligned} \bar{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} &= \frac{((\hat{c}_1 - \bar{g}_1) + \rho(\hat{g}_1 - \bar{g}_1))((\hat{c}_2 - \bar{g}_2) + \rho(\hat{g}_2 - \bar{g}_2))}{(\hat{c}_1 - \bar{g}_1)(\hat{c}_2 - \bar{g}_2) + \rho(\hat{g}_1 - \bar{g}_1)(\hat{g}_2 - \bar{g}_2)} \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1}, \\ \bar{p}_{\boldsymbol{\alpha}} &= \hat{p}_{\boldsymbol{\alpha}} + \hat{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} - \bar{p}_{\boldsymbol{\alpha}+\mathbf{e}_1} \end{aligned}$$

for every $\boldsymbol{\alpha} \in \{0,1\}^K$ s.t. $\alpha_1 = 0$ results in a solution to (11). By continuity, there is $\bar{\mathbf{g}}$ sufficiently close to $\hat{\mathbf{g}}$ so that $\bar{\mathbf{c}}, \bar{\mathbf{g}} \in [0,1]^J$, $\mathbf{c} \succ \mathbf{g}$, and $\bar{\mathbf{p}} \succ \mathbf{0}$. Thus, the model is non-identifiable when condition C3 fails, making it a necessary condition. □

Proof of Theorem 4. Suppose that conditions C1 and C3 hold, and let $(\hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}})$ and $(\bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}})$ be two sets of parameters solving Equation (10). According to condition C1, there is an item requiring solely the kth attribute for each $k \in \{1, \ldots, K\}$. Moreover, by condition C3, there are also least two additional items requiring the kth attribute. We begin the proof of sufficiency by showing that for every k, there exists an item j requiring the kth attribute s.t. $\hat{g}_j = \bar{g}_j$. The case where all these items require solely the kth attribute and the case where at least one requires multiple attributes are treated separately.

Case 1 All items requiring the kth attribute require solely the kth attribute. WLOG, $k = 1$ and the first three rows of Q are as follows:

$$Q_{1:3} = \begin{pmatrix} 1 & \mathbf{0}^\top \\ 1 & \mathbf{0}^\top \\ 1 & \mathbf{0}^\top \end{pmatrix}.$$

By Proposition 2, $T(Q, \hat{\mathbf{c}}, \hat{\mathbf{g}})\hat{\mathbf{p}} = T(Q, \bar{\mathbf{c}}, \bar{\mathbf{g}})\bar{\mathbf{p}}$ iff

$$T(Q, \hat{\mathbf{c}} - \hat{\mathbf{g}}, \mathbf{0})\hat{\mathbf{p}} = T(Q, \bar{\mathbf{c}} - \hat{\mathbf{g}}, \bar{\mathbf{g}} - \hat{\mathbf{g}})\bar{\mathbf{p}}.$$

Then, since

$$\frac{T_{\mathbf{e}_1+\mathbf{e}_3}(\hat{\mathbf{c}} - \hat{\mathbf{g}}, \mathbf{0})\hat{\mathbf{p}}}{T_{\mathbf{e}_1}(\hat{\mathbf{c}} - \hat{\mathbf{g}}, \mathbf{0})\hat{\mathbf{p}}} = \hat{c}_3 - \hat{g}_3 = \frac{T_{\mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3}(\hat{\mathbf{c}} - \hat{\mathbf{g}}, \mathbf{0})\hat{\mathbf{p}}}{T_{\mathbf{e}_1+\mathbf{e}_2}(\hat{\mathbf{c}} - \hat{\mathbf{g}}, \mathbf{0})\hat{\mathbf{p}}},$$

we may conclude that

$$\frac{T_{\mathbf{e}_1+\mathbf{e}_3}(\bar{\mathbf{c}} - \hat{\mathbf{g}}, \bar{\mathbf{g}} - \hat{\mathbf{g}})\bar{\mathbf{p}}}{T_{\mathbf{e}_1}(\bar{\mathbf{c}} - \hat{\mathbf{g}}, \bar{\mathbf{g}} - \hat{\mathbf{g}})\bar{\mathbf{p}}} = \frac{T_{\mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3}(\bar{\mathbf{c}} - \hat{\mathbf{g}}, \bar{\mathbf{g}} - \hat{\mathbf{g}})\bar{\mathbf{p}}}{T_{\mathbf{e}_1+\mathbf{e}_2}(\bar{\mathbf{c}} - \hat{\mathbf{g}}, \bar{\mathbf{g}} - \hat{\mathbf{g}})\bar{\mathbf{p}}}.$$

Let $\tilde{\mathbf{c}} = \bar{\mathbf{c}} - \hat{\mathbf{g}}$ and let $\tilde{\mathbf{g}} = \bar{\mathbf{g}} - \hat{\mathbf{g}}$. In addition, let $\bar{p}_i = \sum_{\boldsymbol{\alpha}:\alpha_1=i} p_{\boldsymbol{\alpha}}$ for $i = 0, 1$. Then the previous equation may be written as

$$\frac{\tilde{g}_1\tilde{g}_3\bar{p}_0 + \tilde{c}_1\tilde{c}_3\bar{p}_1}{\tilde{g}_1\bar{p}_0 + \tilde{c}_1\bar{p}_1} = \frac{\tilde{g}_1\tilde{g}_2\tilde{g}_3\bar{p}_0 + \tilde{c}_1\tilde{c}_2\tilde{c}_3\bar{p}_1}{\tilde{g}_1\tilde{g}_2\bar{p}_0 + \tilde{c}_1\tilde{c}_2\bar{p}_1}$$

and

$$\tilde{g}_1\tilde{c}_1(\tilde{c}_2 - \tilde{g}_2)(\tilde{c}_3 - \tilde{g}_3)\bar{p}_0\bar{p}_1 = 0.$$

By assumption, $\bar{\mathbf{p}} \succ \mathbf{0}$, $\tilde{\mathbf{c}} \succ \tilde{\mathbf{g}}$, so $\hat{g}_1 = \bar{g}_1$ or $\bar{c}_1$. By symmetry, $\bar{g}_1 = \hat{g}_1$ or $\hat{c}_1$. If $\hat{g}_1 \neq \bar{g}_1$, then $\hat{c}_1 = \bar{g}_1$ and $\bar{c}_1 = \hat{g}_1$. This contradicts the assumption that $\hat{\mathbf{c}} \succ \hat{\mathbf{g}}$ and $\bar{\mathbf{c}} \succ \bar{\mathbf{g}}$. Thus $\hat{g}_1 = \bar{g}_1$.

Case 2 At least one item requiring the kth attribute requires multiple attributes. WLOG, $k = 1$ and

$$Q_{1:3} = \begin{pmatrix} 1 & 0 & \mathbf{0}^\top \\ 1 & 1 & \mathbf{v}^\top \\ 0 & 1 & \mathbf{0}^\top \end{pmatrix},$$

for some vector $\mathbf{v} \in \{0, 1\}^{K-2}$. We will show that $\hat{g}_2 = \bar{g}_2$.
Since

$$\frac{T_{\mathbf{e}_1+\mathbf{e}_2}(\hat{\mathbf{c}} - \hat{\mathbf{g}}, \mathbf{0})\bar{\mathbf{p}}}{T_{\mathbf{e}_2}(\hat{\mathbf{c}} - \hat{\mathbf{g}}, \mathbf{0})\bar{\mathbf{p}}} = \hat{c}_1 - \hat{g}_1 = \frac{T_{\mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3}(\hat{\mathbf{c}} - \hat{\mathbf{g}}, \mathbf{0})\bar{\mathbf{p}}}{T_{\mathbf{e}_2+\mathbf{e}_3}(\hat{\mathbf{c}} - \hat{\mathbf{g}}, \mathbf{0})\bar{\mathbf{p}}},$$

we know that

$$\frac{T_{\mathbf{e}_1+\mathbf{e}_2}(\bar{\mathbf{c}} - \hat{\mathbf{g}}, \bar{\mathbf{g}} - \hat{\mathbf{g}})\bar{\mathbf{p}}}{T_{\mathbf{e}_2}(\bar{\mathbf{c}} - \hat{\mathbf{g}}, \bar{\mathbf{g}} - \hat{\mathbf{g}})\bar{\mathbf{p}}} = \frac{T_{\mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3}(\bar{\mathbf{c}} - \hat{\mathbf{g}}, \bar{\mathbf{g}} - \hat{\mathbf{g}})\bar{\mathbf{p}}}{T_{\mathbf{e}_2+\mathbf{e}_3}(\bar{\mathbf{c}} - \hat{\mathbf{g}}, \bar{\mathbf{g}} - \hat{\mathbf{g}})\bar{\mathbf{p}}}.$$

Thus,

$$\frac{\tilde{g}_1\tilde{g}_2\bar{p}_{0,0} + \tilde{c}_1\tilde{g}_2\bar{p}_{1,0} + \tilde{g}_1\tilde{g}_2\bar{p}_{0,1} + \tilde{c}_1\tilde{c}_2\bar{p}_{1,1}}{\tilde{g}_2\bar{p}_{0,0} + \tilde{g}_2\bar{p}_{1,0} + \tilde{g}_2\bar{p}_{0,1} + \tilde{c}_2\bar{p}_{1,1}}$$
$$= \frac{\tilde{g}_1\tilde{g}_3\tilde{g}_2\bar{p}_{0,0} + \tilde{c}_1\tilde{g}_3\tilde{g}_2\bar{p}_{1,0} + \tilde{g}_1\tilde{c}_3\tilde{g}_2\bar{p}_{0,1} + \tilde{c}_1\tilde{c}_3\tilde{c}_2\bar{p}_{1,1}}{\tilde{g}_3\tilde{g}_2\bar{p}_{0,0} + \tilde{g}_3\tilde{g}_2\bar{p}_{1,0} + \tilde{c}_3\tilde{g}_2\bar{p}_{0,1} + \tilde{c}_3\tilde{c}_2\bar{p}_{1,1}},$$

where $\bar{p}_{i,j} = \sum_{\boldsymbol{\alpha}:(\alpha_1,\alpha_2)=(i,j)} \bar{p}_{\boldsymbol{\alpha}}$ for $(i, j) \in \{0, 1\}^2$, $\tilde{g}_j = \bar{g}_j - \hat{g}_j$ for $j = 1, 2, 3$, $\tilde{c}_j = \bar{c}_j - \hat{c}_j$ for $j = 1, 3$, and

$$\tilde{c}_2 = \frac{(\bar{c}_2 - \hat{c}_2)\sum_{\boldsymbol{\alpha}:\boldsymbol{\alpha}\succeq Q_2} p_{\boldsymbol{\alpha}} + (\bar{g}_2 - \hat{g}_2)\sum_{\boldsymbol{\alpha}:\alpha_1=\alpha_2=1,\boldsymbol{\alpha}\nsucceq Q_2} \bar{p}_{\boldsymbol{\alpha}}}{\bar{p}_{1,1}}.$$

Cross-multiply and cancel to obtain that

$$\bar{p}_{0,1}\bar{p}_{1,0}(\tilde{c}_1 - \tilde{g}_1)\tilde{g}_2^2(\tilde{c}_3 - \tilde{g}_3) = \bar{p}_{0,0}\bar{p}_{1,1}(\tilde{c}_1 - \tilde{g}_1)\tilde{c}_2\tilde{g}_2(\tilde{c}_3 - \tilde{g}_3)$$

Now suppose that $\hat{g}_2 \neq \bar{g}_2$. Since $\tilde{c}_j > \tilde{g}_j$ for $j = 1, 2, 3$,

$$\bar{p}_{1,0}\bar{p}_{0,1}(\bar{g}_2 - \hat{g}_2) = \bar{p}_{0,0}\bar{p}_{1,1}(\bar{c}_2 - \hat{g}_2). \tag{13}$$

In addition, by symmetry,

$$\hat{p}_{1,0}\hat{p}_{0,1}(\hat{g}_2 - \bar{g}_2) = \hat{p}_{0,0}\hat{p}_{1,1}(\hat{c}_2 - \bar{g}_2), \tag{14}$$

where $\hat{p}_{i,j} = \sum_{\boldsymbol{\alpha}:(\alpha_1,\alpha_2)=(i,j)} \hat{p}_{\boldsymbol{\alpha}}$ for $(i, j) \in \{0, 1\}^2$.

Taken together, (13) and (14) imply that either $\hat{c}_3 > \hat{g}_3 > \bar{c}_3 > \bar{g}_3$ or $\bar{c}_3 > \bar{g}_3 > \hat{c}_3 > \hat{g}_3$. However, since $T_{\mathbf{e}_2}(\hat{\mathbf{c}}, \hat{\mathbf{g}})\hat{\mathbf{p}} = T_{\mathbf{e}_2}(\bar{\mathbf{c}}, \bar{\mathbf{g}})\bar{\mathbf{p}}$,

$$\hat{g}_2(\hat{p}_{0,0} + \hat{p}_{1,0} + \hat{p}_{0,1}) + \hat{c}_2 p_{1,1} = \bar{g}_2(\bar{p}_{0,0} + \bar{p}_{1,0} + \bar{p}_{0,1}) + \bar{c}_2 p_{1,1}.$$

This is a contradiction; thus $\hat{g}_2 = \bar{g}_2$.

WLOG, the Q-matrix can be written as

$$Q = \begin{pmatrix} \mathcal{I}_K \\ Q' \end{pmatrix}.$$

We have shown that for each $k \in \{1, \ldots, K\}$, there exists some item $j_k > K$ requiring the kth attribute s.t. $\hat{g}_{j_k} = \bar{g}_{j_k}$. For each item $j > K$, let $\mathbf{r}^j = \begin{pmatrix} Q_j^\top & \mathbf{0} \end{pmatrix}$ be the response vector selecting those among the first K items requiring attributes required by the jth item. Then $\mathbf{r}^j$ and $\mathbf{r}^j + \mathbf{e}_j$ denote distinct sets of items with identical attribute requirements and

$$\hat{c}_j - \hat{g}_j = \frac{T_{\mathbf{r}^j+\mathbf{e}_j}(Q, \hat{\mathbf{c}} - \hat{\mathbf{g}}, \mathbf{0})\hat{\mathbf{p}}}{T_{\mathbf{r}^j}(Q, \hat{\mathbf{c}} - \hat{\mathbf{g}}, \mathbf{0})\hat{\mathbf{p}}} = \frac{T_{\mathbf{r}^j+\mathbf{e}_j}(Q, \bar{\mathbf{c}} - \hat{\mathbf{g}}, \bar{\mathbf{g}} - \hat{\mathbf{g}})\bar{\mathbf{p}}}{T_{\mathbf{r}^j}(Q, \bar{\mathbf{c}} - \hat{\mathbf{g}}, \bar{\mathbf{g}} - \hat{\mathbf{g}})\bar{\mathbf{p}}}. = \bar{c}_j - \hat{g}_j$$

Thus, $\hat{c}_j = \bar{c}_j$ if $\hat{g}_j = \bar{g}_j$; by the proof of Case 2, this includes all items j requiring multiple attributes. Otherwise, $Q_j = \mathbf{e}_k$ for some $k \in \{1, \ldots, K\}$, and the response vectors $\mathbf{e}_j + \mathbf{e}_{j_k}$ and $\mathbf{e}_{j_k}$ represent distinct combinations of items with identical attribute requirements, so that

$$\begin{aligned}\hat{c}_j &= \frac{T_{\mathbf{e}_j+\mathbf{e}_{j_k}}(Q, \hat{\mathbf{c}} - \hat{g}_{j_k}\mathbf{e}_{j_k}, \hat{\mathbf{g}} - \hat{g}_{j_k}\mathbf{e}_{j_k})\hat{\mathbf{p}}}{T_{\mathbf{e}_{j_k}}(Q, \hat{\mathbf{c}} - \hat{g}_{j_k}\mathbf{e}_{j_k}, \hat{\mathbf{g}} - \hat{g}_{j_k}\mathbf{e}_{j_k})\hat{\mathbf{p}}} \\ &= \frac{T_{\mathbf{e}_j+\mathbf{e}_{j_k}}(Q, \bar{\mathbf{c}} - \hat{g}_{j_k}\mathbf{e}_{j_k}, \bar{\mathbf{g}} - \hat{g}_{j_k}\mathbf{e}_{j_k})\bar{\mathbf{p}}}{T_{\mathbf{e}_{j_k}}(Q, \bar{\mathbf{c}} - \hat{g}_{j_k}\mathbf{e}_{j_k}, \bar{\mathbf{g}} - \hat{g}_{j_k}\mathbf{e}_{j_k})\bar{\mathbf{p}}} = \bar{c}_j.\end{aligned}$$

Thus, $\hat{c}_j = \bar{c}_j$ for every $j \in \{1, \ldots, J\}$, i.e., $\hat{\mathbf{c}} = \bar{\mathbf{c}}$.

We now consider the identifiability of the remaining g_j. For each $j > K$ s.t. $Q_j = \mathbf{e}_k$ for some $k \in \{1, \ldots, K\}$, let $\mathbf{c}^* = \hat{c}_k \mathbf{e}_k + \hat{c}_j \mathbf{e}_j$. Then

$$\hat{g}_k - \hat{c}_k = \frac{T_{\mathbf{e}_k+\mathbf{e}_j}(\hat{\mathbf{c}} - \mathbf{c}^*, \hat{\mathbf{g}} - \mathbf{c}^*)\hat{\mathbf{p}}}{T_{\mathbf{e}_k}(\hat{\mathbf{c}} - \mathbf{c}^*, \hat{\mathbf{g}} - \mathbf{c}^*)\hat{\mathbf{p}}} = \frac{T_{\mathbf{e}_k+\mathbf{e}_j}(\bar{\mathbf{c}} - \mathbf{c}^*, \bar{\mathbf{g}} - \mathbf{c}^*)\bar{\mathbf{p}}}{T_{\mathbf{e}_k}(\bar{\mathbf{c}} - \mathbf{c}^*, \bar{\mathbf{g}} - \mathbf{c}^*)\bar{\mathbf{p}}} = \bar{g}_k - \hat{c}_k$$

and $\hat{g}_k = \bar{g}_k$. Thus g_j is identifiable for all $j > K$.

To show the identifiability of $g_1, \ldots, g_K$, for each $k \leq K$ let

$$\mathbf{r}^k = \sum_{j=K+1}^{J} \mathbf{e}_j (1 - q_{jk})$$

represent the set of items in Q' not requiring the kth attribute. When condition C4 holds, there is some item $\ell > K$ requiring the kth attribute and no other attributes not required by the set of items denoted by $\mathbf{r}^k$. Let $\mathbf{g}^* = (\hat{c}_1, \ldots, \hat{c}_k, \hat{g}_{K+1}, \ldots, \hat{g}_J)^\top$. Then, for any set of parameters $(\mathbf{c}, \mathbf{g}, \mathbf{p})$ s.t. $g_j = \hat{g}_j$ for all $j > K$,

$$T_{\mathbf{r}}(Q, \mathbf{c} - \mathbf{g}^*, \mathbf{g} - \mathbf{g}^*)\mathbf{p} = \left(\prod_{j=K+1}^{J} (c_j - \hat{g}_j)^{r_j} \right) \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} p_{\boldsymbol{\alpha}} t_{\mathbf{r},\boldsymbol{\alpha}}(Q)$$

for all response vectors $\mathbf{r}$ s.t. $r_j = 0$ for all $j \leq K$. Since $\hat{\mathbf{c}} = \bar{\mathbf{c}}$ and $\bar{g}_j = \hat{g}_j$ for all $j > K$, this implies that

$$\sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \hat{p}_{\boldsymbol{\alpha}} t_{\mathbf{r},\boldsymbol{\alpha}}(Q) = \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \bar{p}_{\boldsymbol{\alpha}} t_{\mathbf{r},\boldsymbol{\alpha}}(Q) \tag{15}$$

for all such $\mathbf{r}$. Consider the row of $T(Q, \mathbf{c} - \mathbf{g}^*, \mathbf{g} - \mathbf{g}^*)$ corresponding to the combination of the kth item with all the items denoted by $\mathbf{r}^k$. The entries of this row-vector are non-zero only for attribute profiles denoting mastery of the skills required by $\mathbf{r}^k$ and non-mastery of the kth attribute. Thus,

$$\begin{aligned} &T_{\mathbf{e}_k+\mathbf{r}^k}(Q, \mathbf{c} - \mathbf{g}^*, \mathbf{g} - \mathbf{g}^*)\mathbf{p} \\ &\quad = (g_k - \hat{c}_k) \left(\prod_{j=K+1}^{J} (c_j - \hat{g}_j)^{r_j^k} \right) \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} p_{\boldsymbol{\alpha}} (t_{\mathbf{r}^k,\boldsymbol{\alpha}}(Q) - t_{\mathbf{e}_k+\mathbf{r}^k,\boldsymbol{\alpha}}(Q)). \end{aligned}$$

When condition C4 holds, there is some $\mathbf{r}$ s.t. $r_j = 0$ for all $j \leq K$ and $T_{\mathbf{e}_k+\mathbf{r}^k}(Q) = T_{\mathbf{r}+\mathbf{r}^k}(Q)$. Then, by (15)

$$\sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \hat{p}_{\boldsymbol{\alpha}} (t_{\mathbf{r}^k}(Q) - t_{\mathbf{e}_k+\mathbf{r}^k,\boldsymbol{\alpha}}(Q)) = \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \bar{p}_{\boldsymbol{\alpha}} (t_{\mathbf{r}^k}(Q) - t_{\mathbf{e}_k+\mathbf{r}^k,\boldsymbol{\alpha}}(Q)).$$

Since $T_{\mathbf{e}_k+\mathbf{r}^k}(Q, \hat{\mathbf{c}} - \mathbf{g}^*, \hat{\mathbf{g}} - \mathbf{g}^*)\hat{\mathbf{p}} = T_{\mathbf{e}_k+\mathbf{r}^k}(Q, \bar{\mathbf{c}} - \mathbf{g}^*, \bar{\mathbf{g}} - \mathbf{g}^*)\bar{\mathbf{p}}$, it must be true that $\hat{g}_k = \bar{g}_k$. Thus, $\mathbf{g}$ is fully identifiable and by Theorem 1 so is $\mathbf{p}$. $\square$

Proof of Proposition 1. The observations follow a multinomial distribution over the set of possible responses $\mathbf{r} \in \{0, 1\}^J$, with probabilities $\pi_{\mathbf{r}}$ as defined in (9). For a particular $(\hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}})$ and $(\bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}})$,

$$\begin{aligned} L(\hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}}; R) &= \prod_{\mathbf{r}\in\{0,1\}^J} \pi_{\mathbf{r}}(Q, \hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}})^{N_{\mathbf{r}}} \\ &= \prod_{\mathbf{r}\in\{0,1\}^J} \pi_{\mathbf{r}}(Q, \bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}})^{N_{\mathbf{r}}} = L(\bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}}; R) \end{aligned}$$

for all observation matrices R iff $\pi_{\mathbf{r}}(Q, \hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}}) = \pi_{\mathbf{r}}(Q, \bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}})$ for all $\mathbf{r} \in \{0, 1\}^J$. Consider a P-matrix $P(Q, \mathbf{c}, \mathbf{g})$ indexed like the T-matrix by item subsets $\mathbf{r} \in \{0, 1\}^J$ and attribute profiles $\boldsymbol{\alpha} \in \{0, 1\}^K$. The entries of $P(Q, \mathbf{c}, \mathbf{g})$ are denoted by the quantities

$$p_{\mathbf{r},\boldsymbol{\alpha}}(Q, \mathbf{c}, \mathbf{g}) = P(\mathbf{R} = \mathbf{r}|Q, \mathbf{c}, \mathbf{g}).$$

Then the statement $\pi_{\mathbf{r}}(Q, \hat{\mathbf{c}}, \hat{\mathbf{g}}, \hat{\mathbf{p}}) = \pi_{\mathbf{r}}(Q, \bar{\mathbf{c}}, \bar{\mathbf{g}}, \bar{\mathbf{p}})$ for all $\mathbf{r} \in \{0, 1\}^J$ can be written in matrix notation as $P(Q, \hat{\mathbf{c}}, \hat{\mathbf{g}})\hat{\mathbf{p}} = P(Q, \bar{\mathbf{c}}, \bar{\mathbf{g}})\bar{\mathbf{p}}$. Since

$$t_{\mathbf{r},\boldsymbol{\alpha}}(Q, \mathbf{c}, \mathbf{g}) = P(\mathbf{R} \succeq \mathbf{r}|Q, \mathbf{c}, \mathbf{g}) = \sum_{\mathbf{r}\succeq\mathbf{r}} p_{\mathbf{r},\boldsymbol{\alpha}}(Q, \mathbf{c}, \mathbf{g})$$

there is a one-to-one linear transformation between $P(Q, \mathbf{c}, \mathbf{g})$ and $T(Q, \mathbf{c}, \mathbf{g})$ that is not dependent on Q, $\mathbf{c}$, $\mathbf{g}$, and

$$P(Q, \hat{\mathbf{c}}, \hat{\mathbf{g}})\hat{\mathbf{p}} = P(Q, \bar{\mathbf{c}}, \bar{\mathbf{g}})\bar{\mathbf{p}} \Leftrightarrow T(Q, \hat{\mathbf{c}}, \hat{\mathbf{g}})\hat{\mathbf{p}} = T(Q, \bar{\mathbf{c}}, \bar{\mathbf{g}})\bar{\mathbf{p}}.$$

□

Proof of the Proposition 2. In what follows, we construct such a D matrix satisfying the conditions in the proposition, i.e., $D(\mathbf{g}^*)$ is a matrix only depending on $\mathbf{g}^*$ such that $D_{\mathbf{g}^*}T(Q, \mathbf{c}, \mathbf{g}) = T_{\mathbf{c}-\mathbf{g}^*,\mathbf{g}-\mathbf{g}^*}(Q)$ for any Q, $\mathbf{c}$, $\mathbf{g}$. Recall that for any Q, $\mathbf{c}$, $\mathbf{g}$,

$$t_{\mathbf{r},\boldsymbol{\alpha}}(Q, \mathbf{c}, \mathbf{g}) = \prod_{j\in S} c_j^{\xi_j(Q,\boldsymbol{\alpha})} g_j^{1-\xi_j(Q,\boldsymbol{\alpha})} \ \forall\, \mathbf{r} \in \{0, 1\}^J, \quad \boldsymbol{\alpha} \in \{0, 1\}^K.$$

We may extend this definition to include $\mathbf{c}, \mathbf{g} \notin [0, 1]^M$, though in such cases the $t_{\mathbf{r},\boldsymbol{\alpha}}$ will no longer correspond to probabilities. Then for any $\mathbf{g}^* \in \mathbb{R}$,

$$t_{\mathbf{r},\boldsymbol{\alpha}}(Q, \mathbf{c} - \mathbf{g}^*, \mathbf{g} - \mathbf{g}^*) = \prod_{j:r_j=1} (c_j - g_j^*)^{\xi_j(Q,\boldsymbol{\alpha})}(g_j - g_j^*)^{1-\xi_j(Q,\boldsymbol{\alpha})} = \prod_{j:r_j=1} (b_j - g_j^*),$$

where $b_j = c_j^{\xi_j(Q,\boldsymbol{\alpha})} g_j^{1-\xi_j(Q,\boldsymbol{\alpha})} = t_{\mathbf{e}_j,\boldsymbol{\alpha}}(Q, \mathbf{c}, \mathbf{g})$. By polynomial expansion,

$$t_{\mathbf{r},\boldsymbol{\alpha}}(Q, \mathbf{c} - \mathbf{g}^*, \mathbf{g} - \mathbf{g}^*) = \sum_{\mathbf{r}'\preceq\mathbf{r}} (-1)^{\sum_{j=1}^J r_j - r_j'} \prod_{j:r_j-r_j'=1} g_j^* \prod_{k:r_k'=1} b_k.$$

Define the entries $d_{\mathbf{r},\mathbf{r}'}(\mathbf{g}^*)$ of $D(\mathbf{g}^*)$ as

$$d_{\mathbf{r},\mathbf{r}'}(\mathbf{g}^*) = \begin{cases} 0 & \mathbf{r}' \npreceq \mathbf{r} \\ (-1)^{\sum_{j=1}^{J} r_j - r'_j} \prod_{j: r_j - r'_j = 1} g^*_j & \mathbf{r}' \prec \mathbf{r} \\ 1 & \mathbf{r}' = \mathbf{r} \end{cases}.$$

Then

$$T(Q, \mathbf{c} - \mathbf{g}^*, \mathbf{g} - \mathbf{g}^*) = D(\mathbf{g}^*) T(Q, \mathbf{c}, \mathbf{g}),$$

where $D(\mathbf{g}^*)$ is a lower triangular matrix depending solely on $\mathbf{g}^*$ with eigenvalues equal to its diagonal. Since $\text{diag}(D(\mathbf{g}^*)) = \mathbf{1}$, $D(\mathbf{g}^*)$ is invertible. $\square$

References

Bechger, T. M., Verstralen, H. H., & Verhelst, N. D. (2002). Equivalent linear logistic test models. *Psychometrika*, *67*(1), 123–136.

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2014). Statistical analysis of q-matrix based diagnostic classification models. *Journal of the American Statistical Association* (to appear).

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

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

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.

de la Torre, J., & Douglas, J. A. (2004). Higher order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333–353.

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

DeCarlo, L. T. (2012). Recognizing uncertainty in the Q-matrix via a bayesian extension of the DINA model. *Applied Psychological Measurement*, *36*(6), 447–468.

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361–390). Hillsdale, NJ: Erlbaum Associates.

Gabrielsen, A. (1978). Consistency and identifiability. *Journal of Econometrics*, *8*(2), 261–263.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*, 215–231.

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.

Koopmans, T. C. (Ed.). (1950). *Statistical inference in dynamic economic models* (Vol. 10). New York: Wiley.

Koopmans, T. C., & Reiersøl, O. (1950). The identification of structural characteristics. *Annals of Mathematical Statistics*, *21*, 165–181.

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy model for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, *41*, 205–237.

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

Liu, J., Xu, G., & Ying, Z. (2013). Theory of self-learning Q-matrix. *Bernoulli*, *19*(5A), 1790–1817.

Maris, G., & Bechger, T. M. (2004). Equivalent MIRID models. *Psychometrika*, *69*(4), 627–639.

Maris, G., & Bechger, T. M. (2009). Equivalent diagnostic classification models. *Measurement*, *7*, 41–46.

McHugh, R. B. (1956). Efficient estimatio and local identification in latent class analysis. *Psychometrika*, *21*, 331–347.

Paulino, C. D. M., & de Bragançca Pereira, C. A. (1994). On identifiability of parametric statistical models. *Journal of the Italian Statistical Society*, *3*(1), 125–151.

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

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

San Martín, E., & Quintana, F. (2002). Consistency and identifiability revisited. *Brazilian Journal of Probability and Statistics*, *16*, 99–106.

San Martín, E., Rolin, J.-M., & Castro, L. M. (2013). Identification of the 1PL model with guessing parameter: Parametric and semi-parametric results. *Psychometrika*, *78*(2), 341–379.

Tatsuoka, C. (2009). Diagnostic models as partially ordered sets. *Measurement*, *7*, 49–53.

Tatsuoka, K. K. (1983). Rule space: An approch 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 misconceptions in the pattern classification approach. *Journal of Educational Statistics*, *12*, 55–73.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453–488). Hillsdale, NJ: Erlbaum.

Tatsuoka, K. K. (1991). *Boolean algebra applied to determination of universal set of knowledge states* (Research report No. RR-91-44-ONR). Princeton, NJ: Educational Testing Service.

Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. New York: Routledge.

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

U.S. Department of Education. (2009, November). Race to the top program executive summary. Washington, D.C. Retrieved May 9, 2014 from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf.

von Davier, M. (2005). *A general diagnostic model applied to language testing data (Research report)*. 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*, 287–307.

von Davier, M., & Yamamoto, K. (2004). A class of models for cognitive diagnosis. In *4th spearman conference*. Philadelphia, PA.

Zhang, S. S., DeCarlo, L. T., & Ying, Z. (2013, March). Non-identifiability, equivalence classes, and attribute-specific classification in Q-matrix based Cognitive Diagnosis Models. ArXiv e-prints.

Manuscript Received: 29 AUG 2014
Final Version Received: 17 JUN 2015
Published Online Date: 9 JUL 2015