Psychometrika (2026), **91**, 131–150
doi:10.1017/psy.2025.10078

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Constructive *Q*-Matrix Identifiability via Novel Tensor Unfolding

Yuqi Gu

Department of Statistics, Columbia University, USA
Email: yuqi.gu@columbia.edu

(Received 27 August 2025; revised 6 December 2025; accepted 8 December 2025)

Abstract

This work establishes a new identifiability theory for a cornerstone of various cognitive diagnostic models (CDMs) popular in psychometrics: the Q-matrix. The key idea is a novel tensor-unfolding proof strategy. Representing the joint distribution of J categorical responses as a J-way tensor, we strategically unfold the tensor into matrices in multiple ways and use their rank properties to identify the unknown Q-matrix. This approach departs fundamentally from all prior identifiability analyses in CDMs. Our proof is constructive, elucidating a population-level procedure to exactly recover the Q-matrix within a parameter space where each latent attribute is measured by at least two "pure" items that solely measure this attribute. The theory has several desirable features: it can constructively identify both the Q-matrix and the number of latent attributes; it applies to broad classes of linear and nonlinear CDMs with main or all saturated effects of attributes; and it accommodates polytomous responses, extending beyond classical binary response settings. The new identifiability result unifies and strengthens identifiability guarantees across diverse CDMs. It provides rigorous theoretical foundations and indicates a future pathway toward using tensor unfolding for practical Q-matrix estimation.

Keywords: Algebraic statistics; Cognitive diagnostic model; Constructive proof; Identifiability; Q-matrix; Tensor unfolding

1. Introduction

Cognitive Diagnostic Models (CDMs), or Diagnostic Classification Models (Rupp and Templin, 2008; von Davier and Lee, 2019), are popular discrete latent variable models widely used in educational and psychological measurement. Based on test respondents' multivariate categorical item responses, one can use a CDM to infer their fine-grained discrete latent attributes, such as cognitive abilities, personality traits, or mental disorders. One key structure in a CDM is the Q-matrix (Tatsuoka, 1983), which is a $J \times K$ binary matrix describing how the J items depend on the K latent attributes. Various CDMs have been proposed with different modeling assumptions. Popular examples include the conjunctive Deterministic Input Noisy Output "And" gate model (DINA; Junker and Sijtsma, 2001), the disjunctive Deterministic Input Noisy Output "Or" gate model (DINO; Templin and Henson, 2006), the main-effect diagnostic models including the linear logistic models (LLM; Maris, 1999), the reduced Reparameterized Unified Model (reduced-RUM; DiBello et al., 2012), and the additive CDM (ACDM; de la Torre, 2011), and all-saturated-effect diagnostic models including the general diagnostic models (GDM; von Davier, 2008), the log-linear CDM (LCDM; Henson et al., 2009), and the generalized DINA model (GDINA; de la Torre, 2011).

© The Author(s), 2026. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

The identifiability of CDMs is crucial to ensuring the validity of their statistical analysis. Especially in exploratory CDMs where the Q-matrix is unknown, uniquely identifying Q is a fundamental prerequisite for reliably estimating it using any method. In the past decade, the identifiability problems of CDMs and Q-matrix have attracted increasing interest. Regarding the *proof techniques*, the existing identifiability studies fall into two categories. The first category (Chen et al., 2020; Culpepper, 2019,2; Fang et al., 2019; Liu and Culpepper, 2024) leveraged Kruskal's Theorem on the uniqueness of three-way tensor factorizations (Kruskal, 1976,7) to establish identifiability. These approaches share a similar spirit to the general proof framework popularized by Allman et al. (2009). The second category (Chen et al., 2015; Gu and Xu, 2021; Liu et al., 2013; Xu and Shang, 2018) exploited the marginal moments of the binary responses and investigated the uniqueness of roots to the polynomial equations defined by these moments. This fine-grained proof technique can sometimes deliver sharper identifiability conditions compared to directly invoking Kruskal's Theorem.

In contrast to all of the above existing work, we adopt a fundamentally different proof strategy based on *tensor unfolding*. Unfolding or flattening a tensor (third-order or higher-order array) means reshaping it into a lower-order tensor or a matrix (Kolda and Bader, 2009). In a CDM, the population distribution of the J observed categorical variables can be represented by a Jth-order tensor, whose entries are elements in the joint probability mass function of the J-dimensional response vector. Our key proof idea is to strategically unfold this tensor to a matrix in various ways and use the rank properties of these resulting matrices to uniquely identify the Q-matrix. Our identifiability proof is a *constructive proof* for the first time in the literature: the proof itself is a population-level procedure and algorithm to reconstruct the Q-matrix from the population distribution in a restricted identifiable Q-matrix space. This identifiable space contains all Q-matrices that have at least two "pure" items solely measuring each latent attribute. Our proof's constructive nature departs from all existing identifiability analyses of the Q-matrix: previous works adopt *existence proofs*, by proving that if two Q-matrices lead to the same population distribution of observed responses, the two Q-matrices must be identical, without indicating how the Q-matrix can be reconstructed from the population distribution.

Our identifiability result enjoys three additional features. First, it can directly and constructively *identify the number of latent attributes* K together with the Q-matrix. Second, it is broadly applicable to various flexible CDMs, including all main-effect CDMs (such as reduced-RUM, LLM, and ACDM) and all-saturated-effect CDMs (such as GDM, LCDM, and GDINA). Third, it applies to CDMs with polytomous responses, in addition to traditional CDMs with binary responses. Recently, variants of CDMs with polytomous responses (Culpepper, 2019; Fang et al., 2019; Liu and Culpepper, 2024; Ma and de la Torre, 2016; Wayman et al., 2025) have attracted increasing interest due to their flexibility. Our theory offers a unified identifiability result for all these CDMs with general categorical responses. Compared to existing identifiability conditions for binary- or polytomous-response CDMs, our new condition is weaker because it only requires the Q-matrix to contain two identity matrices I_K without any additional assumption. Specifically, this implies that if there are exactly $J = 2K$ items and the Q-matrix contains just two I_K, then this Q is still identifiable; whereas to our best knowledge, such a Q-matrix does not satisfy previous identifiability conditions for general CDMs.

In the rest of this article, Section 2 introduces the unfoldings of the probability tensor in CDMs and gives a detailed illustrative example. Section 3 presents the new identifiability theory for both binary CDMs and polytomous CDMs. Section 4 concludes and discusses a future direction. The proofs of the theoretical results are included in the Appendix.

2. Unfolding Population Distribution Tensors in a CDM

2.1. Model Setup of CDMs with Binary Responses and Attributes

Consider a CDM where each subject is associated with J observed variables $\mathbf{R} = (R_1, \ldots, R_J)$ as responses to J test items and K latent attributes $\mathbf{A} = (A_1, \ldots, A_K)$. We start by considering the most commonly used CDMs with binary responses $\mathbf{R} \in \{0,1\}^J$ and binary attributes $\mathbf{A} \in \{0,1\}^K$, and later will generalize our result to polytomous responses and attributes. A key structure in a CDM is the Q-matrix introduced

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

in Tatsuoka (1983), which specifies the relationship between the observed responses and the latent attributes. The Q-matrix is a $J \times K$ matrix $\mathbf{Q} = (q_{jk})$ with binary entries, with rows indexed by the J items and columns by the K attributes. The entry $q_{jk} = 1$ or 0 indicates whether or not the jth item requires/measures the kth latent attribute. The row vectors of the Q-matrix are also called the $\mathbf{q}$-vectors and denoted by $\mathbf{q}_j$ for each $j \in [J]$. Here, for any positive integer M, we denote $[M] = \{1, 2, \ldots, M\}$.

To model the conditional distribution of the observed $\mathbf{R}$ given the latent $\mathbf{A}$, we use the general notation of the *item parameters*

$$\theta_{j,\boldsymbol{\alpha}} = \mathbb{P}(R_j = 1 \mid \mathbf{A} = \boldsymbol{\alpha}), \quad j = 1, \ldots, J; \quad \boldsymbol{\alpha} \in \{0,1\}^K,$$

and denote the collection of them by $\boldsymbol{\Theta}$. These item parameters are subject to certain equality constraints imposed by the Q-matrix in different ways under different models; see Examples 1 and 2 for concrete examples. For the latent attributes, we consider the most commonly adopted saturated attribute model with proportion parameters $\mathbf{p} = (p_{\boldsymbol{\alpha}}; \boldsymbol{\alpha} \in \{0,1\}^K)$, where

$$p_{\boldsymbol{\alpha}} = \mathbb{P}(\mathbf{A} = \boldsymbol{\alpha}), \quad \forall \boldsymbol{\alpha} \in \{0,1\}^K.$$

The proportion parameters $\mathbf{p}$ satisfy that $p_{\boldsymbol{\alpha}} \geq 0$ for all $\boldsymbol{\alpha}$ and $\sum_{\boldsymbol{\alpha} \in \{0,1\}^K} p_{\boldsymbol{\alpha}} = 1$. We use $\mathbf{e}_k$ to denote the K-dimensional canonical basis vector with an "1" in the kth entry and "0" in all other entries.

Under the above general setup, various CDMs were proposed with different diagnostic modeling assumptions and different link functions. *In this work, we focus on the flexible main-effect or all-saturated-effect CDMs*, which are also called the multi-parameter restricted latent class models in Gu and Xu (2021) and encompass a wide range of diagnostic models.

Remark 1. The Boolean-product-based DINA and DINO models (also called two-parameter restricted latent class models) exhibit very different algebraic structures from the multi-parameter CDMs. Such distinctive properties make the marginal-moment-based proof technique (Gu and Xu, 2021; Xu and Zhang, 2016) optimal for deriving the minimal identifiability conditions for the DINA and DINO models. Necessary and sufficient identifiability conditions for the Q-matrix in these models are already well understood in the literature (Gu, 2023; Gu and Xu, 2021). Therefore, we do not consider the DINA and DINO models in this work and instead focus on the general and flexible main-effect or all-saturated-effect CDMs.

In the following, we review concrete examples of multi-parameter CDMs in the following.

Example 1 (Main-Effect Cognitive Diagnosis Models). An important family of cognitive diagnosis models assumes that the $\theta_{j,\boldsymbol{\alpha}}$ depends on the main effects of those attributes required by item j, but not their interactions. This family includes the popular reduced Reparameterized Unified Model (rRUM; DiBello et al., 2012), Additive Cognitive Diagnosis Models (ACDM; de la Torre, 2011), the Linear Logistic Model (LLM; Maris, 1999), and the General Diagnostic Model (GDM; von Davier, 2008). We call them the Main-Effect Cognitive Diagnosis Models. In particular, under the rRUM,

$$\mathbb{P}^{\text{rRUM}}(R_j = 1 \mid \mathbf{A} = \boldsymbol{\alpha}, Q) = \theta_j^+ \prod_{k=1}^{K} r_{jk}^{q_{jk}(1-\alpha_k)},$$

where $\theta_j^+ = \mathbb{P}(R_j = 1 \mid A_k \geq q_{jk} \text{ for all } k = 1, \ldots, K)$ represents the positive response probability of a capable subject of item j (i.e., a subject mastering all required skills of item j), and $r_{jk} \in (0,1)$ is the parameter penalizing not mastering attribute k required by item j. Equivalently, the item parameter in the rRUM can be written as $\log \theta_{j,\boldsymbol{\alpha}}^{rRUM} = \beta_{j0} + \sum_{k=1}^{K} \beta_{jk} q_{jk} \alpha_k$. Similarly, the ACDM assumes the parameter $\theta_{j,\boldsymbol{\alpha}}$ can be written as a linear combination of the main effects of the required attributes:

$$\mathbb{P}^{\text{ACDM}}(R_j = 1 \mid \mathbf{A} = \boldsymbol{\alpha}, Q) = \beta_{j0} + \sum_{k=1}^{K} \beta_{jk} q_{jk} \alpha_k.$$

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

The Linear Logistic Model assumes a logistic link function with

$$\mathbb{P}^{\mathrm{LLM}}(R_j = 1 \mid \mathbf{A} = \boldsymbol{\alpha}, Q) = \sigma\Big(\beta_{j0} + \sum_{k=1}^{K} \beta_{jk} q_{jk} \alpha_k\Big),$$

where $\sigma(x) = 1/(1+e^{-x})$ is the logistic function.

Example 2 (All-saturated-effect Cognitive Diagnosis Models). Another popular type of cognitive diagnostic model assumes that the positive response probability depends on the main effects and all of the interaction effects of the required attributes of the item. We call these models all-saturated-effect cognitive diagnosis models. The GDINA model (de la Torre, 2011), the log-linear cognitive diagnosis models (LCDM; Henson et al., 2009), and the general diagnostic model (GDM; von Davier, 2008) are popular examples of all-saturated-effect models. The item parameters (conditional positive response probabilities) under these models are

$$\begin{aligned}\mathbb{P}^{\mathrm{All-Effect}}(R_j = 1 \mid \mathbf{A} = \boldsymbol{\alpha}) = f\Big(\beta_{j0} + \sum_{k=1}^{K} \beta_{jk} q_{jk} \alpha_k + \sum_{1 \le k_1 < k_2 \le K} \beta_{j,\{k_1,k_2\}} q_{j,k_1} q_{j,k_2} \alpha_{k_1} \alpha_{k_2} \\ + \cdots + \beta_{j,\{1,2,\cdots,K\}} \prod_{k=1}^{K} q_{jk} \prod_{k=1}^{K} \alpha_k\Big).\end{aligned} \tag{1}$$

When f is the identity link $f(x) = x$, the above gives the GDINA model, and when f is the inverse logit link $f(x) = \sigma(x) = 1/(1+e^{-x})$, the above gives the LCDM. von Davier (2014) showed that the GDINA and LCDM can be rewritten as GDMs with an extended skill space.

The local independence assumption is usually imposed in CDMs, meaning that the observed responses $R_1, \ldots, R_J$ are conditionally independent given the latent attributes $\mathbf{A}$. With this assumption, we can write the joint distribution of the observed $\mathbf{R}$ by marginalizing out the latent $\mathbf{A}$ in a CDM with binary responses and binary attributes:

$$\begin{aligned}\mathbb{P}(R_1 = r_1, \ldots, R_J = r_J \mid Q, \boldsymbol{\Theta}, \mathbf{p}) &= \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \mathbb{P}(\mathbf{A} = \boldsymbol{\alpha}) \prod_{j=1}^{J} \mathbb{P}(R_j = r_j \mid \mathbf{A} = \boldsymbol{\alpha}, Q) \\ &= \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} p_{\boldsymbol{\alpha}} \prod_{j=1}^{J} \theta_{j,\boldsymbol{\alpha}}^{r_j} (1 - \theta_{j,\boldsymbol{\alpha}})^{1-r_j},\end{aligned} \tag{2}$$

for any response pattern $\mathbf{r} = (r_1, \ldots, r_J) \in \{0,1\}^J$, where the Q-matrix-induced constraints on parameters $\theta_{j,\boldsymbol{\alpha}}$ are made implicit. The Q-matrix is said to be identifiable, if it can be uniquely identified from the population distribution of $\mathbf{R}$ in (2) up to the column permutation of Q (Liu et al., 2013).

2.2. Population Distribution Tensor under a CDM

Studying the identifiability in CDMs often requires one to take the tensor perspective. This subsection introduces the population tensor under a CDM and explains its importance in studying model identifiability.

We first introduce the notation of the *population distribution tensor* summarizing the population distribution of the observed response vector $\mathbf{R}$. The distribution of $\mathbf{R}$ in (2) can be characterized by a Jth-order tensor $\mathcal{T} = (t_{r_1, r_2, \ldots, r_J})$ of size $2 \times 2 \times \cdots \times 2$ with entries

$$t_{r_1, r_2, \ldots, r_J} = \mathbb{P}(R_1 = r_1, \ldots, R_J = r_J \mid Q, \boldsymbol{\Theta}, \mathbf{p}), \quad \forall r_1, r_2, \ldots, r_J \in \{0,1\} \tag{3}$$

denotes the marginal probability of the response pattern $\mathbf{r} = (r_1, r_2, \ldots, r_J)$ under the specified CDM. Tensor $\mathcal{T}$ is said to have J *modes* indexed by the items $j = 1, \ldots, J$. We call $\mathcal{T}$ the *population distribution tensor* of $\mathbf{R}$. For a vector $\mathbf{R} = (R_1, \ldots, R_M)$ and a subset $S \subseteq [M]$, denote $\mathbf{R}_S = (R_j; j \in S)$ as a $|S|$-dimensional subvector of $\mathbf{R}$ that collects all entries indexed by S.

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

When establishing the identifiability of the Q-matrix or continuous parameters in main-effect or all-saturated-effect CDMs, previous proof approaches based on Kruskal's Theorem often involve a step of rewriting this Jth-order tensor $\mathcal{T}$ into a three-way tensor (third-order tensor) by concatenating certain modes together (see, e.g. Culpepper, 2019; Fang et al., 2019). Even for studies that do not explicitly invoke Kruskal's Theorem (see, e.g. Gu and Xu, 2020; Xu, 2017; Xu and Shang, 2018), the identifiability proofs still implicitly partition the J binary responses into three subsets and investigate the polynomial moment equations arising from this partitioning. In this work, we take a fundamentally different approach but still investigates the population tensor $\mathcal{T}$. Our approach *avoids looking at three-way* tensor decompositions but examines *two-way tensor unfoldings* and hence delivers a strictly more relaxed identifiability condition for multiparameter CDMs than existing studies; the next subsection presents details of the tensor unfolding.

2.3. Tensor Unfoldings

This subsection introduces the key *tensor unfolding* idea behind our new identifiability proof. Unfolding or flattening a tensor (third-order or higher-order array) means reshaping it into a lower-order tensor or a matrix (Kolda and Bader, 2009). To prove the identifiability result for the Q-matrix, we show that it suffices to consider unfolding the Jth-order tensor $\mathcal{T}$ to various matrices. For this purpose, define the *row group* of indices as $S \subseteq [J]$ and the *column group* as $[J]\backslash S$. We denote the matrix resulting from the unfolded tensor as

$$\texttt{unfold}(\mathcal{T};\, S,\, [J]\backslash S) =: [\mathcal{T}]_{S,:}, \tag{4}$$

which is a $2^{|S|} \times 2^{J-|S|}$ matrix with rows indexed by the $2^{|S|}$ configurations of $\mathbf{R}_S \in \{0,1\}^{|S|}$ and columns by the $2^{J-|S|}$ configurations of $\mathbf{R}_{[J]\backslash S} \in \{0,1\}^{J-|S|}$. The entries in $[\mathcal{T}]_{S,:}$ are still those marginal response probabilities $t_{r_1,\ldots,r_J}$ in (3). A rather surprising fact that we will uncover shortly is that the ranks of such matrices (4) contain rich information about the unknown Q-matrix. For example, when S contains two items, the rank of $[\mathcal{T}]_{S,:}$ reflects whether these two items solely measure the same latent attribute; see Section 2.4 for details. So, we will strategically unfold the tensor $\mathcal{T}$ and use the rank properties of the resulting matrices as certificates to identify and recover the unknown Q-matrix.

We also introduce the notion of *marginal tensor* and *marginal unfolding*. Given a subset $S \subseteq [J]$, define the *marginal probability tensor* for $\mathbf{R}_S$ as a $|S|$-way tensor $\texttt{marginal}(\mathcal{T}, S)$ with entries specifying the joint PMF of the random vector $\mathbf{R}_S$. This marginal tensor can be obtained from the aforementioned full tensor $\mathcal{T}$ by appropriately summing up its entries: each new entry in $[\mathcal{T}]_S$ takes the form of $\sum_{j \notin S} \sum_{r_j=0,1} t_{r_1,\ldots,r_J}$. So, $\mathcal{T}$ and S uniquely defines the marginal probability tensor for $\mathbf{R}_S$ and we denote it by $\texttt{marginal}(\mathcal{T}, S)$. For two disjoint index sets $S_1, S_2 \subseteq [J]$ whose union does not equal $[J]$, we define a *marginal unfolding*

$$\texttt{unfold}(\texttt{marginal}(\mathcal{T}, S_1 \cup S_2),\, S_1,\, S_2) =: [\mathcal{T}]_{S_1,S_2},$$

which characterizes the joint probability table between random vectors $\mathbf{R}_{S_1}$ and $\mathbf{R}_{S_2}$. The entries in $[\mathcal{T}]_{S_1,S_2}$ are summations of $t_{r_1,\ldots,r_J}$ in the form of $\sum_{j \notin S_1 \cup S_2} \sum_{r_j=0,1} t_{r_1,\ldots,r_J}$, corresponding to the PMF of $\mathbf{R}_{S_1 \cup S_2}$ by marginalizing out other random variables $\mathbf{R}_{[J]\backslash(S_1 \cup S_2)}$. We introduce the marginal tensor $[\mathcal{T}]_{S_1,S_2}$ in order to focus on the items belonging to $S_1 \cup S_2$ and exclude the effects of latent attributes not manifested in $S_1 \cup S_2$.

We introduce the definition of a *conditional probability table* (CPT). Let $\mathbb{P}(R_j \mid \mathbf{A}_L)$ denote the CPT that specifies the conditional distribution of R_j given the latent attributes A_k for $k \in L$. In the general case where each observed R_j has C categories, the CPT $\mathbb{P}(R_j \mid \mathbf{A}_L)$ is a matrix with size $C \times 2^{|L|}$. Its $2^{|L|}$ columns are indexed by all possible configurations of the latent vector $\mathbf{A}_L$ ranging in $\{0,1\}^{|L|}$, and its C rows are indexed by the C categories of R_j in $\{0,1,\ldots,C-1\}$.

Finally, we introduce a useful notation of the *joint probability table* between two random vectors. Consider general categorical latent $A_1,\ldots,A_K \in \{0,1\}$. For two sets $S_1, S_2 \subseteq [K]$, we describe the joint

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

distribution of $\mathbf{A}_{S_1} = (A_k)_{k\in S_1}$ and $\mathbf{A}_{S_2} = (A_k)_{k\in S_2}$ using

$$\mathbb{P}(\mathbf{A}_{S_1}, \mathbf{A}_{S_2}) : \text{ a } 2^{|S_1|} \times 2^{|S_2|} \text{ matrix.}$$

The rows of $\mathbb{P}(\mathbf{A}_{S_1}, \mathbf{A}_{S_2})$ are indexed by the configurations of $\mathbf{A}_{S_1} \in \{0,1\}^{|S_1|}$, and columns by those of $\mathbf{A}_{S_2} \in \{0,1\}^{|S_2|}$. Importantly, we emphasize that S_1 and S_2 *do not need to be disjoint and can overlap*. At a high level, allowing overlapping sets will enable us to study the rank properties of unfoldings potentially involving multi-attribute items. When $S_1 \cap S_2 \neq \varnothing$, $\mathbb{P}(\mathbf{A}_{S_1}, \mathbf{A}_{S_2})$ has some zero entries corresponding to impossible configurations of $\mathbf{A}_{S_1 \cup S_2}$. An extreme example is if $S_1 = S_2$, then $\mathbb{P}(\mathbf{A}_{S_1}, \mathbf{A}_{S_1})$ is a $2^{|S_1|} \times 2^{|S_1|}$ diagonal matrix, because $\mathbb{P}(\mathbf{A}_{S_1} = \mathbf{a}, \mathbf{A}_{S_1} = \mathbf{b}) = 0$ for any $\mathbf{a} \neq \mathbf{b}$. In this case, the diagonal entries of $\mathbb{P}(\mathbf{A}_{S_1}, \mathbf{A}_{S_1})$ are given by the PMF of $\mathbf{A}_{S_1}$. Another example is when $S_1 \subseteq S_2$, then the matrix $\mathbb{P}(\mathbf{A}_{S_1}, \mathbf{A}_{S_2})$ has orthogonal row vectors because $\mathbb{P}(\mathbf{A}_{S_1} = \mathbf{a}, \mathbf{A}_{S_2} = \mathbf{b}) \neq 0$ only if $\mathbf{a}$ is a subvector of $\mathbf{b}$ indexed by integers in S_1. This general matrix notation of $\mathbb{P}(\mathbf{A}_{S_1}, \mathbf{A}_{S_2})$ for potentially overlapping sets S_1 and S_2 turns out to be very useful to facilitate our identifiability proofs.

2.4. A Toy Example Illustrating the Tensor Unfolding Insight

The next toy example illustrates the tensor unfolding idea and reveals useful insights on how to utilize unfoldings to identify and recover the Q-matrix.

Consider $J = 5$, $K = 2$, and the following Q-matrix

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

If considering binary responses $\mathbf{R} \in \{0,1\}^5$, then the population distribution tensor $\mathcal{T}$ has size $2 \times 2 \times 2 \times 2 \times 2$. Following the definition of CDMs in Section 2.1, we can write

$$\begin{aligned}
&\mathbb{P}(R_1 = r_1, R_2 = r_2, R_3 = r_3, R_4 = r_4, R_5 = r_5) \\
&= \sum_{(a_1,a_2)\in\{0,1\}^2} \mathbb{P}(\mathbf{A} = (a_1,a_2)) \prod_{j=1}^{5} \mathbb{P}(R_j = r_j \mid \mathbf{A} = (a_1,a_2)) \\
&= \sum_{(a_1,a_2)\in\{0,1\}^2} \mathbb{P}(\mathbf{A} = (a_1,a_2)) \mathbb{P}(\mathbf{R}_{1,2} = \mathbf{r}_{1,2} \mid A_1 = a_1) \mathbb{P}(\mathbf{R}_{3,4,5} = \mathbf{r}_{3,4,5} \mid A_1 = a_1, A_2 = a_2) \\
&= \sum_{(a_1,a_2)\in\{0,1\}^2} \sum_{a_3\in\{0,1\}} \mathbb{P}(\mathbf{A} = (a_1,a_2), A_1 = a_3) \mathbb{P}(\mathbf{R}_{1,2} = \mathbf{r}_{1,2} \mid A_1 = a_3) \\
&\quad \times \mathbb{P}(\mathbf{R}_{3,4,5} = \mathbf{r}_{3,4,5} \mid A_1 = a_1, A_2 = a_2),
\end{aligned} \tag{5}$$

where we have purposefully introduced the term $\mathbb{P}(\mathbf{A} = (a_1,a_2), A_1 = a_3)$ to facilitate the following matrix factorization perspective for the unfolded tensor. To be more specific, the last equality in the above expression holds because although a_3 can freely range in $\{0,1\}$, the probability $\mathbb{P}(\mathbf{A} = (a_1,a_2), A_1 = a_3)$ will be zero for any $a_3 \neq a_1$.

First, we unfold the tensor $\mathcal{T}$ so that the row group contains item indices $\{1,2\}$, and the column group contains item indices $\{3,4,5\}$. Let $t_{ab,cde} = \mathbb{P}(R_1 = a, R_2 = b, R_3 = c, R_4 = d, R_5 = e)$ denote entries from the joint distribution of the five observed variables, where $a,b,c,d,e \in \{0,1\}$. With the row group $S = \{1,2\}$ and column group $[J]\backslash S = \{3,4,5\}$, the unfolded tensor is a $2^2 \times 2^3$ joint probability table of

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

$\mathbf{R}_{1,2} = (R_1, R_2)$ and $\mathbf{R}_{3,4,5} = (R_3, R_4, R_5)$:

$$\begin{aligned}
&\texttt{unfold}(\mathcal{T};\, S,\, [J]\backslash S) = [\mathcal{T}]_{\{1,2\},\,:} \qquad (6)\\
&= \begin{pmatrix}
t_{00,000} & t_{00,001} & t_{00,010} & t_{00,011} & t_{00,100} & t_{00,101} & t_{00,110} & t_{00,111}\\
t_{01,000} & t_{01,001} & t_{01,010} & t_{01,011} & t_{01,100} & t_{01,101} & t_{01,110} & t_{01,111}\\
t_{10,000} & t_{10,001} & t_{10,010} & t_{10,011} & t_{10,100} & t_{10,101} & t_{10,110} & t_{10,111}\\
t_{11,000} & t_{11,001} & t_{11,010} & t_{11,011} & t_{11,100} & t_{11,101} & t_{11,110} & t_{11,111}
\end{pmatrix}\\
&= \underbrace{\mathbb{P}(\mathbf{R}_{1,2} \mid A_1)}_{2^2\times 2} \cdot \underbrace{\mathbb{P}(A_1, \mathbf{A}_{1,2})}_{2\times 2^2} \cdot \underbrace{\mathbb{P}(\mathbf{R}_{3,4,5} \mid \mathbf{A}_{1,2})^{\top}}_{2^2\times 2^3}, \quad \text{(following from the decomposition (2))}\\
&\Longrightarrow \operatorname{rank}([\mathcal{T}]_{\{1,2\},\,:}) \le 2.
\end{aligned}$$

Specifically, $\mathbb{P}(\mathbf{R}_{1,2} \mid A_1)$ is a $2^2 \times 2$ matrix that collects the following conditional probabilities:

$$\mathbb{P}(\mathbf{R}_{1,2} \mid A_1) = \begin{pmatrix}
\mathbb{P}(\mathbf{R}_{1,2} = (0,0) \mid A_1 = 0) & \mathbb{P}(\mathbf{R}_{1,2} = (0,0) \mid A_1 = 1)\\
\mathbb{P}(\mathbf{R}_{1,2} = (0,1) \mid A_1 = 0) & \mathbb{P}(\mathbf{R}_{1,2} = (0,1) \mid A_1 = 1)\\
\mathbb{P}(\mathbf{R}_{1,2} = (1,0) \mid A_1 = 0) & \mathbb{P}(\mathbf{R}_{1,2} = (1,0) \mid A_1 = 1)\\
\mathbb{P}(\mathbf{R}_{1,2} = (1,1) \mid A_1 = 0) & \mathbb{P}(\mathbf{R}_{1,2} = (1,1) \mid A_1 = 1)
\end{pmatrix}_{4\times 2},$$

and the $2^3 \times 2^2$ matrix $\mathbb{P}(\mathbf{R}_{3,4,5} \mid \mathbf{A}_{1,2})$ is defined similarly. In summary, the above derivation shows that we can factorize $\texttt{unfold}(\mathcal{T};\, \{1,2\},\, \{3,4,5\})$ into the product of three matrices, which have size 4×2, 2×4, and 4×8, respectively. Therefore, the rank of the 4×8 matrix $\texttt{unfold}(\mathcal{T};\, \{1,2\},\, \{3,4,5\})$ is at most 2. Similarly, $\operatorname{rank}(\texttt{unfold}(\mathcal{T};\, \{3,4\},\, \{1,2,5\})) \le 2$ also holds due to symmetry because items 3 and 4 both solely measure the second attribute.

In contrast, if two items j_1, j_2 do not solely measure the same latent attribute, the unfolding $[\mathcal{T}]_{\{j_1,j_2\},:}$ will show distinct rank behavior. Specifically, if we unfold the tensor $\mathcal{T}$ with the row group being $\{1,3\}$ and the column group being $\{2,4,5\}$, then as will be shown rigorously in the proof of the main theorem, the unfolding $\texttt{unfold}(\mathcal{T};\, \{1,3\},\, \{2,4,5\})$ has rank strictly larger than 2; more generally, $\operatorname{rank}(\texttt{unfold}(\{j_1,j_2\},\, [J]\backslash\{j_1,j_2\})) > 2$ holds as long as items j_1 and j_2 are not solely measuring the same attribute. In other words, whether the rank of $\texttt{unfold}(\{j_1,j_2\},\, [J]\backslash\{j_1,j_2\})$ exceeds two is a perfect certificate that reveals whether j_1 and j_2 are measuring the same latent attribute. By exhaustively examining the rank of the unfoldings with the row group consisting of every pair of items, we can exactly tell which items are solely measuring the same attribute as well as the number of latent attributes: items 1,2 solely measure one latent attribute, items 3,4 solely measure the other latent attribute, and there are two latent attributes. Note that if there is an additional item 6 that solely measures the first latent attribute, then $\texttt{unfold}(\{1,2\},\, [J]\backslash\{1,2\})$, $\texttt{unfold}(\{1,6\},\, [J]\backslash\{1,6\})$, and $\texttt{unfold}(\{2,6\},\, [J]\backslash\{2,6\})$ all have ranks at most two, so all three items 1, 2, and 6 will be correctly assigned as solely measuring the same latent attribute.

Since item 5 with $\mathbf{q}_5 = (1,1)$ measures both latent attributes, identifying such $\mathbf{q}$-vectors calls for a more challenging analysis than the pairwise unfolding illustrated in the previous paragraph. But surprisingly and fortunately, we rigorously prove (in Theorems 1 and 2 in Section 3) that examining strategically constructed unfolded matrices' ranks can still identify the $\mathbf{q}$-vectors that vary arbitrarily in $\{0,1\}^K$. The proof for these multi-attribute items is much more technical than the single-attribute items described above, so we omit the details here and refer interested readers to the proof of the theorems.

In summary, the toy example above illustrates that the ranks of unfolded tensors can help reveal the underlying Q-matrix. Our theorem applies to general Q-matrix dimensions and structures far beyond this toy example, as will be laid out in the next section.

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

3. Identifiability Theory

3.1. CDMs with Binary Responses

We next rigorously formalize the mild technical assumptions we impose for identifiability. If an item j solely measures attribute k (that is, $\mathbf{q}_j = \mathbf{e}_k$ where $\mathbf{e}_k$ is the kth canonical basis vector), then we call it a *single-attribute item*.

Definition 1 (Non-degenerate Single-attribute Items: Binary-CDMs). A single-attribute item j measuring attribute k is non-degenerate if the CPT $\mathbb{P}(R_j \mid A_k)$ has full rank 2.

Requiring an item to satisfy Definition 1 is a mild assumption. Mathematically, if item j solely requires attribute k, then $\mathbb{P}(R_j \mid \mathbf{A}) \equiv \mathbb{P}(R_j \mid A_k)$ and the 2×2 conditional probability table is:

$$\begin{pmatrix} \mathbb{P}(R_j = 0 \mid A_k = 0) & \mathbb{P}(R_j = 0 \mid A_k = 1) \\ \mathbb{P}(R_j = 1 \mid A_k = 0) & \mathbb{P}(R_j = 1 \mid A_k = 1) \end{pmatrix}.$$

In this case, requiring $\mathbb{P}(R_j \mid A_k)$ to have full rank 2 as specified in Definition 1 is equivalent to requiring R_j not to be independent of A_k. We next present the first main theorem.

Theorem 1. *Consider any CDM in Examples 1 and 2 with binary responses. Consider the following identifiable parameter space for the Q-matrix:*

$$\mathcal{Q} = \{Q \in \{0,1\}^{J \times K} : PQ = (I_K, I_K, Q^{*\top})^\top \text{ for some } J \times J \text{ permutation matrix } P\}, \tag{7}$$

where I_K is the $K \times K$ identity matrix and Q^* is an arbitrary binary matrix or an empty matrix (meaning $J = 2K$). Assume all single-attribute items are non-degenerate and $p_{\boldsymbol{\alpha}} > 0$ for all $\boldsymbol{\alpha} \in \{0,1\}^K$. If the true and alternative Q-matrices $Q_{true}, Q_{alternative} \in \mathcal{Q}$ lead to the same marginal distribution of the observed response vector $\mathbf{R}$, then Q_{true} is identifiable and $Q_{alternative}$ must be identical to Q_{true} up to a column permutation. Moreover, Q_{true} can be constructively identified from the population distribution tensor $\mathcal{T}$ via strategically unfolding it and examining the ranks.

The identifiable Q-matrix space in (7) includes all $J \times K$ binary matrices that contain at least two identity submatrices after some column and row permutation. Specifically, we prove Theorem 1 by establishing the following two key technical propositions, which explicitly lay out the tensor unfolding procedures along with the corresponding rank certificates to identify and recover the Q-matrix.

Proposition 1. *Under the condition in Theorem 1, for arbitrary $j_1 \neq j_2 \in [J]$, it holds that*

$$\operatorname{rank}(\underbrace{[\mathcal{T}]_{\{j_1,j_2\},:}}_{2^2 \times 2^{J-2}}) \leq 2 \quad \text{if and only if} \quad \mathbf{q}_{j_1} = \mathbf{q}_{j_2} = \mathbf{e}_k \text{ for some } k \in [K].$$

Proposition 1 formalizes the intuition explained in the toy example in Section 2.4, and precisely characterizes how to identify the $\mathbf{q}$-vectors for all single-attribute items in the Q-matrix. In words, Proposition 1 states that the rank of $[\mathcal{T}]_{\{j_1,j_2\},:}$ reflects whether the two items j_1 and j_2 solely measure the same latent attribute. Interestingly, Proposition 1 also implies we can constructively identify K, the number of latent attributes: K is equal to the maximum number of mutually disjoint item sets $S_1, S_2, \ldots \subseteq [J]$ such that the rank of the unfoldings $[\mathcal{T}]_{\{j_1,j_2\},:}$ is at most two when considering all $j_1, j_2 \in S_k$ for each k.

Remark 2. In principle, the tensor unfolding techniques can also be applied to the DINA and DINO models, but the Boolean product structure in these two models (i.e., the conjunctive assumption in DINA and the disjunctive assumption in DINO) makes the rank properties of the unfolded tensor different from those under the main-effect and all-saturated-effect CDMs. More specifically, take Proposition 1 for example, which states that for arbitrary $j_1 \neq j_2$ it holds that $\operatorname{rank}([\mathcal{T}]_{\{j_1,j_2\},:}) \leq 2$ if and only if $\mathbf{q}_{j_1} = \mathbf{q}_{j_2} = \mathbf{e}_k$ for some $k \in [K]$. Actually, under the DINA and DINO models, the "if" part in the above statement still holds (that is, the single-attribute items still lead to low-rank unfoldings), but the "only if" part no longer holds because of the Boolean-product structure (that is, the low-rank

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

unfoldings are not necessarily caused by those single-attribute items). Therefore, under the DINA or DINO model, the rank of $[\mathcal{T}]_{\{j_1,j_2\},:}$ being at most 2 is not a valid "certificate" for concluding that items j_1 and j_2 both only measure the same latent attribute. Based on the above reason, we focus on the non-DINA/DINO CDMs, for which the tensor unfolding technique can weaken the existing identifiability conditions of the Q-matrix and provide new theoretical understanding.

Having identified K and all single-attribute items, the next proposition characterizes a more nuanced tensor unfolding strategy along with their rank certificates to identify the challenging multi-attribute items. These items can have arbitrary corresponding $\mathbf{q}$-vectors ranging in $\{0,1\}^K \backslash \{\mathbf{e}_1, \ldots, \mathbf{e}_K\}$.

Proposition 2. *Under the condition in Theorem 1, assume without loss of generality that the first* $2K$ *rows of the Q-matrix are identified to be* $(I_K; I_K)^\top$. *The following holds for any attribute* $k \in [K]$ *and any item* j *that requires at least two attributes (i.e., with* $\sum_{k=1}^K q_{jk} \geq 2$*):*

$$\operatorname{rank}(\underbrace{[\mathcal{T}]_{([K]\backslash\{k\})\cup\{j\},\ \{K+1,\ldots,2K\}}}_{2^K\times 2^K}) > 2^{K-1} \quad \textit{if and only if} \quad q_{jk} = 1.$$

The proof of Proposition 2 is challenging. The difficulties include: first, to come up with an appropriate row group S_1 and a corresponding column group S_2 of variables from $R_1, \ldots, R_J$ such that the marginal unfolding $[\mathcal{T}]_{S_1,S_2}$ contains as much as possible useful information about the multi-attribute $\mathbf{q}$-vectors; and second, to carefully investigate the rank of these unfoldings to establish the certificate for recovering those $\mathbf{q}$-vectors.

The identifiability condition in Theorem 1 is weaker than existing strict identifiability conditions on the Q-matrix for main-effect or all-saturated-effect, binary-response or polytomous response CDMs (e.g., Culpepper, 2019; Fang et al., 2019; Xu and Shang, 2018). Specifically, the condition in Theorem 1 is satisfied if, after some row permutation, Q takes the form of $Q = (I_K, I_K)^\top$ with exactly $J = 2K$ items; that is, when each latent attribute is exactly measured by two pure items, and there are no additional items in the test. To our best knowledge, such a Q-matrix does not satisfy the identifiability condition in the existing studies for general CDMs. The high-level intuition behind why our condition relaxes existing conditions in the literature is that, existing studies either explicitly or implicitly leverage the *three-way* decompositions of the population distribution tensor to establish identifiability. In contrast, we work with the rank of *matrices* to establish identifiability despite starting with tensors, and these matrices essentially describe the joint distributions of *two* random vectors instead of *three* random vectors. This intuitively explains why our approach delivers a weaker identifiability condition that only requires each latent attribute to be measured twice in the Q-matrix.

It is worth noting that Culpepper (2023) relaxed the condition that Q needs to contain identity submatrices by introducing and studying the notion of "dyads", which consist of pairs of items. Specifically, the identifiability theorem in Culpepper (2023) applies even if no single-attribute items exist by allowing *saturated dyads*. This significantly broadens the applicability of the identifiability theory to practical cognitive diagnostic assessment settings. However, that work still leverages Kruskal's Theorem on the uniqueness of three-way tensor decompositions to establish identifiability with the dyad structure and requires $J > 2K$. So, the theorem in Culpepper (2023) does not apply to the case where the Q-matrix contains exactly $2K$ rows; moreover, the proof of that theorem is still an existence proof, instead of a constructive proof like our tensor unfolding approach.

3.2. CDMs with Polytomous Responses

In this section, we consider CDMs with polytomous item responses. Suppose item $R_j \in \{0,1,\ldots,C_j-1\}$ has C_j categories for each $j \in \{1,\ldots,J\}$, where $C_j \geq 2$ is a general integer. Here, we allow different items to have different numbers of response categories (i.e., C_j across different j can be different), similar to the setting considered in Liu and Culpepper (2024) and Wayman et al. (2025). We still use the same notation for the proportion parameters $\mathbf{p} = (p_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\{0,1\}^K}$ to describe the joint distribution of the binary latent attributes. To describe the conditional distribution of the responses $\mathbf{R}$ given the latent attributes

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

$\mathbf{A}$, we introduce item parameters $\mathbf{\Lambda} = (\lambda_{j,c,\boldsymbol{\alpha}})$:

$$\lambda_{j,c,\boldsymbol{\alpha}} = \mathbb{P}(R_j = r_j \mid \mathbf{A} = \boldsymbol{\alpha}), \quad j \in [J], \ r_j \in \{0,1,\ldots,C_j-1\}, \ \boldsymbol{\alpha} \in \{0,1\}^K.$$

The Q-matrix induces constraints on the $\mathbf{\Lambda}$ parameters similar to the binary CDM case; here, $\lambda_{j,c,\boldsymbol{\alpha}}$ depends only on those attributes that are measured by the jth item according to the Q-matrix. Mathematically, this means

$$\mathbb{P}(R_j = r_j \mid \mathbf{A} = \boldsymbol{\alpha}) \equiv \mathbb{P}(R_j = r_j \mid \{A_k = \alpha_k\}_{q_{jk}=1}), \quad \forall j \in [J], \ \boldsymbol{\alpha} \in \{0,1\}^K.$$

The above general assumption summarizes the key property shared by many polytomous-response CDMs. Our considered setting hence covers existing models for polytomous responses, such as Chen and de la Torre (2018), Fang et al. (2019), Culpepper (2019), and Liu and Culpepper (2024).

The population distribution of the observed response vector $\mathbf{R}$ can be written as

$$\begin{aligned}\mathbb{P}(R_1 = r_1,\ldots,R_J = r_J \mid Q,\mathbf{\Lambda},\mathbf{p}) &= \sum_{\boldsymbol{\alpha}\in\{0,1\}^K} \mathbb{P}(\mathbf{A} = \boldsymbol{\alpha}) \prod_{j=1}^{J} \mathbb{P}(R_j = r_j \mid \mathbf{A} = \boldsymbol{\alpha}) \\ &= \sum_{\boldsymbol{\alpha}\in\{0,1\}^K} p_{\boldsymbol{\alpha}} \prod_{j=1}^{J} \prod_{c=0}^{C_j-1} \lambda_{j,c,\boldsymbol{\alpha}}^{\mathbb{1}(r_j=c)},\end{aligned}$$

for any multivariate polytomous response pattern $\mathbf{r} = (r_1,\ldots,r_J)$ with $r_j \in \{0,\ldots,C-1\}$ for each j. The J-way population distribution tensor $\mathcal{T} = (t_{r_1,r_2,\ldots,r_J})$ characterizing the distribution of $\mathbf{R}$ has size $C_1 \times C_2 \times \cdots \times C_J$, with entries

$$t_{r_1,r_2,\ldots,r_J} = \mathbb{P}(R_1 = r_1,\ldots,R_J = r_J \mid Q,\mathbf{\Lambda},\mathbf{p}), \quad \forall r_j \in \{0,1,\ldots,C_j-1\}, \ j = 1,\ldots,J.$$

We will unfold this tensor similarly to the binary CDM case. Consider the *row group* of indices as $S \subseteq [J]$ and the *column group* as $[J]\backslash S$, then the matrix resulting from the unfolded tensor is

$$[\mathcal{T}]_{S,:} = \texttt{unfold}(\mathcal{T};\, S,\, [J]\backslash S) : \text{ size } \prod_{j\in S} C_j \times \prod_{j\in[J]\backslash S} C_j,$$

The notations of marginal tensors and marginal unfoldings are defined similarly.

To formalize the requirements on the polytomous CDMs, we introduce the following two definitions.

Definition 2 (Non-degenerate Single-attribute Items in Polytomous-CDMs). A single-attribute item j solely measuring attribute k is said to be non-degenerate if the following conditional probability table, denoted by $\mathbb{P}(R_j \mid A_k)$, has full column rank 2:

$$\begin{pmatrix} \mathbb{P}(R_j = 0 \mid A_k = 0) & \mathbb{P}(R_j = 0 \mid A_k = 1) \\ \vdots & \vdots \\ \mathbb{P}(R_j = C_j - 1 \mid A_k = 0) & \mathbb{P}(R_j = C_j - 1 \mid A_k = 1) \end{pmatrix}$$

Definition 2 is a direct extension of the earlier Definition 1 to the polytomous CDMs and covers that as a special case.

As mentioned earlier in Section 2.1, we do not consider the Boolean-product-based CDMs (e.g., DINA or DINO models) in this work. We next formalize the non-DINA/DINO assumption in the polytomous CDM.

Definition 3 (Non-DINA/DINO Assumption). For any response category $c \in \{0,1,\ldots,C_j-1\}$ and any latent attributes configuration $\boldsymbol{\alpha} \in \{0,1\}^K$, the conditional probability $\mathbb{P}(R_j = c \mid \mathbf{A} = \boldsymbol{\alpha})$ does not conform to the Boolean-product-based DINA or DINO model, but rather conforms to any linear or nonlinear main-effect or all-saturated-effect-based CDM.

We have the following main theorem for polytomous CDMs with essentially the same identifiability condition as the binary CDMs, again proved via tensor unfoldings.

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

Theorem 2 (Polytomous-response CDMs). *Consider a polytomous-response CDM under Definitions 2 and 3. Still consider the restricted identifiable Q-matrix space $\mathcal{Q}$ defined in (7). Assume all single-attribute items are non-degenerate and $p_{\boldsymbol{\alpha}} > 0$ for all $\boldsymbol{\alpha} \in \{0,1\}^K$. If the true and alternative Q-matrices $Q, \tilde{Q} \in \mathcal{Q}$ lead to the same distribution of the observed polytomous response vector $\mathbf{R} \in \times_{j=1}^{J}\{0,\ldots,C_j-1\}$, then the Q-matrix is identifiable and can be constructively identified from the population distribution tensor $\mathcal{T}$ via unfoldings.*

Theorem 2 is proved by establishing the following proposition on the ranks of the unfolded tensors in polytomous CDMs.

Proposition 3. *Under the condition in Theorem 1, the following rank statements hold.*

(a) *For arbitrary $j_1 \neq j_2 \in [J]$, it holds that*

$$\operatorname{rank}\Big(\underbrace{[\mathcal{T}]_{\{j_1,j_2\},:}}_{C_{j_1}C_{j_2}\times\prod_{j\neq j_1,j_2}C_j}\Big) \leq 2 \quad \textit{if and only if} \quad \mathbf{q}_{j_1} = \mathbf{q}_{j_2} = \mathbf{e}_k \textit{ for some } k \in [K].$$

(b) *Assume without loss of generality that the first $2K$ rows of the Q-matrix are identified to be $(I_K; I_K)^\top$. The following holds for any attribute $k \in [K]$ and any item $j \in \{2K+1,\ldots,J\}$ that requires at least two attributes (i.e., with $\sum_{k=1}^{K} q_{jk} \geq 2$):*

$$\operatorname{rank}\big([\mathcal{T}]_{([K]\setminus\{k\})\cup\{j\},\ \{K+1,\ldots,2K\}}\big) > 2^{K-1} \quad \textit{if and only if} \quad q_{jk} = 1.$$

The above proposition generalizes Proposition 1 from the binary setting to the general polytomous setting. The two scenarios (a) and (b) in Proposition 3 give explicit and precise characterizations of the $\mathbf{q}$-vectors of single-attribute items and multiple-attribute items, respectively. These conditions are simple rank constraints of the matrices resulting from unfolding the $C_1 \times C_2 \times \cdots \times C_J$ tensor $\mathcal{T}$, and they serve as certificates to reveal the unknown $\mathbf{q}$-vectors for all J items.

It is worth mentioning that prior works have studied *generic identifiability* in CDMs, which is a slightly weaker notion than strict identifiability studied in this work. Generic identifiability means the parameters are almost everywhere identifiable in the parameter space, possibly excluding a Lebesgue mesasure zero set where identifiability breaks down. For main-effect or all-saturated-effect CDMs, generic identifiablity has been established under weaker conditions on the Q-matrix (e.g., Chen et al., 2020; Gu and Xu, 2020), without requiring Q to contain two identity submatrices I_K. We remark that it is not technically straightforward to establish generic identifiability using the tensor unfolding technique. To clarify, most previous studies that established generic identifiability for CDMs mainly utilized Kruskal's Theorem by adapting the technique popularized by Allman et al. (2009). Roughly speaking, in those works, generic identifiability is often proved by finding one possible configuration of the parameters and showing that identifiability holds under this configuration using Kruskal's Theorem; then, generic identifiability follows by concluding that the identifiable parameter configuration is "generic" in the parameter space, so that non-identifiability occurs only in a negligible subset of the parameter space. Therefore, generic identifiability is inherently tied to the Kruskal Theorem-based *existence proofs*, while not entirely aligned with the tensor-unfolding-based *constructive proofs*. Relaxing the condition of "Q contains at least two I_K matrices" would distort the rank properties of the unfolded tensor, which are important certificates to recover the Q-matrix in our proof.

4. Discussion

We have established a new identifiability theory of the Q-matrix for various main-effect and all-saturated-effect, binary and polytomous-response CDMs. Our proof exploits a novel tensor unfolding technique and is the first constructive identifiability proof in the literature.

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

Previous studies such as Chen et al. (2020) and Liu and Culpepper (2024) leveraged Kruskal's Theorem to study the "Δ-matrix" in diagnostic models, which has size $J \times 2^K$, with columns corresponding to both main-effects and all possible higher-order interaction effects of latent attributes. Compared to the Q-matrix, the Δ-matrix provides a more fine-grained summary of the measurement model structure and distinguishes the main-effect models (such as the ACDM and reduced-RUM) from the all-saturated-effect models (such as the GDINA and LCDM). On the other hand, however, the Q-matrix can be viewed as a more general and high-level summary of the *statistical dependence* relationship between the observed item responses and the latent attributes. In the language of probabilistic graphical models, the Q-matrix encodes the bipartite graph structure between the observed variables and the latent ones, where $q_{jk} = 1$ or 0 indicates whether or not the jth observed response *directly depends on* (has a directed arrow from) the kth latent attribute. One interesting property of our tensor-unfolding-based identifiability result is that the proof is *agnostic* to the measurement model structure, regardless of whether the model is main-effect or all-saturated-effect, and always recovers the fundamental dependence relationship between the observed and latent variables encoded by the Q-matrix. More specifically, the rank properties of the unfolded tensor reflect whether each observed variable depends on each latent variable, rather than the detailed form of this dependence (i.e., being main-effect only or involving interaction effects). With that said, we believe that it may be possible to adapt the tensor unfolding technique to study the identifiability of the Δ-matrix in the future.

To facilitate understanding the constructive proof, we next provide a summary of how we leverage tensor unfolding to uniquely identify and recover the Q-matrix entries based on the probabilities of the response patterns. This proof outline follows from Propositions 1 and 2 for binary-response CDMs, and Proposition 3 for polytomous-response CDMs. First, we identify all single-attribute items by exhaustively considering all pairs of items j_1, j_2 ranging in the item set $\{1, \ldots, J\}$ and examining the rank of the unfolding $[\mathcal{T}]_{\{j_1, j_2\},:}$. If the rank is smaller than or equal to 2, then items j_1, j_2 must be measuring the same latent attribute. After this first step, all $\mathbf{q}$-vectors corresponding to single-attribute items have been recovered, giving K item sets $S_1, \ldots, S_K \subseteq \{1, \ldots, J\}$, where each set S_k contains items that solely measure latent attribute k. Our identifiability condition ensures that each S_k contains at least two items, and we denote these two item indices by j_{k1} and j_{k2}. Second, for any other item j that does not belong to $\cup_{k=1}^{K} S_k$ and for any $k \in [K]$, we recover the entry q_{jk} by considering the unfolding $[\mathcal{T}]_{(\{j_{11}, \ldots, j_{K1}\} \setminus \{j_{k1}\}) \cup \{j\},\ \{j_{12}, \ldots, j_{K2}\}}$. If the rank of this unfolding is larger than 2^{K-1}, then we conclude $q_{jk} = 1$; otherwise, conclude $q_{jk} = 0$. After this second step, all $\mathbf{q}$-vectors corresponding to multi-attribute items have also been recovered. This completes the constructive proof strategy for identifying and recovering the whole Q-matrix.

Our proof indicates an interesting future direction toward practically estimating the Q-matrix via tensor unfolding. To clarify, the identifiability proof in this paper operates on the *population distribution tensor* summarizing the population distribution of $\mathbf{R}$; this identifiability notion is the same as the classical notion of population identifiability in all existing studies of CDMs and the Q-matrix (e.g., Chen et al., 2020,1; Culpepper, 2019; Gu and Xu, 2021; Liu et al., 2013; Xu and Shang, 2018). In practice, the available data are a finite sample, and one needs to use the samples to estimate the Q-matrix. Previous identifiability results are inherently disconnected from the estimation methods due to their existence proof nature: after proposing the identifiability conditions for CDMs, existing studies often proceed to estimate the Q-matrix and parameters using likelihood-based or Bayesian procedures. More specifically, frequentist methods often impose a sparsity-inducing penalty on continuous parameters and let the estimated sparsity pattern inform the Q-matrix structure (Chen et al., 2015; Xu and Shang, 2018), while Bayesian methods incorporate carefully designed MCMC sampling steps to sample only from the identifiable space of Q-matrices (Chen et al., 2020,1; Culpepper, 2019). Although identifiability conditions serve as a guideline for navigating Q-matrix estimation in these studies, the identifiability theory itself does not directly imply any specific estimation procedure.

In contrast, our constructive proof implies the possibility of developing a practical algorithm that takes the *empirical distribution tensor* $\widehat{\mathcal{T}}$ (constructed from sample proportions of observed response

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

patterns) as input and leverages and generalizes Proposition 3 to directly estimate the Q-matrix. Specifically, $\widehat{\mathcal{T}}$ has the same size as its population counterpart $\mathcal{T}$, but its entries are empirical proportions of each response pattern in the sample, instead of the population probabilities of the response patterns $\mathbb{P}(R_1 = r_1, \ldots, R_J = r_J)$. To this end, important modifications must be made to replace the rank constraints of the unfolded population tensor with surrogate singular value conditions on the unfolded empirical tensor. Developing this procedure is out of the scope of this article, which focuses on the highly nontrivial identifiability theory as a crucial first step. We will study this direction in the future to fully realize the potential of tensor unfolding and enrich the Q-matrix estimation literature with efficient tensor-based methods.

Financial support. The author is partially supported by the NSF Grant DMS-2210796.

Competing interest. None to declare.

References

Allman, E. S., Matias, C., and Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, *37*(6A):3099–3132.

Chen, J. and de la Torre, J. (2018). Introducing the general polytomous diagnosis modeling framework. *Frontiers in Psychology*, page 1474.

Chen, Y., Culpepper, S., and Liang, F. (2020). A sparse latent class model for cognitive diagnosis. *Psychometrika*, *85*(1):1–33.

Chen, Y., Culpepper, S. A., Chen, Y., and Douglas, J. (2018). Bayesian estimation of the DINA Q matrix. *Psychometrika*, *83*(1):89–108.

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

Culpepper, S. A. (2019). An exploratory diagnostic model for ordinal responses with binary attributes: identifiability and estimation. *Psychometrika*, *84*(4):921–940.

Culpepper, S. A. (2023). A note on weaker conditions for identifying restricted latent class models for binary responses. *Psychometrika*, *88*(1):158–174.

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

DiBello, L. V., Stout, W. F., and Roussos, L. A. (2012). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In *Cognitively diagnostic assessment*, pages 361–389. Routledge.

Fang, G., Liu, J., and Ying, Z. (2019). On the identifiability of diagnostic classification models. *Psychometrika*, *84*(1):19–40.

Gu, Y. (2023). Generic identifiability of the DINA model and blessing of latent dependence. *Psychometrika*, *88*(1):117–131.

Gu, Y. and Xu, G. (2020). Partial identifiability of restricted latent class models. *Annals of Statistics*, *48*(4):2082–2107.

Gu, Y. and Xu, G. (2021). Sufficient and necessary conditions for the identifiability of the Q-matrix. *Statistica Sinica*, *31*:449–472.

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.

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, *51*(3):455–500.

Kruskal, J. B. (1976). More factors than subjects, tests and treatments: an indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika*, *41*:281–293.

Kruskal, J. B. (1977). Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications*, *18*(2):95–138.

Liu, J., Xu, G., and Ying, Z. (2013). Theory of the self-learning Q-matrix. *Bernoulli: official journal of the Bernoulli Society for Mathematical Statistics and Probability*, *19*(5A):1790–1817.

Liu, Y. and Culpepper, S. A. (2024). Restricted latent class models for nominal response data: Identifiability and estimation. *Psychometrika*, *89*(2):592–625.

Ma, W. and de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology*, *69*(3):253–275.

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

Rupp, A. A. and Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement*, *6*(4):219–262.

Stegeman, A. and Sidiropoulos, N. D. (2007). On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition. *Linear Algebra and its Applications*, *420*:540–552.

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

Tatsuoka, K. K. (1983). Rule space: an approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*:345–354.
Templin, J. L. and Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3):287.
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. (2014). The log-linear cognitive diagnostic model (LCDM) as a special case of the general diagnostic model (GDM). *ETS Research Report Series*, *2014*(2):1–13.
von Davier, M. and Lee, Y.-S. (2019). *Handbook of diagnostic classification models*. Cham : Springer International Publishing.
Wayman, E. A., Culpepper, S. A., Douglas, J., and Bowers, J. (2025). A restricted latent class model with polytomous attributes and respondent-level covariates. *Behaviormetrika*, pages 1–29.
Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *The Annals of Statistics*, *45*(2):675–707.
Xu, G. and Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association*, *113*(523):1284–1295.
Xu, G. and Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, *81*(3):625–649.

Appendix

In this Appendix, Section A provides the proofs of the main theoretical results, and Section B provides additional proofs of supporting lemmas and corollaries. Since Theorem 1 (along with Propositions 1 and 2) for the binary CDM is essentially a special case of Theorem 2 (along with Proposition 3) for the polytomous-response CDM, we focus on proving Theorem 2 and Proposition 3.

For each item $j \in [J]$, introduce a notation $\mathrm{pa}(j) := \{k \in [K] : q_{jk} = 1\}$, which denotes the set of attributes required/measured by item j. Here, "pa" is short for "parent", which reflects the fact that the CDM can be viewed from a directed graphical model perspective, and the set of attributes required by an item j can be viewed as the "parent" variables of this item that have directed edges pointing to it.

A. Proofs of the Main Results

We will need to frequently use the matrix Kronecker product "$\otimes$" and matrix Khatri-Rao product "$\odot$" in the proofs. The Khatri-Rao product of matrices is the column-wise Kronecker product (Kolda and Bader, 2009). In particular, consider matrices $\mathbf{A} = (a_{i,j}) \in \mathbb{R}^{m\times r}$, $\mathbf{B} = (b_{i,j}) \in \mathbb{R}^{s\times t}$; and matrices $\mathbf{C} = (c_{i,j}) = (\boldsymbol{c}_{:,1} \mid \cdots \mid \boldsymbol{c}_{:,k}) \in \mathbb{R}^{n\times k}$, $\mathbf{D} = (d_{i,j}) = (\boldsymbol{d}_{:,1} \mid \cdots \mid \boldsymbol{d}_{:,k}) \in \mathbb{R}^{\ell\times k}$, then $\mathbf{A} \bigotimes \mathbf{B} \in \mathbb{R}^{ms\times rt}$ and $\mathbf{C} \bigodot \mathbf{D} \in \mathbb{R}^{n\ell\times k}$:

$$\mathbf{A} \bigotimes \mathbf{B} = \begin{pmatrix} a_{1,1}\mathbf{B} & \cdots & a_{1,r}\mathbf{B} \\ \vdots & \vdots & \vdots \\ a_{m,1}\mathbf{B} & \cdots & a_{m,r}\mathbf{B} \end{pmatrix}, \qquad \mathbf{C} \bigodot \mathbf{D} = \left(\boldsymbol{c}_{:,1} \otimes \boldsymbol{d}_{:,1} \mid \cdots \mid \boldsymbol{c}_{:,k} \otimes \boldsymbol{d}_{:,k}\right).$$

For technical convenience, we will slightly abuse notations when writing the Kronecker product and Khatri-Rao product, in that columns and rows in those products could be subject to permutations. Any of such permutations will not impact the rank properties of a matrix, so they do not affect any proof argument in this paper. Furthermore, we next state the proof of the theoretical results under the special case where all items have the same number of response categories $C_1 = C_2 = \cdots = C_J \geq 2$. We emphasize that this is only for notational simplicity, and all proof arguments in the following indeed remain exactly the same when the C_j's are allowed to differ across different j.

We first state three useful lemmas.

Lemma A.1. *Suppose $M \subseteq [J]$ is an index set satisfying that when j ranges in M, the M sets $\mathrm{pa}(j)$ are mutually disjoint. Then, we have*

$$\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_{\mathrm{pa}(M)}) = \bigotimes_{j\in M} \mathbb{P}(R_j \mid A_{\mathrm{pa}(j)}). \tag{A.1}$$

Now consider an arbitrary set $M \subseteq [p]$. For any set S such that $\mathrm{pa}(M) \subseteq S \subseteq [K]$ with $\mathrm{pa}(M) \neq S$, we have

$$\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_S) = \mathbf{1}^{\mathsf{T}}_{C^{|S|-|\mathrm{pa}(M)|}} \bigotimes \mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_{\mathrm{pa}(M)}). \tag{A.2}$$

Moreover, for any $M_1, M_2 \subseteq [J]$ and $M_1 \cap M_2 = \varnothing$ and any $S \supseteq \mathrm{pa}(M_1) \cup \mathrm{pa}(M_2)$, we have

$$\mathbb{P}(\mathbf{R}_{M_1\cup M_2} \mid \mathbf{A}_S) = \mathbb{P}(\mathbf{R}_{M_1} \mid \mathbf{A}_S) \bigodot \mathbb{P}(\mathbf{R}_{M_2} \mid \mathbf{A}_S). \tag{A.3}$$

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

Lemma A.2. *Suppose* $S_1, S_2 \subseteq [J]$ *are two disjoint sets that satisfy* $\text{pa}(S_1) \subseteq \text{pa}(S_2)$, $\text{pa}(S_1) \neq \text{pa}(S_2)$ *and* $\mathbb{P}(\mathbf{R}_{S_1} \mid \mathbf{A}_{\text{pa}(S_1)})$ *has full column rank* $2^{|\text{pa}(S_1)|}$. *Then*

$$\text{rank}(\mathbb{P}(\mathbf{R}_{S_1 \cup S_2} \mid \mathbf{A}_{\text{pa}(S_2)})) > 2^{|\text{pa}(S_1)|}. \tag{A.4}$$

Lemma A.3. *For* $j_1 \neq j_2 \in [J]$ *with* $|\text{pa}(j_1) \cup \text{pa}(j_2)| > 1$, *the following holds for generic parameters:*

$$\text{rank}(\mathbb{P}(R_{\{j_1,j_2\}} \mid \mathbf{A}_{\text{pa}(j_1)\cup\text{pa}(j_2)})) > 2.$$

Lemma A.4 (Lemma 3.3 in Stegeman and Sidiropoulos (2007)). *Consider two matrices* $\mathbf{A}$ *of size* $I \times R$ *and* $\mathbf{B}$ *of size* $J \times R$.

(a) *If* $krank(\mathbf{A}) = 0$ *or* $krank(B) = 0$, *then* $krank(\mathbf{A} \odot \mathbf{B}) = 0$.
(b) *If* $krank(\mathbf{A}) \geq 1$ *and* $krank(B) \geq 1$, *then*

$$krank(\mathbf{A} \odot \mathbf{B}) \geq \min(krank(\mathbf{A}) + krank(\mathbf{B}) - 1,\ R).$$

A.1. Proof of Proposition 3(a) (and Proposition 1)

The size of the unfolding $[\mathcal{T}]_{\{j_1,j_2\},:}$ is $C^2 \times C^{J-2}$. If $\text{pa}(j_1) = \text{pa}(j_2) = \{k\}$ (that is, $q_{j_1,k} = q_{j_2,k} = 1$ and $q_{j_1,m} = q_{j_2,m} = 0$ for all $m \in [K]\backslash\{k\}$), then we can write $[\mathcal{T}]_{\{j_1,j_2\},:}$ as

$$[\mathcal{T}]_{\{j_1,j_2\},:} = \underbrace{\mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid A_k)}_{C^2\times 2} \cdot \underbrace{\mathbb{P}(A_k, \mathbf{A})}_{2\times 2^K} \cdot \underbrace{\mathbb{P}(\mathbf{R}_{-\{j_1,j_2\}} \mid \mathbf{A})^\top}_{2^K \times C^{J-2}}.$$

Due to the above matrix factorization of the unfolded tensor, $\text{rank}([\mathcal{T}]_{\{j_1,j_2\},:}) \leq D$ must hold, which proves the "if" part of Proposition 1. Next consider the "only if" part. Suppose $\text{pa}(j_1) = \text{pa}(j_2) = \{k\}$ does not hold for any $k \in [K]$, then $\text{pa}(j_1) \cup \text{pa}(j_2)$ is not a singleton set, so $|\text{pa}(j_1) \cup \text{pa}(j_2)| \geq 2$. In this case,

$$\begin{aligned}[\mathcal{T}]_{\{j_1,j_2\},:} &= \underbrace{\mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid \mathbf{A}_{\text{pa}(j_1)\cup\text{pa}(j_2)})}_{\mathbf{P}_1:\ C^2\times 2^{|\text{pa}(j_1)\cup\text{pa}(j_2)|}} \cdot \underbrace{\mathbb{P}(\mathbf{A}_{\text{pa}(j_1)\cup\text{pa}(j_2)}, \mathbf{A})}_{\mathbf{P}_2:\ 2^{\text{pa}(j_1)\cup\text{pa}(j_2)}\times 2^K} \cdot \underbrace{\mathbb{P}(\mathbf{R}_{[J]\backslash\{j_1,j_2\}} \mid \mathbf{A})^\top}_{\mathbf{P}_3:\ 2^K \times C^{J-2}} \\ &= \mathbf{P}_1 \cdot \mathbf{P}_2 \cdot \mathbf{P}_3\end{aligned} \tag{A.5}$$

We next examine the rank of each of the three matrices $\mathbf{P}_1$, $\mathbf{P}_2$, and $\mathbf{P}_3$ on the right-hand side of the above expression.

First, consider $\mathbf{P}_1 = \mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid \mathbf{A}_{\text{pa}(j_1)\cup\text{pa}(j_2)})$, which is a $C^2 \times 2^{|\text{pa}(j_1)\cup\text{pa}(j_2)|}$ matrix. We apply Lemma A.3 to obtain $\text{rank}(\mathbf{P}_1) > H$ holds for generic parameters.

Second, consider $\mathbf{P}_2 = \mathbb{P}(\mathbf{A}_{\text{pa}(j_1)\cup\text{pa}(j_2)}, \mathbf{A})$. For generic distribution of the latent variables $\mathbf{A}$, the matrix $\mathbf{P}_2$ has rank $2^{|\text{pa}(j_1)\cup\text{pa}(j_2)|} \geq 2^2 > 2$. That is, $\mathbf{P}_2$ in (A.5) generically has full rank 2^2, which is greater than 2.

Third, consider $\mathbf{P}_3$. we have an important observation that when Q contains two disjoint copies of I_K and $|\text{pa}(j_1) \cup \text{pa}(j_2)| \geq 2$, the $(J-2) \times K$ submatrix $Q_{[J]\backslash\{j_1,j_2\},:}$ contains at least one submatrix of I_K after some row permutation. To see this, note that if $|\text{pa}(j_1) \cup \text{pa}(j_2)| \geq 2$, then *either* $\text{pa}(j_1)$ and $\text{pa}(j_2)$ are both singleton sets with $\text{pa}(j_1) \neq \text{pa}(j_2)$, *or* at least one of $\text{pa}(j_1)$ and $\text{pa}(j_2)$ is not a singleton set. In either of these two cases, there exists a set $S \subseteq [J]\backslash\{j_1,j_2\}$ such that $|S| = K$ and $Q_{S,:} = I_K$. Applying Lemma A.1 gives

$$\mathbb{P}(\mathbf{R}_S \mid \mathbf{A}) = \bigotimes_{j\in S} \mathbb{P}(R_j \mid A_{\text{pa}(j)}),$$

so $\mathbb{P}(\mathbf{R}_S \mid \mathbf{A})$ generically has full column rank 2^K. Since $S \subseteq [J]\backslash\{j_1,j_2\}$, we can write

$$\mathbb{P}(\mathbf{R}_{[J]\backslash\{j_1,j_2\}} \mid \mathbf{A}) = \mathbb{P}(\mathbf{R}_S \mid \mathbf{A}) \odot \mathbb{P}(\mathbf{R}_{[J]\backslash(S\cup\{j_1,j_2\})} \mid \mathbf{A}).$$

Next, we will use Lemma A.4 to proceed with the proof. One implication of Lemma A.4(b) is that for two matrices $\mathbf{A}$ and $\mathbf{B}$ with the same number of columns, if $\mathbf{A}$ has full column rank and $\mathbf{B}$ does not contain any zero columns, then $\text{krank}(\mathbf{A} \odot \mathbf{B}) \geq \min(R + 1 - 1,\ R) = R$, so $\mathbf{A} \odot \mathbf{B}$ has full column rank. Note that $\mathbb{P}(\mathbf{R}_S \mid \mathbf{A})$ has full column rank and $\mathbb{P}(\mathbf{R}_{[J]\backslash(S\cup\{j_1,j_2\})} \mid \mathbf{A})$ does not contain any zero column vectors since each column of it is a conditional probability vector. So by Lemma A.4, we obtain that $\mathbb{P}(\mathbf{R}_{[J]\backslash\{j_1,j_2\}} \mid \mathbf{A})$ has full column rank 2^K generically. So, $\mathbf{P}_3$ in (A.5) has full row rank 2^K generically.

Going back to (A.5), since both $\mathbf{P}_2$ and $\mathbf{P}_3$ have full rank, the rank of $[\mathcal{T}]_{\{j_1,j_2\},:}$ equals the rank of $\mathbf{P}_1$, which is greater than 2. Now we have proved the "only if" part of the proposition that $\text{rank}([\mathcal{T}]_{\{j_1,j_2\},:}) > 2$ generically if $\text{pa}(j_1) \cup \text{pa}(j_2)$ is not a singleton set. The proof of Proposition 3(a) and Proposition 1 is complete.

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

A.2. Proof of Proposition 3(b) (and Proposition 2)

Define notations B_j and D:

$$B_j = ([K]\backslash\{k\}) \cup \{j\}, \quad D = \{K+1, \ldots, 2K\}.$$

We first prove the "only if" part of Proposition 3(b) (and Proposition 2). It suffices to show that if $k \notin \text{pa}(j)$, then $\text{rank}([\mathcal{T}]_{B_j, D}) \le 2^{K-1}$. Suppose $k \notin \text{pa}(j)$, then $\text{pa}(j) \subseteq \{1, \ldots, k-1, k+1, \ldots, K\}$ and hence $B_j = \{1, \ldots, k-1, k+1, \ldots, K\}$. In this case, we have the following decomposition,

$$[\mathcal{T}]_{B_j, D} = \underbrace{\mathbb{P}(\mathbf{R}_{B_j} \mid \mathbf{A}_{[K]\backslash\{k\}})}_{C^{|B_j|} \times 2^{K-1}} \cdot \underbrace{\mathbb{P}(\mathbf{A}_{[K]\backslash\{k\}}, \mathbf{A}_{[K]})}_{2^{K-1} \times 2^K} \cdot \underbrace{\mathbb{P}(\mathbf{R}_C \mid \mathbf{A}_{[K]})^\top}_{2^K \times C^{|D|}}.$$

Since the first matrix factor above has 2^{K-1} columns, the above decomposition clearly shows that $\text{rank}([\mathcal{T}]_{B_j, D}) \le 2^{K-1}$.

We next prove the "if" part of Proposition 3(b) (and Proposition 2). Since $B_j = ([K]\backslash\{k\}) \cup \{j\}$ and $k \in \text{pa}(j)$, we can rewrite $B_j = \big\{[K]\backslash\text{pa}(j)\big\} \cup \big\{(\text{pa}(j)\backslash\{k\}) \cup \{j\})\big\}$. So we have

$$[\mathcal{T}]_{B_j, D} = \underbrace{\mathbb{P}(\mathbf{R}_{B_j} \mid \mathbf{A}_{1:K})}_{\text{denoted as } \mathbf{P}_1} \cdot \underbrace{\text{diag}(\mathbb{P}(\mathbf{A}_{1:K}))}_{\text{denoted as } \mathbf{P}_2} \cdot \underbrace{\mathbb{P}(\mathbf{R}_D \mid \mathbf{A}_{1:K})^\top}_{\text{denoted as } \mathbf{P}_3}, \tag{A.6}$$

$$\begin{aligned} \text{where } \mathbf{P}_1 &= \mathbb{P}(\mathbf{R}_{[K]\backslash\text{pa}(j)} \mid \mathbf{A}_{[K]\backslash\text{pa}(j)}) \bigotimes \mathbb{P}(\mathbf{R}_{(\text{pa}(j)\backslash\{k\})\cup\{j\})} \mid \mathbf{A}_{\text{pa}(j)}) \\ &= \underbrace{\Big(\bigotimes_{m \in [K]\backslash\text{pa}(j)} \mathbb{P}(R_m \mid A_m)\Big)}_{\text{denoted as } \mathbf{P}_{1,1}} \bigotimes \underbrace{\mathbb{P}(\mathbf{R}_{(\text{pa}(j)\backslash\{k\})\cup\{j\})} \mid \mathbf{A}_{\text{pa}(j)})}_{\text{denoted as } \mathbf{P}_{1,2}}. \end{aligned}$$

First, the matrix factor $\mathbf{P}_{1,1}$ is a Kronecker product of $K - |\text{pa}(j)|$ matrices, each of which has size $C \times H$ and has full column rank H for generic parameters. So, $\mathbf{P}_{1,1}$ has full rank $2^{K-|\text{pa}(j)|}$ generically. Since $\text{rank}(\mathbf{P}_1) = \text{rank}(\mathbf{P}_{1,1}) \cdot \text{rank}(\mathbf{P}_{1,2})$, we next will prove

$$\text{rank}(\mathbf{P}_{1,2}) > 2^{|\text{pa}(j)|-1} \tag{A.7}$$

in order to show $\text{rank}(\mathbf{P}_1) > 2^{K-|\text{pa}(j)|} \cdot 2^{|\text{pa}(j)|-1} = 2^{K-1}$ in Proposition 2. Denote by $Q_{\text{pa}(j)\cup\{j\},\ \text{pa}(j)}$ the submatrix of Q with row indices ranging in $\text{pa}(j) \cup \{j\}$ and column indices ranging in $\text{pa}(j)$. Without loss of generality, after some column permutation $Q_{\text{pa}(j)\cup\{j\},\ \text{pa}(j)}$ can be written as follows and we denote it as $\widetilde{Q}$ for convenience:

$$\widetilde{Q} = Q_{\text{pa}(j)\cup\{j\},\ \text{pa}(j)} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix};$$

in other words, after some column permutation we can always write $Q_{\text{pa}(j)\cup\{j\},\ \text{pa}(j)}$ such that its first column corresponds to the latent variable A_k under our current consideration, and its remaining columns correspond to all the other parent latent variables of R_j. Write $\mathbf{P}_{1,2}$ as

$$\mathbf{P}_{1,2} = \mathbb{P}(\mathbf{R}_{\text{pa}(j)\backslash\{k\}} \mid \mathbf{A}_{\text{pa}(j)}) \bigodot \mathbb{P}(\mathbf{R}_j \mid \mathbf{A}_{\text{pa}(j)}),$$

where the second matrix factor above $\mathbb{P}(\mathbf{R}_j \mid \mathbf{A}_{\text{pa}(j)})$ has size $C \times 2^{|\text{pa}(j)|}$. We next use Lemma A.3 to lower bound the rank of $\mathbf{P}_{1,2}$. Recall that $Q_{1:K,:} = I_K$. Define $S_1 = \text{pa}(j)\backslash\{k\}$ and $S_2 = \{j\}$, then $\text{pa}(S_1) = \text{pa}(j)\backslash\{k\}$ and $\text{pa}(S_2) = \text{pa}(j)$. The sets S_1 and S_2 satisfy that $\text{pa}(S_1) \subseteq \text{pa}(S_2)$ and $\text{pa}(S_2)\backslash\text{pa}(S_1)$ is a singleton set $\{k\}$; additionally,

$$\mathbb{P}(\mathbf{R}_{S_1} \mid \mathbf{A}_{\text{pa}(S_1)}) = \bigotimes_{\ell \in \text{pa}(j)\backslash\{k\}} \mathbb{P}(R_\ell \mid A_\ell)$$

has full column rank $2^{|\text{pa}(S_1)|}$. Therefore, we can use Lemma A.3 to obtain $\text{rank}(\mathbf{P}_{1,2}) > 2^{|\text{pa}(S_1)|}$. According to the argument right after (A.7), it holds that

$$\text{rank}(\mathbf{P}_1) > 2^{K-1}.$$

We now go back to the decomposition in (A.6) that $[\mathcal{T}]_{B_j, D} = \mathbf{P}_1\mathbf{P}_2\mathbf{P}_3$. The rank of the $2^K \times 2^K$ diagonal matrix $\mathbf{P}_2 = \text{diag}(\mathbb{P}(\mathbf{A}_{1:K}))$ is equal to 2^K for generic parameters. The rank of the $2^K \times C^K$ matrix $\mathbf{P}_3$ is also equal to 2^K for generic parameters, because $\mathbf{P}_3 = \bigotimes_{j=K+1}^{2K} \mathbb{P}(R_j \mid \mathbf{A}_{j-K})^\top$. Both $\mathbf{P}_2$ and $\mathbf{P}_3$ have full rank generically by Assumption 2. Therefore, $\text{rank}([\mathcal{T}]_{B_j, D}) = \text{rank}(\mathbf{P}_1\mathbf{P}_2\mathbf{P}_3) = \text{rank}(\mathbf{P}_1) > 2^{K-1}$ generically. This proves the "if" part of Proposition 3(b) (and Proposition 2).

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

Proof of Theorems 1 and 2. It suffices to prove Theorem 2 for the polytomous CDMs because the binary CDMs in Theorem 1 are special cases of the former. Proposition 3 has the following implications. If one exhaustively investigate all pairs of items indexed by $j_1, j_2 \in [J]$, then we can exactly obtain K disjoint sets of item indices $S_1, \ldots, S_K \subseteq [J]$ with the following property: For any $k \in [K]$, it holds that

$$\text{rank}([\mathcal{T}]_{\{j_1,j_2\},:}) \le H, \quad \text{for any } j_1, j_2 \in S_k.$$

The above argument can also uniquely identify the maximum number of latent variables K. This is because K is equal to the number of all the nonoverlapping sets $S_1, S_2, \ldots \subseteq [J]$ such that for any $j_1, j_2 \in S_k$ for each k, it holds that $\text{rank}([\mathcal{T}]_{\{j_1,j_2\},:}) \le 2$. After exhaustively inspecting $\text{rank}([\mathcal{T}]_{\{j_1,j_2\},:})$ for all pairs of item indices $j_1, j_2 \in [J]$, the number of such nonoverlapping sets is uniquely determined and K is uniquely determined constructively. Thus far, we have identified all the row vectors indexed by $j \in \bigcup_{k=1}^{K} S_k$. This is exactly the set of all single-attribute items among the J items. Now, using the assumption that the true Q-matrix contains at least two single-attribute items for each attribute, we can permute the already identified row vectors in Q so that it can be written as $Q = (I_K; I_K, Q^{*\top})^\top$.

Further, for any item j that is not a single-attribute item, in order to determine whether $q_{jk} = 1$ for each $k \in [K]$, we consider the following marginalized unfolding

$$[\mathcal{T}]_{B,D}, \text{ where } B := ([K] \backslash \{k\}) \cup \{j\}, \quad C := \{K+1, \ldots, 2K\}.$$

Here both B and D are non-overlapping sets with cardinality K, so $[\mathcal{T}]_{B,D}$ is a $2^K \times 2^K$ matrix. Since $\sum_{k=1}^{K} q_{jk} \ge 2$, part (b) of Proposition 1 implies that $\text{rank}([\mathcal{T}]_{B,D}) > 2^{K-1}$ holds if and only if $q_{jk} = 1$. Therefore, examining whether the rank of $[\mathcal{T}]_{B,D}$ exceeds 2^{K-1} exhaustively for all non-single-attribute items j and all $k \in [K]$ will identify those row vectors of the Q-matrix corresponding to these items. Now, we have identified all quantities, including K and the Q-matrix, and the proof is complete. □

B. Proof of Supporting Results

B.1. Proof of Lemma A.1

We first prove (A.1). By the conditional independence of the observed variables given the latent parents, we have

$$\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_{\text{pa}(M)}) = \prod_{j=1}^{M} \mathbb{P}(R_j \mid A_{\text{pa}(M)}) = \prod_{j=1}^{M} \mathbb{P}(R_j \mid A_{\text{pa}(j)}),$$

which implies for any observed pattern $\mathbf{r}_M \in \{0, \ldots, V-1\}^{|M|}$ and any latent pattern $\mathbf{a}_{\text{pa}(M)} \in \{0, \ldots, H-1\}^{|\text{pa}(M)|}$, it holds that

$$\mathbb{P}(\mathbf{R}_M = \mathbf{r}_M \mid \mathbf{A}_{\text{pa}(M)} = \boldsymbol{\alpha}_{\text{pa}(M)}) = \prod_{j=1}^{M} \mathbb{P}(R_j = y_j \mid A_{\text{pa}(j)} = \alpha_{\text{pa}(j)}).$$

Since $\text{pa}(j)$ for $j \in M$ are disjoint singleton sets, by following the definition of the Kronecker product of matrices, the above factorization implies $\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_{\text{pa}(M)}) = \bigotimes_{j \in M} \mathbb{P}(R_j \mid A_{\text{pa}(j)})$ and proves (A.1).

We next prove (A.2). First, note that $\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_S)$ has size $C^{|M|} \times 2^{|S|}$, while $\mathbf{1}^\top_{2^{|S|-|\text{pa}(M)|}}$ has size $1 \times 2^{|S|-|\text{pa}(M)|}$ and $\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_{\text{pa}(M)})$ has size $C^{|M|} \times 2^{|\text{pa}(M)|}$. So, the sizes of matrices on the left-hand side and the right-hand side of (A.2) match each other. Further, since $\text{pa}(M) \subsetneq S$, we know that the conditional distribution of $\mathbf{R}_M$ given $\mathbf{A}_S$ only depends on those latent variables belonging to the index set $\text{pa}(M)$; in other words,

$$\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_S) = \mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_{\text{pa}(M)}, \mathbf{A}_{S \backslash \text{pa}(M)}) = \mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_{\text{pa}(M)}),$$

where $\mathbf{A}_{\text{pa}(M)}$ and $\mathbf{A}_{S \backslash \text{pa}(M)}$ are a subvectors of $\mathbf{A}_S$. The above fact implies that, although $\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_S)$ has $2^{|S|}$ columns indexed by all possible configurations of $\mathbf{A}_S \in \{0, \ldots, H-1\}^{|S|}$, actually $\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_S)$ contains only $2^{|\text{pa}(M)|} (< 2^{|S|})$ unique columns. Each unique column is indexed by a unique pattern the subvector $\mathbf{A}_{\text{pa}(M)}$ takes in $\{0, \ldots, H-1\}^{|\text{pa}(M)|}$. So, the matrix $\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_S)$ horizontally stacks $2^{|S|-|\text{pa}(M)|}$ repetitive blocks, each of which equals $\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_{\text{pa}(M)})$. So, we have

$$\begin{aligned}\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_S) &= \underbrace{\left(\mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_{\text{pa}(M)}), \ \cdots, \ \mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_{\text{pa}(M)})\right)}_{2^{|S|-|\text{pa}(M)|} \text{ copies}} \\ &= \mathbf{1}^\top_{2^{|S|-|\text{pa}(M)|}} \bigotimes \mathbb{P}(\mathbf{R}_M \mid \mathbf{A}_S).\end{aligned}$$

We next prove (A.3). Since $S \supseteq \text{pa}(M_1) \cup \text{pa}(M_2)$ holds, $\mathbf{R}_{M_1}$ and $\mathbf{R}_{M_2}$ are conditionally independent given $\mathbf{A}_S$. So, we can use Lemma 12 in Allman et al. (2009) to direct obtain the conclusion that $\mathbb{P}(\mathbf{R}_{M_1 \cup M_2} \mid \mathbf{A}_S) = \mathbb{P}(\mathbf{R}_{M_1} \mid \mathbf{A}_S) \odot \mathbb{P}(\mathbf{R}_{M_2} \mid \mathbf{A}_S)$. The proof of Lemma A.1 is complete.

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

B.2. Proof of Lemma A.2

Since $\text{pa}(S_1) \subsetneq \text{pa}(S_2)$, we use Lemma A.1 to write

$$\mathbb{P}(R_{S_1} \mid \mathbf{A}_{\text{pa}(S_2)}) = \mathbf{1}_{2^{|\text{pa}(S_2)\setminus\text{pa}(S_1)|}}^\top \bigotimes \mathbb{P}(R_{S_1} \mid \mathbf{A}_{\text{pa}(S_2)}),$$

and

$$\mathbb{P}(\mathbf{R}_{S_1\cup S_2} \mid \mathbf{A}_{\text{pa}(S_2)}) = \left(\mathbf{1}_{2^{|\text{pa}(S_2)\setminus\text{pa}(S_1)|}}^\top \bigotimes \mathbb{P}(R_{S_1} \mid \mathbf{A}_{\text{pa}(S_1)})\right) \bigodot \mathbb{P}(\mathbf{R}_{S_2} \mid \mathbf{A}_{\text{pa}(S_2)}). \tag{B.1}$$

Define notation

$$\mathbf{F} = \mathbb{P}(R_{S_1} \mid \mathbf{A}_{\text{pa}(S_1)});$$
$$\mathbf{E}_{\boldsymbol{\alpha}_\ell} = \mathbb{P}(\mathbf{R}_{S_2} \mid \mathbf{A}_{\text{pa}(S_1)}, \mathbf{A}_{\text{pa}(S_2)\setminus\text{pa}(S_1)} = \boldsymbol{\alpha}_\ell), \quad \boldsymbol{\alpha}_\ell \in \{0,1,\ldots,H-1\}^{|\text{pa}(S_2)\setminus\text{pa}(S_1)|}.$$

For each fixed pattern $\boldsymbol{\alpha}_\ell \in \{0,1\}^{|\text{pa}(S_2)\setminus\text{pa}(S_1)|}$, the $\mathbf{E}_{\boldsymbol{\alpha}_\ell}$ is a $C^{|S_2|} \times 2^{|\text{pa}(S_1)|}$ matrix with rows indexed by all possible configurations of R_{S_2} and columns by all possible configurations of $\mathbf{A}_{\text{pa}(S_1)}$. By the condition in the lemma, matrix $\mathbf{F}$ has full column rank $2^{|\text{pa}(S_1)|}$. For notational simplicity, denote this number $2^{|\text{pa}(S_1)|} =: R$. After possibly rearranging the columns of $\mathbf{1}_{2^{\text{pa}(S_2)\setminus\text{pa}(S_1)}}^\top \otimes \mathbb{P}(R_{S_1} \mid \mathbf{A}_{\text{pa}(S_1)})$ and $\mathbb{P}(\mathbf{R}_{S_2} \mid \mathbf{A}_{\text{pa}(S_2)})$ via a common permutation, we can write

$$\mathbf{1}_{2^{\text{pa}(S_2)\setminus\text{pa}(S_1)}}^\top \bigotimes \mathbb{P}(R_{S_1} \mid \mathbf{A}_{\text{pa}(S_1)}) = (\underbrace{\mathbf{F}, \mathbf{F}, \ldots, \mathbf{F}}_{R=2^{\text{pa}(S_2)\setminus\text{pa}(S_1)} \text{ copies}});$$
$$\mathbb{P}(\mathbf{R}_{S_2} \mid \mathbf{A}_{\text{pa}(S_2)}) = (\mathbf{E}_{\boldsymbol{\alpha}_1}, \mathbf{E}_{\boldsymbol{\alpha}_2}, \ldots, \mathbf{E}_{\boldsymbol{\alpha}_R}).$$

Therefore, the matrix $\mathbb{P}(\mathbf{R}_{S_1\cup S_2} \mid \mathbf{A}_{\text{pa}(S_2)})$ in (B.1) can be written as

$$\mathbb{P}(\mathbf{R}_{S_1\cup S_2} \mid \mathbf{A}_{\text{pa}(S_2)}) = \left(\mathbf{F} \bigodot \mathbf{E}_{\boldsymbol{\alpha}_1}, \mathbf{F} \bigodot \mathbf{E}_0, \ \ldots, \ \mathbf{F} \bigodot \mathbf{E}_{\boldsymbol{\alpha}_R}\right). \tag{B.2}$$

Next, we will prove that the matrix $\mathbb{P}(\mathbf{R}_{S_1\cup S_2} \mid \mathbf{A}_{\text{pa}(S_2)})$ has rank at least $R+1$ under Assumption 2. Denote the columns of $\mathbf{F}$ by $\mathbf{F}_1, \ldots, \mathbf{F}_R$, and the columns of $\mathbf{E}_{\boldsymbol{\alpha}_h}$ by $\mathbf{E}_{\boldsymbol{\alpha}_h,1}, \ldots, \mathbf{E}_{\boldsymbol{\alpha}_h,R}$. Note that by definition, each $\mathbf{E}_{\boldsymbol{\alpha}_h,m}$ is a vector describing the conditional distribution of R_{S_2} given a specific configuration of $\mathbf{A}_{\text{pa}(S_2)}$ when $\mathbf{A}_{\text{pa}(S_2)\setminus\text{pa}(S_1)}$ is fixed to be $\boldsymbol{\alpha}_h$. Since $\text{pa}(S_1) \subsetneq \text{pa}(S_2)$, there exists some $k \in \text{pa}(S_1)$ but $k \notin \text{pa}(S_2)$. Under the non-DINA/DINO assumption in Definition 3, there exist 2 different vectors $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_R\}$ where $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ differ only in the entry corresponding to $A_k = 0,1$, in which there exist $h_1 \neq h_2 \in \{0,1\}$ and some $r \in \{1,\ldots,R\}$ such that

$$\mathbf{E}_{\boldsymbol{\beta}_{h_1},r} \neq \mathbf{E}_{\boldsymbol{\beta}_{h_2},r}. \tag{B.3}$$

Suppose that there exists a vector $\mathbf{x} = (x_1, \ldots, x_R, x_{R+1})^\top$ such that

$$x_1\mathbf{F}_1 \bigotimes \mathbf{E}_{\boldsymbol{\beta}_{h_1},1} + \cdots + x_R\mathbf{F}_R \bigotimes \mathbf{E}_{\boldsymbol{\beta}_{h_1},R} + x_{R+1}\mathbf{F}_r \bigotimes \mathbf{E}_{\boldsymbol{\beta}_{h_2},r} = \mathbf{0}.$$

Rearranging the terms above gives

$$\sum_{m\in[R]\setminus\{r\}} x_m\mathbf{F}_m \bigotimes \mathbf{E}_{\boldsymbol{\beta}_{h_1},m} + \mathbf{F}_r \bigotimes (x_r\mathbf{E}_{\boldsymbol{\beta}_{h_1},r} + x_{R+1}\mathbf{E}_{\boldsymbol{\beta}_{h_2},r}) = \mathbf{0}. \tag{B.4}$$

Define a matrix $\widetilde{\mathbf{E}}$ whose r-th column vector is $x_r\mathbf{E}_{\boldsymbol{\beta}_{h_1},r} + x_{R+1}\mathbf{E}_{\boldsymbol{\beta}_{h_2},r}$, and for any $m \in [R]$ and $m \neq r$, the m-th column vector of $\widetilde{\mathbf{E}}$ is $\mathbf{E}_{\boldsymbol{\beta}_{h_1},m}$. Define a R-dimensional vector $\widetilde{\mathbf{y}}$ with $y_r = 1$ and $y_m = x_m$ for any $m \in [R]$ and $m \neq r$. Then (B.4) can be rewritten as

$$(\mathbf{F} \bigodot \widetilde{\mathbf{E}})\mathbf{y} = \mathbf{0}. \tag{B.5}$$

Since $\mathbf{y}$ above is not a zero vector, we have that $\text{rank}(\mathbf{F} \bigodot \widetilde{\mathbf{E}}) < R$. We next will show that this necessarily implies $x_r\mathbf{E}_{\boldsymbol{\beta}_{h_1},r} + x_{R+1}\mathbf{E}_{\boldsymbol{\beta}_{h_2},r} = \mathbf{0}$.

We introduce the definition of the Kruskal rank. The Kruskal rank of a matrix is the largest integer M such that every M columns of this matrix are linearly independent. Denote the Kruskal rank of any matrix $\mathbf{B}$ by $\text{krank}(\mathbf{B})$. It is easy to see that if a matrix has full column rank, then its Kruskal rank equals its rank. Also, a matrix $\mathbf{B}$ contains a zero column vector if and only if $\text{krank}(\mathbf{B}) = 0$.

Now consider the matrix $\mathbf{F} \bigodot \widetilde{\mathbf{E}}$ in (B.5). Since $\mathbf{F}$ has full column rank R, then $\text{krank}(\mathbf{F}) = \text{rank}(\mathbf{F}) = R$. For any $m \in [R]\setminus\{r\}$, the m-th column vector of $\widetilde{\mathbf{E}}$ is a conditional probability vector for R_j given a fixed pattern of $\mathbf{A}_{\text{pa}(j)}$, therefore $\mathbf{E}_{\boldsymbol{\beta}_{h_1},m} \neq \mathbf{0}$. So, if $x_r\mathbf{E}_{\boldsymbol{\beta}_{h_1},r} + x_{R+1}\mathbf{E}_{\boldsymbol{\beta}_{h_2},r} \neq \mathbf{0}$, then $\widetilde{\mathbf{E}}$ does not have any zero column vectors and $\text{krank}(\widetilde{\mathbf{E}}) \geq 1$. In this case, the following holds by Lemma A.4(b),

$$\begin{aligned}\text{krank}(\mathbf{F} \bigodot \widetilde{\mathbf{E}}) &\geq \min(\text{krank}(\mathbf{F}) + \text{krank}(\widetilde{\mathbf{E}}) - 1,\ R)\\ &= \min(R + \text{krank}(\widetilde{\mathbf{E}}) - 1,\ R)\\ &\geq \min(R + 1 - 1,\ R) = R.\end{aligned}$$

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

Then, $\text{rank}(\mathbf{F}\odot\widetilde{\mathbf{E}}) = \text{krank}(\mathbf{F}\odot\widetilde{\mathbf{E}}) = R$, which contradicts with (B.5) because the vector $\mathbf{y}$ there is not a zero vector with $y_r = 1$ and hence $\mathbf{F}\odot\widetilde{\mathbf{E}}$ should not have full column rank. This contradiction implies that

$$x_r\mathbf{E}_{\boldsymbol{\beta}_{h_1},r} + x_{R+1}\mathbf{E}_{\boldsymbol{\beta}_{h_2},r} = \mathbf{0}.$$

Now note that each of $\mathbf{E}_{\boldsymbol{\beta}_{h_1},r}$ and $\mathbf{E}_{\boldsymbol{\beta}_{h_2},r}$ is a conditional probability vector describing the distribution of R_{S_2} given a fixed pattern of $\mathbf{A}_{\text{pa}(S_2)}$. Therefore, $\mathbf{1}^\top\mathbf{E}_{\boldsymbol{\beta}_{h_1},r} = \mathbf{1}^\top\mathbf{E}_{\boldsymbol{\beta}_{h_1},r} = 0$, and we further have $x_r + x_{R+1} = 0$. Plugging the above expression $x_r\mathbf{E}_{\boldsymbol{\beta}_{h_1},r} + x_{R+1}\mathbf{E}_{\boldsymbol{\beta}_{h_2},r} = \mathbf{0}$ back into (B.4) further gives

$$\sum_{m\in[R]\backslash\{r\}} x_m\mathbf{F}_m\bigotimes\mathbf{E}_{\boldsymbol{\beta}_{h_1},m} = \mathbf{0}. \tag{B.6}$$

We next use Lemma A.4 to show that $\{\mathbf{F}_m\otimes\mathbf{E}_{\boldsymbol{\beta}_{h_1},m}:\ m\in[R]\backslash\{r\}\}$ are $R-1$ linearly independent vectors. To show this, first note that (B.6) can be written as

$$(\mathbf{F}_{-r}\bigodot\widetilde{\mathbf{E}}_{-r})\mathbf{x}_{-r} = \mathbf{0},$$

where matrix $\mathbf{F}_{-r}$ contains all but the r-th column of matrix $\mathbf{F}$, matrix $\widetilde{\mathbf{E}}_{-r}$ contains all but the r-th column of matrix $\widetilde{\mathbf{E}}$, and vector $\mathbf{x}_{-r}$ contains all but the r-th entry of the vector $\mathbf{x}$. Now, because $\text{krank}(\mathbf{F}_{-r}) = \text{rank}(\mathbf{F}_{-r}) = R-1$ and $\text{krank}(\widetilde{\mathbf{E}}_{-r}) \geq 1$, Lemma A.4(b) implies that

$$\text{krank}(\mathbf{F}_{-r}\bigodot\widetilde{\mathbf{E}}_{-r}) \geq \min(R-1+\text{krank}(\widetilde{\mathbf{E}}_{-r})-1,\ R-1) \geq R-1.$$

Since $\mathbf{F}_{-r}\odot\widetilde{\mathbf{E}}_{-r}$ contains $R-1$ columns, the above inequality gives that $\text{rank}(\mathbf{F}_{-r}\odot\widetilde{\mathbf{E}}_{-r}) = \text{krank}(\mathbf{F}_{-r}\odot\widetilde{\mathbf{E}}_{-r}) = R-1$. In other words, the $R-1$ vectors in $\{\mathbf{F}_m\otimes\mathbf{E}_{\boldsymbol{\beta}_{h_1},m}:\ m\in[R]\backslash\{r\}\}$ are linearly independent. Therefore, from (B.6) we have that $x_m = 0$ for all $m\in[R]\backslash\{r\}$. Recall we aim to examine whether there exists a nonzero vector $\mathbf{x} = (x_1,\ldots,x_R,x_{R+1})$ such that (B.4) holds, and we have already shown that under (B.4), the only two potentially nonzero entries in the vector $\mathbf{x}$ are x_r and x_{R+1}. If $(x_r,x_{R+1}) \neq (0,0)$, then $x_r = -x_{R+1} \neq 0$, and it indicates

$$\mathbf{E}_{\boldsymbol{\beta}_{h_1},m} = \mathbf{E}_{\boldsymbol{\beta}_{h_2},m}.$$

The above equality contradicts the earlier (B.3) that $\mathbf{E}_{\boldsymbol{\beta}_{h_1},m} \neq \mathbf{E}_{\boldsymbol{\beta}_{h_2},m}$ which follows from Assumption 2, therefore $x_r = x_{R+1} = 0$ must hold. In summary, we have shown that the $R+1$ vectors $\mathbf{F}_1\otimes\mathbf{E}_{\boldsymbol{\beta}_{h_1},1},\ \ldots,\ \mathbf{F}_R\otimes\mathbf{E}_{\boldsymbol{\beta}_{h_1},R},\ \mathbf{F}_r\otimes\mathbf{E}_{\boldsymbol{\beta}_{h_2},r}$ are linearly independent, which means the matrix $\mathbb{P}(R_{S_1\cup S_2}\mid\mathbf{A}_{\text{pa}(S_2)})$ in (B.2) contains at least $R+1$ linearly independent column vectors and $\text{rank}(\mathbb{P}(R_{S_1\cup S_2}\mid\mathbf{A}_{\text{pa}(S_2)})) \geq R+1 = 2^{|\text{pa}(S_1)|}+1$. The proof of Lemma A.2 is complete.

B.3. Proof of Lemma A.3

Consider all possible cases: Case (a), $\text{pa}(j_1)\cap\text{pa}(j_2) = \varnothing$; Case (b), $\text{pa}(j_1) = \text{pa}(j_2)$ with $|\text{pa}(j_1)| = |\text{pa}(j_2)| \geq 2$; Case (c), $\text{pa}(j_1)\subsetneq\text{pa}(j_2)$ with $|\text{pa}(j_2)| \geq 2$; and Case (d), $\text{pa}(j_1)\nsubseteq\text{pa}(j_2)$ and $\text{pa}(j_2)\nsubseteq\text{pa}(j_1)$ with $|\text{pa}(j_1)| \geq 2$, $|\text{pa}(j_2)| \geq 2$. Denote $\mathbf{P} = \mathbb{P}(R_{\{j_1,j_2\}}\mid\mathbf{A}_{\text{pa}(j_1)\cup\text{pa}(j_2)})$.

Case (a): Since $\text{pa}(j_1)\cap\text{pa}(j_2) = \varnothing$, we have

$$\mathbf{P} = \mathbb{P}(R_{\{j_1,j_2\}}\mid\mathbf{A}_{\text{pa}(j_1)\cup\text{pa}(j_2)}) = \mathbb{P}(\mathbf{R}_{j_1}\mid\mathbf{A}_{\text{pa}(j_1)})\bigotimes\mathbb{P}(\mathbf{R}_{j_2}\mid\mathbf{A}_{\text{pa}(j_2)}),$$

so $\text{rank}(\mathbf{P}) = \text{rank}(\mathbb{P}(\mathbf{R}_{j_1}\mid\mathbf{A}_{\text{pa}(j_1)}))\cdot\text{rank}(\mathbb{P}(\mathbf{R}_{j_2}\mid\mathbf{A}_{\text{pa}(j_2)})) \geq 2^2 > 2$.

Case (b): We have $\text{pa}(j_1)\cup\text{pa}(j_2) = \text{pa}(j_1) = \text{pa}(j_2)$, so both $\mathbb{P}(\mathbf{R}_{j_1}\mid\mathbf{A}_{\text{pa}(j_1)}) = \mathbb{P}(\mathbf{R}_{j_1}\mid\mathbf{A}_{\text{pa}(j_1)\cup\text{pa}(j_2)})$ and $\mathbb{P}(\mathbf{R}_{j_2}\mid\mathbf{A}_{\text{pa}(j_2)}) = \mathbb{P}(\mathbf{R}_{j_2}\mid\mathbf{A}_{\text{pa}(j_1)\cup\text{pa}(j_2)})$ are generic conditional probability matrices without particular structures (i.e., without equality between certain columns). We can write

$$\mathbf{P} = \mathbb{P}(\mathbf{R}_{j_1}\mid\mathbf{A}_{\text{pa}(j_1)\cup\text{pa}(j_2)})\bigodot\mathbb{P}(\mathbf{R}_{j_2}\mid\mathbf{A}_{\text{pa}(j_1)\cup\text{pa}(j_2)}).$$

Under the non-DINA/DINO assumption in Definition 3, we apply Lemma 13 in Allman et al. (2009) to obtain that $\text{rank}(\mathbf{P}) = \min(C^2, 2^{|\text{pa}(j_1)\cup\text{pa}(j_2)|}) = 2^{|\text{pa}(j_1)\cup\text{pa}(j_2)|} > 2$ holds for generic parameters.

Case (c): $\text{pa}(j_1)\subsetneq\text{pa}(j_2)$ with $|\text{pa}(j_2)| \geq 2$. Denote $L = 2^{|\text{pa}(j_2)\backslash\text{pa}(j_1)|}$. We can apply Lemma A.1 to write

$$\begin{aligned}\mathbf{P} &= \mathbb{P}(\mathbf{R}_{j_1}\mid\mathbf{A}_{\text{pa}(j_2)})\bigodot\mathbb{P}(\mathbf{R}_{j_2}\mid\mathbf{A}_{\text{pa}(j_2)})\\ &= \left(\mathbf{1}_{2^{|\text{pa}(j_2)\backslash\text{pa}(j_1)|}}^\top\bigotimes\mathbb{P}(\mathbf{R}_{j_1}\mid\mathbf{A}_{\text{pa}(j_1)})\right)\bigodot\mathbb{P}(\mathbf{R}_{j_2}\mid\mathbf{A}_{\text{pa}(j_2)})\\ &= \left(\mathbb{P}(\mathbf{R}_{j_1}\mid\mathbf{A}_{\text{pa}(j_1)})\bigodot\mathbf{E}_{\boldsymbol{\alpha}_1},\ \ldots,\ \mathbb{P}(\mathbf{R}_{j_1}\mid\mathbf{A}_{\text{pa}(j_1)})\bigodot\mathbf{E}_{\boldsymbol{\alpha}_L}\right),\end{aligned} \tag{B.7}$$

where each $\mathbf{E}_{\boldsymbol{\alpha}_\ell}$ denotes a $C\times 2^{|\text{pa}(j_1)|}$ submatrix of $\mathbb{P}(\mathbf{R}_{j_2}\mid\mathbf{A}_{\text{pa}(j_2)})$. Here, $\boldsymbol{\alpha}_\ell\in\{0,1\}^{|\text{pa}(j_2)\backslash\text{pa}(j_1)|}$ indexes a specific configuration of $\mathbf{A}_{\text{pa}(j_2)\backslash\text{pa}(j_1)}$, and the $2^{|\text{pa}(j_1)|}$ columns of $\mathbf{E}_{\boldsymbol{\alpha}_\ell}$ are indexed by all possible configurations $\mathbf{A}_{\text{pa}(j_1)}$ can take. Therefore, each matrix $\mathbf{E}_{\boldsymbol{\alpha}_\ell}$ collections the conditional probability vectors of R_{j_2} when fixing $\mathbf{A}_{\text{pa}(j_2)\backslash\text{pa}(j_1)} = \boldsymbol{\alpha}_\ell$ and varying $\mathbf{A}_{\text{pa}(j_1)}$. Note that

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press

$\mathbf{E}_{\boldsymbol{\alpha}_\ell}$ is a generic matrix without equality constraints between certain columns. Therefore, we apply Lemma 13 in Allman et al. (2009) to obtain that the following holds for generic parameters:

$$\operatorname{rank}(\mathbb{P}(\mathbf{R}_{j_1} \mid \mathbf{A}_{\operatorname{pa}(j_1)}) \odot \mathbf{E}_{\boldsymbol{\alpha}_\ell}) = \min(C^2, 2^{|\operatorname{pa}(j_1)|}).$$

Now, if $|\operatorname{pa}(j_1)| > 1$, then the above inequality gives $\operatorname{rank}(\mathbb{P}(\mathbf{R}_{j_1} \mid \mathbf{A}_{\operatorname{pa}(j_1)}) \odot \mathbf{E}_{\boldsymbol{\alpha}_\ell}) \geq 2^2 > 2$, which further implies $\operatorname{rank}(\mathbf{P}) \geq \operatorname{rank}(\mathbb{P}(\mathbf{R}_{j_1} \mid \mathbf{A}_{\operatorname{pa}(j_1)}) > 2$. So, it remains to consider the case with $|\operatorname{pa}(j_1)| = 1$. In this case, $\mathbb{P}(\mathbf{R}_{j_1} \mid \mathbf{A}_{\operatorname{pa}(j_1)})$ has full column rank 2 by Assumption 2. Define $S_1 = \{j_1\}$ and $S_2 = \{j_2\}$, then the conditions $S_1 \subsetneq S_2$ and $\mathbb{P}(\mathbf{R}_{S_1} \mid \mathbf{A}_{\operatorname{pa}(S_1)})$ in Lemma A.2 are satisfied. So Lemma A.2 gives that the rank of $\mathbf{P}$ is greater than $2^{|\operatorname{pa}(S_1)|} = 2$. In summary, in Case (c), $\operatorname{rank}(\mathbf{P}) > 2$.

Case (d): $\operatorname{pa}(j_1) \nsubseteq \operatorname{pa}(j_2)$ and $\operatorname{pa}(j_2) \nsubseteq \operatorname{pa}(j_1)$ with $|\operatorname{pa}(j_1)| \geq 2$, $|\operatorname{pa}(j_2)| \geq 2$. Note that $\operatorname{pa}(j_1) \subsetneq \operatorname{pa}(S_2)$ and $\operatorname{pa}(j_2) \subsetneq \operatorname{pa}(S_1)$ imply that both $\operatorname{pa}(j_1)$ and $\operatorname{pa}(j_2)$ contain at least two elements. First consider the $C^2 \times 2^{|\operatorname{pa}(j_2)|}$ matrix

$$\mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid \mathbf{A}_{\operatorname{pa}(j_2)}) = \mathbf{1}^\top_{2^{|\operatorname{pa}(j_2)\setminus\operatorname{pa}(j_1)|}} \otimes \mathbb{P}(\mathbf{R}_{j_1} \mid \mathbf{A}_{\operatorname{pa}(j_1)\cap\operatorname{pa}(j_2)}) \odot \mathbb{P}(\mathbf{R}_{j_2} \mid \mathbf{A}_{\operatorname{pa}(j_2)}),$$

which takes the same form as (B.7) in case (c) and falls into the same setting considered in case (c). Therefore, we know that $\operatorname{rank}(\mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid \mathbf{A}_{\operatorname{pa}(j_2)})) > 2$ holds generically. Now, we claim that this matrix $\mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid \mathbf{A}_{\operatorname{pa}(j_2)})$ has the following linear transformation relationship with the original matrix $\mathbf{P} = \mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid \mathbf{A}_{\operatorname{pa}(j_1)\cup\operatorname{pa}(j_2)})$:

$$\mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid \mathbf{A}_{\operatorname{pa}(j_2)}) = \mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid \mathbf{A}_{\operatorname{pa}(j_1)\cup\operatorname{pa}(j_2)}) \cdot \mathbf{S} = \mathbf{P} \cdot \mathbf{S}, \tag{B.8}$$

where $\mathbf{S}$ has mutually orthogonal columns and hence full column rank. To see that this claim is true, we only need to note that for any vectors $\mathbf{y}_{\{j_1,j_2\}} \in \{0,\ldots,C-1\}^2$ and $\boldsymbol{\alpha}_{\operatorname{pa}(j_2)} \in \{0,1\}^{|\operatorname{pa}(j_2)|}$, it holds that

$$\begin{aligned}
&\mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} = \mathbf{y}_{\{j_1,j_2\}} \mid \mathbf{A}_{\operatorname{pa}(j_2)} = \boldsymbol{\alpha}_{\operatorname{pa}(j_2)}) \\
= &\sum_{\boldsymbol{\alpha}_{\operatorname{pa}(j_1)} \in \{0,1\}^{|\operatorname{pa}(j_1)|}} \mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} = \mathbf{y}_{\{j_1,j_2\}} \mid \mathbf{A}_{\operatorname{pa}(j_2)} = \boldsymbol{\alpha}_{\operatorname{pa}(j_2)}, \mathbf{A}_{\operatorname{pa}(j_1)} = \boldsymbol{\alpha}_{\operatorname{pa}(j_1)}).
\end{aligned}$$

This equality means that to get the smaller conditional probability table $\mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid \mathbf{A}_{\operatorname{pa}(j_2)})$, we can just sum up appropriate columns in the larger conditional probability table $\mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid \mathbf{A}_{\operatorname{pa}(j_1)\cup\operatorname{pa}(j_2)})$, where the $2^{|\operatorname{pa}(j_1)\cup\operatorname{pa}(j_2)|} \times 2^{|\operatorname{pa}(j_2)|}$ matrix $\mathbf{S}$ has binary entries that reflect this summation. Specifically, for any $\boldsymbol{\alpha}_{\operatorname{pa}(j_1)\cup\operatorname{pa}(j_2)}$ and $\boldsymbol{\beta}_{\operatorname{pa}(j_2)}$, the corresponding entry in S is

$$S_{\boldsymbol{\alpha}_{\operatorname{pa}(j_1)\cup\operatorname{pa}(j_2)}, \boldsymbol{\beta}_{\operatorname{pa}(j_2)}} = \mathbb{1}(\boldsymbol{\alpha}_{\operatorname{pa}(j_1)\cup\operatorname{pa}(j_2)} \text{ restricted to } \operatorname{pa}(j_2) \text{ equals } \boldsymbol{\beta}_{\operatorname{pa}(j_2)}).$$

Since each column in $\mathbf{S}$ is indexed by a different configuration $\boldsymbol{\beta}_{\operatorname{pa}(j_2)}$, the column vectors of $\mathbf{S}$ have mutually disjoint support and are orthogonal. Now we have shown that the claim about (B.8) is correct. Therefore, (B.8) implies that

$$\operatorname{rank}(\mathbf{P}) \geq \operatorname{rank}(\mathbb{P}(\mathbf{R}_{\{j_1,j_2\}} \mid \mathbf{A}_{\operatorname{pa}(j_2)})) > 2.$$

Now we have proved the conclusion of Lemma A.3 for all possible cases (a), (b), (c), and (d). The proof is complete.

Cite this article: Gu, Y. (2026). Constructive *Q*-Matrix Identifiability via Novel Tensor Unfolding. *Psychometrika*, 131–150. https://doi.org/10.1017/psy.2025.10078

https://doi.org/10.1017/psy.2025.10078 Published online by Cambridge University Press