

Identifiability of Hierarchical Latent Attribute Models

Yuqi Gu¹ and Gongjun Xu²

¹*Columbia University* and ²*University of Michigan*

Abstract: Hierarchical Latent Attribute Models (HLAMs) are a family of discrete latent variable models that are attracting increasing attention in educational, psychological, and behavioral sciences. The key ingredients of an HLAM include a binary structural matrix and a directed acyclic graph specifying hierarchical constraints on the configurations of latent attributes. These components encode practitioners' design information and carry important scientific meanings. Despite the popularity of HLAMs, the fundamental identifiability issue remains unaddressed. The existence of the attribute hierarchy graph leads to degenerate parameter space, and the potentially unknown structural matrix further complicates the identifiability problem. This paper addresses this issue of identifying the latent structure and model parameters underlying an HLAM. We develop sufficient and necessary identifiability conditions. These results directly and sharply characterize the different impacts on identifiability cast by different attribute types in the graph. The proposed conditions not only provide insights into diagnostic test designs under the attribute hierarchy, but also serve as tools to assess the validity of an estimated HLAM.

Key words and phrases: Identifiability, Attribute hierarchy graph, \mathbf{Q} -matrix, Cognitive diagnosis.

1. Introduction

Latent attribute models are a family of discrete latent variable models popular in multiple scientific disciplines, including cognitive diagnosis in educational assessments (Junker and Sijtsma, 2001; von Davier, 2008; Henson et al., 2009; Rupp et al., 2010; de la Torre, 2011; Wang et al., 2018), psychiatric diagnosis of mental disorders (Templin and Henson, 2006;

de la Torre et al., 2018), and epidemiological and medical measurement studies (Wu et al., 2017; O'Brien et al., 2019). Based on subjects' responses (often binary) to a set of items, a latent attribute model enables fine-grained inference on subjects' statuses of an underlying set of latent traits; this further allows for clustering the population into interpretable subgroups based on the inferred attribute patterns. In a latent attribute model, each attribute is often assumed binary and carries specific scientific meaning. For example, in an educational assessment, the observed responses are students' correct or wrong answers to a set of test items, and the latent attributes indicate students' binary states of mastery or deficiency of certain skills measured by the assessment (Junker and Sijtsma, 2001; von Davier, 2008; Rupp et al., 2010). On top of this, the dependence among the latent attributes can be further modeled to incorporate practitioners' prior knowledge. A particularly popular and powerful way of modeling attribute dependence in educational and psychological studies is to enforce hard constraints on the hierarchical configurations of the attributes. Specifically, educational experts often postulate some prerequisite relations exist among the binary skill attributes, such that mastering some skills serve as a prerequisite for mastering some others (Leighton et al., 2004). Such a family of *Hierarchical Latent Attribute Models* (HLAMs) are [attracting](#) increasing attention in cognitive diagnostic applications in recent years; see Leighton et al. (2004); Gierl et al. (2007); Templin and Bradshaw (2014); Wang and Lu (2020). Despite the popularity, the fundamental identifiability issue of HLAMs remains unaddressed. This paper fills this gap and provides the identifiability theory for HLAMs.

HLAMs have close connections with many other popular statistical and machine learning models. Since each possible configuration of the discrete attributes represents a pattern defining a latent subpopulation, the HLAM can be viewed as a structured mixture model

(McLachlan and Peel, 2004) and gives rises to model-based clustering (Fraley and Raftery, 2002) of multivariate categorical data. HLAMs are related to several multivariate discrete latent variable models in the machine learning literature, including latent tree graphical models (Choi et al., 2011), restricted Boltzmann machines (Hinton, 2002), latent feature models (Ghahramani and Griffiths, 2006), but with the following two key differences. First, the observed variables are assumed to have certain structured dependence on the latent attributes. This dependence is summarized by a structural matrix, the so-called **Q**-matrix (Tatsuoka, 1990), to encode scientific interpretations. The second key feature is that HLAMs incorporate the hierarchical structure among the latent attributes. For instance, in educational cognitive diagnosis, the possession of certain skill attributes are often assumed to be the prerequisite for possessing some others (Leighton et al., 2004; Templin and Bradshaw, 2014).

The real-world applications of HLAMs are challenged by the identifiability issues of the attribute hierarchy, the structural **Q**-matrix, and other model parameters. First, in many applications, the attribute hierarchy and the structural **Q**-matrix are specified by the domain experts based on their understanding of the diagnostic tests. Such specification could be subjective and may not reflect the underlying truth. Second, the attribute hierarchy and the **Q**-matrix may even be entirely unknown in exploratory data analysis, where researchers hope to identify and estimate these quantities directly from the observed data. In both of the above situations, a fundamental yet open question is whether and when the attribute hierarchy and even the structural **Q**-matrix are identifiable. The identifiability of HLAMs has a close connection to the uniqueness of tensor decompositions, as the probability distribution of an HLAM can be written as a mixture of highly constrained higher-order tensors. Particularly,

HLAMs can be viewed as a special family of restricted latent class models, with the **Q**-matrix imposing constraints on the model parameters. However, related works on the identifiability of latent class models and uniqueness of tensor decompositions (e.g. Allman et al., 2009; Anandkumar et al., 2014) cannot be directly applied to HLAMs due to the constraints induced by the **Q**-matrix.

To tackle identifiability under such structural constraints, some recent works (Xu, 2017; Xu and Shang, 2018; Gu and Xu, 2019b; Fang et al., 2019; Gu and Xu, 2020, 2019a; Chen et al., 2020) proposed identifiability conditions for latent attribute models. However, most of them (Xu, 2017; Xu and Shang, 2018; Gu and Xu, 2019b; Fang et al., 2019; Chen et al., 2020) considered scenarios without any attribute hierarchy; Gu and Xu (2020) assumed both the true **Q**-matrix and true configurations of attribute patterns are known and fixed; Gu and Xu (2019a) considered the problem of learning the set of truly existing attribute patterns but assumed the **Q**-matrix is correctly specified beforehand. All these previous works did not directly take into account the hierarchical graphical structure of the attribute hierarchy, therefore their results can not provide explicit and sharp identifiability conditions for an HLAM. On the other hand, in the cognitive diagnostic modeling literature, researchers (Köhn and Chiu, 2019; Cai et al., 2018) recently studied the “completeness” of the **Q**-matrix, a relevant concept to be revisited in Section 3, under attribute hierarchy. But these results can not ensure identifying uniquely the model parameters that determine the probabilistic HLAM. In summary, establishing identifiability without assuming any knowledge of the **Q**-matrix and the attribute hierarchy still remains unaddressed in the literature, and it is indeed a technically challenging task.

This paper addresses this identifiability question for popular HLAMs under an arbitrary

attribute hierarchy. We develop explicit sufficient conditions for identifying the attribute hierarchy, the \mathbf{Q} -matrix, and all the model parameters in an HLAM. These sufficient conditions become also necessary when the latent pattern space is saturated with no hierarchy. While for cases where there is a nonempty hierarchy, we discuss the necessity of these individual conditions and relax them in several nontrivial and interesting ways. Based on these and going further, we then establish the fully general necessary and sufficient identifiability conditions for the attribute hierarchy and all the model parameters under a fixed \mathbf{Q} -matrix. Our results in this regard sharply characterize the different roles played by different types of attributes in the attribute hierarchy graph. The theoretical developments can be used to assess the validity of an estimated HLAM obtained from any estimation method. They also provide insights into designing useful diagnostic tests under attribute hierarchy with minimal restrictions.

The rest of the paper is organized as follows. In Section 2, we introduce the model setup of the HLAMs. In Section 3, we present sufficient conditions on identifiability of \mathbf{Q} , attribute hierarchy, and model parameters. In Section 4, to thoroughly investigate how to close the gap between the necessity and sufficiency of the identifiability conditions, we focus on the case where \mathbf{Q} is fixed and derive the fully general necessary and sufficient conditions for identifying the attribute hierarchy and model parameters. In Section 5, we provide an extension of the identifiability result to other types of HLAMs that have potentially more parameters than that studied in Sections 3-4. We give a brief discussion in Section 6. All the technical proofs are presented in the Supplementary Material.

2. Model Setup and Examples

This section introduces the model setup of HLAMs. We first introduce some notation. For an integer m , denote $[m] = \{1, 2, \dots, m\}$. For a set \mathcal{A} , denote its cardinality by $|\mathcal{A}|$. Denote the $K \times K$ identity matrix by I_K and the K -dimensional all-one and all-zero vectors by $\mathbf{1}_K$ and $\mathbf{0}_K$, respectively.

An HLAM consists of two types of subject-specific binary variables, the observed responses $\mathbf{r} = (r_1, \dots, r_J) \in \{0, 1\}^J$ to J *items*; and the latent attribute pattern $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K) \in \{0, 1\}^K$, with α_k indicating the mastery or deficiency of the k th *attribute*. In this work, K is assumed known and fixed. This assumption is well suited for the motivating applications in cognitive diagnosis, where the number and also the real-world meanings of the latent attributes are usually known in the context of the application, and it is of interest to identify and learn other quantities from data. Next, we first describe the distribution of the latent attributes. Attribute k is said to be the prerequisite of attribute ℓ and denoted by $k \rightarrow \ell$, if any pattern $\boldsymbol{\alpha}$ with $\alpha_k = 0$ and $\alpha_\ell = 1$ is “forbidden” to exist. This is a common assumption in applications such as cognitive diagnosis to model subjects’ learning process (Leighton et al., 2004; Templin and Bradshaw, 2014). A subject’s latent pattern $\boldsymbol{\alpha}$ is assumed to follow a categorical distribution of population proportion parameters $\mathbf{p} = (p_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \{0, 1\}^K)$, with $p_{\boldsymbol{\alpha}} \geq 0$ and $\sum_{\boldsymbol{\alpha}} p_{\boldsymbol{\alpha}} = 1$. In particular, any pattern $\boldsymbol{\alpha}$ not respecting the hierarchy is deemed impossible to exist with population proportion $p_{\boldsymbol{\alpha}} = 0$. An attribute hierarchy is a set of prerequisite relations among the K attributes:

$$\mathcal{E} = \{k \rightarrow \ell : \text{attribute } k \text{ is a prerequisite for } \ell\}.$$

Generally, an attribute hierarchy \mathcal{E} implies a directed acyclic graph among the K attributes with no directed cycles; this graph constrains which attribute patterns are permissible or

forbidden. Specifically, any \mathcal{E} would induce a set of allowable configurations of attribute patterns out of $\{0, 1\}^K$, which we denote by $\mathcal{A}(\mathcal{E})$, or simply \mathcal{A} when it causes no confusion. For an arbitrary \mathcal{E} , the all-zero and all-one attribute patterns $\mathbf{0}_K$ and $\mathbf{1}_K$ always belong to the induced \mathcal{A} . This is because any prerequisite relation among attributes would not rule out the existence of the pattern possessing no attributes or the pattern possessing all attributes. When there is no attribute hierarchy among the K attributes, $\mathcal{E} = \emptyset$ and $\mathcal{A} = \{0, 1\}^K$. The set \mathcal{A} is a proper subset of $\{0, 1\}^K$ if $\mathcal{E} \neq \emptyset$. An attribute hierarchy determines the sparsity pattern of the vector of proportion parameters \mathbf{p} , because $p_{\alpha} > 0$ if and only if $\alpha \in \mathcal{A}(\mathcal{E})$, that is, if and only if α is permissible under \mathcal{E} . In this sense, a nonempty attribute hierarchy necessarily leads to degenerate parameter space for \mathbf{p} , as certain entries of \mathbf{p} will be constrained to zero.

In the practice of studying the attribute hierarchy in cognitive diagnosis, the case of $k \rightarrow \ell$ and $\ell \rightarrow k$ would indicate the two skill attributes α_k and α_ℓ are prerequisites for each other, which is not interpretable and hence is not used in modeling. Similarly, the case of having any cycle in the attribute hierarchy graph in the form of $k_1 \rightarrow k_2 \rightarrow \dots \rightarrow k_m \rightarrow k_1$ is also not interpretable. Therefore, a directed acyclic graph (DAG) structure among the latent attributes is well suited to describe the hierarchical nature of attributes that carry these substantive meanings. We emphasize here that the DAG of attribute hierarchy in an HLAM has a different nature from that in a Bayesian network (Pearl, 1986). This is because the DAG of attribute hierarchy encodes hard constraints on what variable patterns are permissible/forbidden, while the DAG in a Bayesian network encodes the conditional independence relations among the variables.

Remark 1. *Our attribute hierarchy constraints that “ $k \rightarrow \ell$ implies $\alpha_k = 0$ and $\alpha_\ell = 1$*

is impossible” have interesting connections to some other constraints in the statistics literature. In variable selection where the main effects of variables and their interaction effects may be present, the effect heredity principle (Hamada and Wu, 1992) posits that only if the main effects of variables exist will their corresponding interaction effect potentially exist. In particular, with θ_i and θ_j denoting the continuous regression coefficients associated with two heredity terms, Yuan et al. (2009) used a linear inequality $\theta_i \leq \theta_j$ (continuous relaxation of the hard constraints on the binary indicators of variable inclusions) to cleverly enforce the heredity constraint and facilitate computation. In causal inference, the monotonicity constraint in instrumental variable analyses (Hernán and Robins, 2006; Swanson et al., 2015) posits that if the instrumental variable satisfies $z_1 < z_2$ then the counterfactual treatment is a non-decreasing function of the instrument, i.e., $X_i^{z_1} \leq X_i^{z_2}$ for all subjects i . A key difference of the attribute hierarchy constraints from the aforementioned constraints is the involvement of many latent variables in HLAMs; indeed, all the $\alpha_1, \dots, \alpha_K$ among which the hierarchical constraints exist are latent. The binary patterns α that respect the attribute hierarchy \mathcal{E} follow an unknown categorical distribution with parameters $\mathbf{p} = (p_\alpha)$ with $\sum_\alpha p_\alpha = 1$, and the observed data distribution is obtained upon marginalizing out the latent structure and is quite complicated as a result.

Example 1. Fig 1 presents several hierarchies with the size of the associated \mathcal{A} , where a dotted arrow from α_k to α_ℓ indicates $k \rightarrow \ell$ and k is a direct prerequisite for ℓ . Note that under the hierarchy in Fig 1(a), the prerequisite $1 \rightarrow 3$ is an indirect prerequisite implied by $1 \rightarrow 2$ (or 4) and 2 (or 4) $\rightarrow 3$.

On top of the model of the latent attributes, an HLAM uses a $J \times K$ binary matrix $\mathbf{Q} = (q_{j,k})$ to encode the structural relationship between the J observed response variables

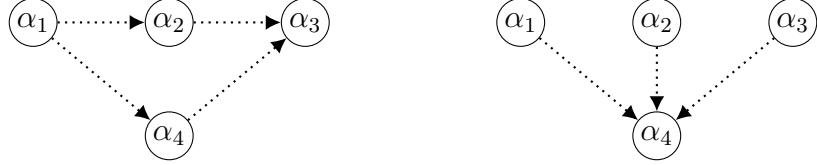


Figure 1: Different attribute hierarchies among binary attributes for $K = 4$ where $|\{0, 1\}^4| = 16$. For example, the set of allowed attribute patterns under hierarchy (a) is $\mathcal{A}_1 = \{\mathbf{0}_4, (1000), (1100), (1001), (1101), \mathbf{1}_4\}$.

$$\mathbf{Q}_{6 \times 3} := \begin{pmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \\ \mathbf{q}_3 \\ \mathbf{q}_4 \\ \mathbf{q}_5 \\ \mathbf{q}_6 \end{pmatrix} := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix};$$

(b) graphical model & attribute hierarchy

$\alpha \in \{0, 1\}^3$

$\mathbf{r} \in \{0, 1\}^6$

Figure 2: A binary structural matrix and the corresponding graphical model with (solid) directed edges from the latent to the observed variables representing dependencies. Below the observed variables in (b) are the row vectors of $\mathbf{Q}_{6 \times 3}$, i.e., the item loading vectors. The dotted arrows indicate the attribute hierarchy with $\mathcal{E} = \{1 \rightarrow 2, 1 \rightarrow 3\}$ and $\mathcal{A} = \{\mathbf{0}_3, (100), (110), (101), \mathbf{1}_3\}$.

and the K latent attributes. In cognitive diagnostic assessments, the matrix \mathbf{Q} is often specified by domain experts to summarize which abilities each test item targets on (Tatsuoka, 1990; von Davier, 2008; Rupp et al., 2010; de la Torre, 2011). Specifically, $q_{j,k} = 1$ if and only if the response r_j to the j th item has statistical dependence on latent variable α_k . The distribution of r_j , i.e., $\theta_{j,\alpha} := \mathbb{P}(r_j = 1 \mid \alpha)$, only depends on its “parent” latent attributes α_k ’s that are connected to r_j , i.e., $\{\alpha_k : q_{j,k} = 1\}$. The structural matrix \mathbf{Q} naturally induces a bipartite graph connecting the latent and the observed variables, with edges corresponding to entries of “1” in $\mathbf{Q} = (q_{j,k})$. Fig 2 presents an example of a structural matrix \mathbf{Q} and its corresponding directed graphical model between the $K = 3$ latent attributes and $J = 6$ observed variables. The solid edges from the latent attributes to the observed variables are specified by $\mathbf{Q}_{6 \times 3}$. As also can be seen from the graphical model, the observed responses to

the J items are conditionally independent given the latent attribute pattern α .

In the psychometrics literature, various HLAMs adopting the \mathbf{Q} -matrix concept have been proposed with the goal of diagnosing targeted attributes (Junker and Sijtsma, 2001; Templin and Henson, 2006; von Davier, 2008; Henson et al., 2009; de la Torre, 2011). They are often called the cognitive diagnostic models. The general family of latent attribute models are also widely used in other scientific areas including psychiatric evaluation (Templin and Henson, 2006; de la Torre et al., 2018) with the goal of diagnosing patients' various mental disorders, and epidemiological diagnosis of disease etiology (Wu et al., 2016, 2017; O'Brien et al., 2019). These applications share the common key interest in identifying the multivariate discrete latent attributes.

In this work, we mainly focus on a popular and fundamental type of modeling assumptions under such a framework; as to be revealed soon, this modeling assumption also has close connections to Boolean matrix factorization (Ravanbakhsh et al., 2016; Rukat et al., 2017). Specifically, we mainly consider the HLAMs that assume a logical *ideal response* $\Gamma_{\mathbf{q}_j, \alpha}$ given an attribute pattern α and an item loading vector \mathbf{q}_j in the noiseless case. Then item-level noise parameters are further introduced to account for uncertainty of observations. The following are two popular ways to define the ideal response.

The first is the Deterministic Input Noisy output “And” gate (DINA) model (Junker and Sijtsma, 2001; de la Torre and Douglas, 2004; von Davier, 2014). The DINA model assumes a conjunctive relationship among the attributes. The ideal response of attribute pattern α to item j is

$$(DINA \text{ ideal response}) \quad \Gamma_{\mathbf{q}_j, \alpha}^{\text{AND}} = \prod_{k=1}^K \alpha_k^{q_{j,k}}, \quad (2.1)$$

where the convention $0^0 \equiv 1$ is adopted. It is not hard to check that the above definition is equivalent to

$$\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}^{\text{AND}} = \mathbb{1}(\alpha_k \geq q_{j,k} \text{ for all the } k \in [K]). \quad (2.2)$$

The above equivalent definition intuitively and explicitly explains that the DINA adopts the *conjunctive* modeling assumption, because only if a subject with attribute pattern $\boldsymbol{\alpha}$ possesses all of the attributes required by the loading vector \mathbf{q}_j would he/she be considered as capable of this item j and have $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} = 1$. Such a conjunctive relationship is often assumed for diagnosis of students' mastery or deficiency of skill attributes in educational assessments, and $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}$ naturally indicates whether a student with $\boldsymbol{\alpha}$ has mastered all the attributes required by the test item j . With $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}$ in (2.1), the uncertainty of the responses is further modeled by the item-specific Bernoulli parameters

$$\theta_j^+ = \mathbb{P}(r_j = 1 \mid \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} = 1), \quad \theta_j^- = \mathbb{P}(r_j = 1 \mid \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} = 0), \quad (2.3)$$

where $\theta_j^+ > \theta_j^-$ is assumed for identifiability. For each item j , the ideal response $\Gamma_{\mathbf{q}_j, \cdot}$, if viewed as a function of attribute patterns, divides the patterns into two latent classes $\{\boldsymbol{\alpha} : \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} = 1\}$ and $\{\boldsymbol{\alpha} : \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} = 0\}$; and for these two latent classes, respectively, the item parameters quantify the noise levels of the response to item j that deviates from the ideal response. Note that the $\theta_{j, \boldsymbol{\alpha}}$ equals either θ_j^+ or θ_j^- , depending on the ideal response $\Gamma_{j, \boldsymbol{\alpha}}$. Denote the item parameter vectors by $\boldsymbol{\theta}^+ = (\theta_1^+, \dots, \theta_J^+)^{\top}$ and $\boldsymbol{\theta}^- = (\theta_1^-, \dots, \theta_J^-)^{\top}$.

The second model is the Deterministic Input Noisy output “Or” gate (DINO) model

(Templin and Henson, 2006). The DINO model assumes the following ideal response

$$(\text{DINO ideal response}) \quad \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}^{\text{OR}} = I(q_{j,k} = \alpha_k = 1 \text{ for at least one } k). \quad (2.4)$$

Such a disjunctive relationship is often assumed in psychiatric measurement of mental disorders (Templin and Henson, 2006; de la Torre et al., 2018). With $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}$ in (2.4), the uncertainty of the responses is modeled by the item-specific parameters as defined in (2.3). In the Boolean matrix factorization literature, a similar model was proposed (Ravanbakhsh et al., 2016). Adapted to the terminology here, Rukat et al. (2017) assumes the ideal response takes the form

$$(\text{equivalent to (2.4)}) \quad \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}^{\text{OR}} = 1 - \prod_{k=1}^K (1 - \alpha_k q_{j,k}), \quad (2.5)$$

which is equivalent to the definition in (2.4), while the model in Rukat et al. (2017) constrains all the item-level noise parameters to be the same.

We next first focus on the asymmetric DINA-based HLAMs, as they are very popular and fundamental models widely used in the motivating applications of educational cognitive diagnosis. We also study the identifiability of DINO-based HLAMs and another type of HLAMs in Section 5. For notational simplicity, we next write $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}^{\text{AND}}$ simply as $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}$. Denote by $\Gamma(\mathbf{Q}, \mathcal{E})$ the $J \times |\mathcal{A}(\mathcal{E})|$ ideal response matrix with the $(j, \boldsymbol{\alpha})$ th entry being $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}$ for $\boldsymbol{\alpha} \in \mathcal{A}(\mathcal{E})$. Under the introduced setup of DINA-based HLAMs, the probability mass function of the J -dimensional random response vector \mathbf{R} takes the form of

$$P(\mathbf{R} = \mathbf{r} \mid \mathbf{Q}, \mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}(\mathcal{E})} p_{\boldsymbol{\alpha}} \prod_{j=1}^J [\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} \theta_j^+ + (1 - \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}) \theta_j^-]^{r_j}$$

$$\times [1 - \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} \theta_j^+ - (1 - \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}) \theta_j^-]^{1-r_j},$$

where $\mathbf{r} \in \{0, 1\}^J$ is an arbitrary response pattern.

3. Identifiability of \mathbf{Q} , Attribute Hierarchy, and Model Parameters: Establishing Sufficiency

This section presents one main result on the sufficient conditions for identifiability of \mathbf{Q} , \mathcal{E} , and model parameters $\boldsymbol{\theta}^+$, $\boldsymbol{\theta}^-$, and \mathbf{p} . Following the definition of identifiability in the statistics literature, we say that $(\mathbf{Q}, \mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$ of an HLAM are identifiable if for any $(\bar{\mathbf{Q}}, \bar{\mathcal{E}}, \bar{\boldsymbol{\theta}}^+, \bar{\boldsymbol{\theta}}^-, \bar{\mathbf{p}})$ in the parameter space constrained by \mathbf{Q} and \mathcal{E} , there exist no $(\bar{\mathbf{Q}}, \bar{\mathcal{E}}, \bar{\boldsymbol{\theta}}^+, \bar{\boldsymbol{\theta}}^-, \bar{\mathbf{p}}) \neq (\mathbf{Q}, \mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$ such that

$$\mathbb{P}(\mathbf{R} = \mathbf{r} \mid \bar{\mathbf{Q}}, \bar{\mathcal{E}}, \bar{\boldsymbol{\theta}}^+, \bar{\boldsymbol{\theta}}^-, \bar{\mathbf{p}}) = \mathbb{P}(\mathbf{R} = \mathbf{r} \mid \mathbf{Q}, \mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p}), \quad \forall \mathbf{r} \in \{0, 1\}^J. \quad (3.6)$$

We point out that in the above definition of identifiability, the alternative vector of proportion parameters $\bar{\mathbf{p}}$ is not constrained to have support on $\mathcal{A}(\mathcal{E})$. Instead, the vector $\bar{\mathbf{p}}$ should be allowed to have an arbitrary support $\bar{\mathcal{A}}$ potentially resulting from an arbitrary $\bar{\mathcal{E}}$; the goal of establishing identifiability is indeed to develop conditions to ensure that as long as (3.6) holds, one must have $\bar{\mathbf{p}} = \mathbf{p}$ and $\bar{\mathcal{E}} = \mathcal{E}$ from the equations in (3.6).

We further introduce some notation and important concepts. Since an attribute hierarchy is a directed acyclic graph, the K attributes $\{1, 2, \dots, K\}$ can be arranged in a topological order such that the prerequisite relation “ \rightarrow ” only happens in one direction; in other words, we can assume without loss of generality that $k \rightarrow \ell$ only if $k < \ell$. Define the following *reachability matrix* \mathbf{E} among the K attributes under the attribute hierarchy. The

$\mathbf{E} = (e_{k,\ell})$ is a $K \times K$ binary matrix, where $e_{k,k} = 1$ for all $k \in [K]$ and $e_{\ell,k} = 1$ if attribute k is a direct or indirect prerequisite for attribute ℓ . In cognitive diagnosis, the concept of the reachability matrix was first considered in Tatsuoka (1986) to represent the direct and indirect relationships between attributes. It is not hard to see that if the attributes $1, 2, \dots, K$ are in a topological order described earlier, the reachability matrix \mathbf{E} is a lower-triangular matrix with all the diagonal entries being one.

Under DINA-based HLAMs, any non-empty attribute hierarchy \mathcal{E} defines an equivalence relation on the set of all the \mathbf{Q} -matrices. To see this, recall $\Gamma(\mathbf{Q}, \mathcal{E})$ denotes the $J \times |\mathcal{A}(\mathcal{E})|$ ideal response matrix. If $\Gamma(\mathbf{Q}_1, \mathcal{E}) = \Gamma(\mathbf{Q}_2, \mathcal{E})$, then \mathbf{Q}_1 and \mathbf{Q}_2 are said to be in the same \mathcal{E} -induced equivalence class and we denote this by $\mathbf{Q}_1 \stackrel{\mathcal{E}}{\sim} \mathbf{Q}_2$. To interpret, if under a certain hierarchy \mathcal{E} , two different \mathbf{Q} -matrices lead to identical ideal responses for all the permissible latent patterns in $\mathcal{A}(\mathcal{E})$, then these two \mathbf{Q} -matrices are indistinguishable based on the response data; therefore they should be treated as equivalent. The following example illustrates how an attribute hierarchy determines a set of equivalent \mathbf{Q} -matrices.

Example 2. Consider the attribute hierarchy $\mathcal{E} = \{1 \rightarrow 2, 1 \rightarrow 3\}$ in Fig 2, which results in $\mathcal{A}(\mathcal{E}) = \{\mathbf{0}_3, (100), (110), (101), \mathbf{1}_3\}$. The identity matrix I_3 is equivalent to the reachability matrix \mathbf{E} under \mathcal{E} and

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \stackrel{\mathcal{E}}{\sim} \mathbf{E} = \begin{pmatrix} 1 & 0 & 0 \\ \textcolor{blue}{1} & 1 & 0 \\ \textcolor{blue}{1} & 0 & 1 \end{pmatrix} \stackrel{\mathcal{E}}{\sim} \begin{pmatrix} 1 & 0 & 0 \\ * & 1 & 0 \\ * & 0 & 1 \end{pmatrix}, \quad (3.7)$$

where the “*’s in the third matrix above indicate unspecified values, any of which can be either 0 or 1. This equivalence is due to that attribute α_1 serves as the prerequisite for both α_2 and α_3 , and any item loading vector \mathbf{q}_j measuring α_2 or α_3 is equivalent to a modified one that also measures α_1 , in terms of classifying the patterns in \mathcal{A} into two categories $\{\boldsymbol{\alpha} : \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} = 1\}$ and $\{\boldsymbol{\alpha} : \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} = 0\}$. Note that any \mathbf{Q} -matrix equivalent to

I_K under the $\mathcal{E} = \{1 \rightarrow 2, 1 \rightarrow 3\}$ must take the form of the third \mathbf{Q} -matrix in (3.7). Under a DINA-based HLAM, if the true \mathbf{Q} -matrix \mathbf{Q}^{true} is not known, then any other \mathbf{Q} with $\mathbf{Q} \stackrel{\mathcal{E}}{\sim} \mathbf{Q}^{\text{true}}$ can not be distinguished from \mathbf{Q}^{true} based on the observations, even if the continuous parameters $(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$ are all known. This is because the ideal response matrix $\Gamma(\mathbf{Q}, \mathcal{E})$ is the key latent structure underlying a DINA-based HLAM, and that if $\mathbf{Q} \stackrel{\mathcal{E}}{\sim} \mathbf{Q}^{\text{true}}$ (equivalently, $\Gamma(\mathbf{Q}, \mathcal{E}) = \Gamma(\mathbf{Q}^{\text{true}}, \mathcal{E})$), then \mathbf{Q} and \mathbf{Q}^{true} are inherently not distinguishable.

Given any hierarchy \mathcal{E} , the equivalence $I_K \stackrel{\mathcal{E}}{\sim} \mathbf{E}$ is always true, for which Eq. (3.7) in Example 2 is an example. Before presenting the theorem on sufficient conditions for identifiability, we introduce two useful operations on a \mathbf{Q} -matrix given an attribute hierarchy \mathcal{E} : the “densifying” operation $\mathcal{D}^{\mathcal{E}}(\cdot)$ and the “sparsifying” operation $\mathcal{S}^{\mathcal{E}}(\cdot)$, as follows.

Definition 1. *Given an attribute hierarchy \mathcal{E} and a matrix \mathbf{Q} , do the following: for any $q_{j,h} = 1$ and $k \rightarrow h$, set $q_{j,k}$ to “1” and obtain a modified matrix $\mathcal{D}^{\mathcal{E}}(\mathbf{Q})$. This $\mathcal{D}^{\mathcal{E}}(\mathbf{Q})$ is said to be the “densified” version of \mathbf{Q} .*

Definition 2. *Given an attribute hierarchy \mathcal{E} and a matrix \mathbf{Q} , do the following: for any $q_{j,h} = 1$ and $k \rightarrow h$, set $q_{j,k}$ to “0” and obtain a modified matrix $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$. This $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$ is said to be the “sparsified” version of \mathbf{Q} .*

Under the above two definitions, given any attribute hierarchy \mathcal{E} , the two statements $\mathcal{D}^{\mathcal{E}}(I_K) = \mathbf{E}$ and $\mathcal{S}^{\mathcal{E}}(\mathbf{E}) = I_K$ always hold. Specifically, $\mathcal{D}^{\mathcal{E}}(I_K) = \mathbf{E}$ means that in the special case where $J = K$ and \mathbf{Q} takes the form of an identity matrix I_K , densifying such a \mathbf{Q} always gives a $K \times K$ reachability matrix \mathbf{E} under the hierarchy \mathcal{E} . Similarly, $\mathcal{S}^{\mathcal{E}}(\mathbf{E}) = I_K$ means that in another special case where $J = K$ and \mathbf{Q} takes the form of the reachability matrix \mathbf{E} , sparsifying such a \mathbf{Q} always gives the identity matrix I_K . These two special examples illustrate the definitions of the sparsifying/densifying operations on \mathbf{Q} -matrices

and the relationship between \mathbf{E} and I_K . In cognitive diagnosis, the densified \mathbf{Q} -matrix with all the row vectors respecting the attribute hierarchy \mathcal{E} is also said to satisfy the “restricted \mathbf{Q} -matrix design” (e.g., Cai et al., 2018; Tu et al., 2019); for such \mathbf{Q} , it holds that $\mathbf{Q} = \mathcal{D}^{\mathcal{E}}(\mathbf{Q})$. It is worth pointing out that either the sparsifying or the densifying operation modifies \mathbf{Q} only within a same equivalence class. Indeed, $\mathcal{D}^{\mathcal{E}}(\mathbf{Q})$ denotes the densest \mathbf{Q} with the largest number of “1”s in the equivalence class, while $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$ denotes the sparsest \mathbf{Q} with the largest number of “0”s in the equivalence class. In the special case with an empty attribute hierarchy, each equivalence class of \mathbf{Q} contains only one element which is \mathbf{Q} itself, so $\mathbf{Q} = \mathcal{D}^{\mathcal{E}}(\mathbf{Q}) = \mathcal{S}^{\mathcal{E}}(\mathbf{Q})$ for $\mathcal{E} = \emptyset$. As will be revealed in the following theorem, our identifiability conditions are essentially requirements on the equivalence class of \mathbf{Q} described using the densifying and sparsifying operations.

Theorem 1. *Consider an HLAM under the DINA model an attribute hierarchy \mathcal{E} . Then $(\Gamma(\mathbf{Q}, \mathcal{E}), \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$ are jointly identifiable if the true \mathbf{Q} satisfies the following conditions.*

A. The \mathbf{Q} contains $K \times K$ submatrix \mathbf{Q}^0 that is equivalent to the identity matrix I_K under the hierarchy \mathcal{E} .

(Without loss of generality, assume the first K rows of \mathbf{Q} form \mathbf{Q}^0 , and denote the remaining submatrix of \mathbf{Q} by \mathbf{Q}^ .)*

B. The $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$, sparsified version of \mathbf{Q} , has at least three entries of “1”s in each column.

C. The $\mathcal{D}^{\mathcal{E}}(\mathbf{Q}^)$, densified version of the submatrix \mathbf{Q}^* , contains K distinct column vectors.*

Furthermore, Conditions A, B and C are necessary and sufficient when there exists no hierarchy with $p_{\alpha} > 0$ for all $\alpha \in \{0, 1\}^K$.

We make several remarks on the relationship between the new theory and existing works.

Remark 2. In the cognitive diagnostic modeling literature, a \mathbf{Q} -matrix is said to be “complete” if it can distinguish all the 2^K latent attribute profiles (Chiu et al., 2009). When the latent pattern space \mathcal{A} is saturated with $\mathcal{A} = \{0, 1\}^K$, the completeness of \mathbf{Q} is a natural necessary requirement for identifiability. When $\mathcal{A} = \{0, 1\}^K$, the \mathbf{Q} -matrix is complete if it contains all the K distinct standard basis vectors as row vectors, that is, \mathbf{Q} contains an I_K . When there exists a certain attribute hierarchy \mathcal{E} leading to some $\mathcal{A} \subsetneq \{0, 1\}^K$, the requirement for the “completeness” of \mathbf{Q} will change. Recently, Köhn and Chiu (2019) and Cai et al. (2018) studied conditions for the completeness of \mathbf{Q} under the attribute hierarchy. But these conditions can not ensure the entire probabilistic model structure involving \mathbf{Q} , \mathcal{E} , and parameters \mathbf{p} , $\boldsymbol{\theta}^+$ and $\boldsymbol{\theta}^-$ are identifiable and estimable from data. To our knowledge, Theorem 1 establishes the first identifiability result under the attribute hierarchy in the literature. Condition A in Theorem 1 is equivalent to requiring $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$ contains an I_K . Therefore, Conditions A and B combined are equivalent to the following statement about $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$: the $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$ contains an I_K and each column of it has at least three entries of “1”s.

Remark 3. As stated in the last part of Theorem 1, when there is no attribute hierarchy with $\mathcal{E} = \emptyset$, Conditions A, B, and C become necessary and sufficient for the identifiability of both \mathbf{Q} and $(\mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$. In such a special case with $\mathcal{E} = \emptyset$, Gu and Xu (2021) established the necessary and sufficient identifiability conditions termed as “*completeness*” that requires the true \mathbf{Q} to contain an identity submatrix I_K , “*repeated-measurement*” that requires \mathbf{Q} to have at least three entries of “1” in each column, and “*distinctiveness*” requiring that in addition to containing an I_K , the \mathbf{Q} should contain distinct column vectors in the remaining submatrix; we denote these three requirements by Conditions A^0 , B^0 , and C^0 , respectively. Our current conditions A, B, and C in Theorem 1 can be thought of as “ \mathcal{E} -

“completeness”, “ \mathcal{E} -repeated-measurement”, “ \mathcal{E} -distinctiveness” given an attribute hierarchy \mathcal{E} . When $\mathcal{E} = \emptyset$, the $\mathcal{S}^{\mathcal{E}}(\mathbf{Q}) = \mathcal{D}^{\mathcal{E}}(\mathbf{Q}) = \mathbf{Q}$ holds; as a result, Condition A exactly becomes requiring \mathbf{Q} itself to contain a submatrix I_K ; similarly, Conditions B and C exactly reduce to the conditions B^0 and C^0 on \mathbf{Q} itself. Indeed, in such cases with $\mathcal{E} = \emptyset$, the current conditions of “ \mathcal{E} -completeness”, “ \mathcal{E} -repeated-measurement”, “ \mathcal{E} -distinctiveness” just reduce to the “completeness”, “repeated-measurement”, “distinctiveness” conditions proposed in Gu and Xu (2021). Establishing identifiability under an arbitrary attribute hierarchy \mathcal{E} as done in Theorem 1 is technically much more challenging than the existing result for $\mathcal{E} = \emptyset$. Moreover, in the later Section 4, we will thoroughly study that under a fixed \mathbf{Q} -matrix, how the necessity of the identifiability conditions changes when there is a nonempty hierarchy.

Theorem 1 ensures the discrete ideal response structure $\Gamma(\mathbf{Q}, \mathcal{E})$ and all the associated model parameters $(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$ are identifiable. The following proposition complements this conclusion and further establishes identifiability of \mathcal{E} and \mathbf{Q} based on Theorem 1.

Proposition 1. *Consider a DINA-based HLAM. In addition to Conditions A–C in Theorem 1, if the true \mathbf{Q} is known to contain an I_K , then $(\mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$ are identifiable. On the other hand, it is indeed necessary for \mathbf{Q} to contain an I_K to ensure an arbitrary \mathcal{E} is identifiable.*

Proposition 2. *Consider a DINA-based HLAM. If Conditions A–C in Theorem 1 are satisfied and the true \mathbf{Q} is known in part to contain a submatrix I_K for certain K items, then the equivalence class of \mathbf{Q} defined by the attribute hierarchy \mathcal{E} is identifiable. That is, the specific \mathbf{Q} is not strictly identifiable within its equivalence class under any $\mathcal{E} \neq \emptyset$, but the densified $\mathcal{D}^{\mathcal{E}}(\mathbf{Q})$ and the sparsified $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$ are identifiable.*

The statement in Proposition 2 that \mathbf{Q} is identifiable only up to its equivalence class is inherent to all the DINA- or DINO-type HLAMs and it is an inevitable consequence of any

nonempty attribute hierarchy $\mathcal{E} \neq \emptyset$; see Example 2. But this statement will not undermine the efficacy of the identifiability conclusion, because $\mathcal{D}^{\mathcal{E}}(\mathbf{Q})$ and $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$ themselves are still identifiable and provide practical interpretability of the structural matrix. We next present a toy example illustrating how to apply Theorem 1 to check identifiability.

Example 3. Consider the attribute hierarchy $\{\alpha_1 \rightarrow \alpha_2, \alpha_1 \rightarrow \alpha_3\}$ among $K = 3$ attributes as in Fig 2. The following 8×3 structural matrix \mathbf{Q} satisfies Conditions A, B and C in Theorem 1. In particular, the first 3 rows of \mathbf{Q} serve as \mathbf{Q}^0 in Condition A, and the last 5 rows serve as \mathbf{Q}^* . In the following display, the matrix entries modified by the sparsifying operation in Condition B and the densifying operation in Condition C are highlighted. The resulting $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$ and $\mathcal{D}^{\mathcal{E}}(\mathbf{Q})$ satisfy the requirements in Conditions B and C. So the HLAM associated with \mathbf{Q} is identifiable.

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}^0 \\ \mathbf{Q}^* \end{pmatrix} = \begin{pmatrix} I_3 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \xrightarrow{\text{Sparsify}} \mathcal{S}^{\mathcal{E}}(\mathbf{Q}) = \begin{pmatrix} I_3 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ \textcolor{orange}{0} & 1 & 0 \\ 0 & 0 & 1 \\ \textcolor{orange}{0} & 1 & 1 \end{pmatrix}; \quad (3.8)$$

$$\xrightarrow{\text{Densify}} \mathcal{D}^{\mathcal{E}}(\mathbf{Q}) = \begin{pmatrix} \mathbf{E} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ \textcolor{blue}{1} & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \quad (3.9)$$

When estimating an HLAM with the goal of recovering the ideal response structure $\Gamma(\mathbf{Q}, \mathcal{E})$ and continuous parameters, Theorem 1 guarantees Conditions A, B and C suffice and are close to being necessary. If the goal is to uniquely determine the attribute hierarchy from the identified $\Gamma(\mathbf{Q}, \mathcal{E})$, the additional condition that \mathbf{Q} contains an I_K becomes necessary. This phenomenon can be better understood if related to the identification of the factor

loading matrix in factor analysis (Anderson, 2009; Bai and Li, 2012); the loading matrix there is often required to include an identity submatrix or satisfy certain rank constraints, since otherwise the loading matrix can not be identifiable due to rotational indeterminacy.

The existing results of identifiability for *non-hierarchical* latent attribute models (i.e., with an empty graph $\mathcal{E} = \emptyset$ in Gu and Xu, 2021) adopt a key assumption that $p_{\alpha} > 0$ for any possible binary pattern $\alpha \in \{0, 1\}^K$, and the proofs in Gu and Xu (2021) heavily relied on this assumption on \mathbf{p} to derive the identifiability conditions. Importantly, when the assumption that “ $p_{\alpha} > 0$ for any $\alpha \in \{0, 1\}^K$ ” fails to hold, the proof arguments in Gu and Xu (2021) also do not hold and hence the conclusions there cannot be simply modified. Rather, a careful analysis of the polynomial systems arising from the probability mass function (PMF) of the observed \mathbf{R} is required to derive suitable identifiability conditions.

We would like to point out that dealing with such a degenerate parameter space of \mathbf{p} under an attribute hierarchy \mathcal{E} requires quite delicate algebraic work. Specifically, our proof technique of identifiability is based on investigating under what conditions, the highly complex and \mathbf{Q} -matrix-constrained polynomial equations given by the PMF of the vector \mathbf{R} has unique roots; uniqueness of roots indicates identifiability of parameters. When using this proof technique, we start by inspecting polynomial equations $\mathbb{P}(\mathbf{R} = \mathbf{r} \mid \bar{\mathbf{Q}}, \bar{\mathcal{E}}, \bar{\boldsymbol{\theta}}^+, \bar{\boldsymbol{\theta}}^-, \bar{\mathbf{p}}) = \mathbb{P}(\mathbf{R} = \mathbf{r} \mid \mathbf{Q}, \mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$, $\forall \mathbf{r} \in \{0, 1\}^J$ for unknown true parameters $(\mathbf{Q}, \mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$ and arbitrary alternative parameters $(\bar{\mathbf{Q}}, \bar{\mathcal{E}}, \bar{\boldsymbol{\theta}}^+, \bar{\boldsymbol{\theta}}^-, \bar{\mathbf{p}})$, and investigate what conditions guarantee the alternative parameters are identical to the true ones. Under an unknown attribute hierarchy \mathcal{E} , certain true proportions p_{α} equal zero but we do not know which ones equal zero. Therefore complex constraints on polynomial equations will occur because certain terms vanish from the one hand side of the equation (corresponding to the true parameters

p_{α} 's) but do not vanish from the other hand side of the equation (corresponding to the unknown alternative parameters \bar{p}_{α} 's; we do not know which $\bar{p}_{\alpha} = 0$ out of all the possible $\alpha \in \{0, 1\}^K$). This fact makes the study of the identifiability issues in the current work considerably harder and quite different from existing results for non-hierarchical latent attribute models (e.g., Gu and Xu, 2021).

As stated in the end of Theorem 1, Conditions A, B, and C become not only sufficient but also necessary for identifiability when there is no hierarchy among attributes. Interestingly, the necessity of these conditions will subtly change when a nonempty attribute hierarchy comes into play. Our next section thoroughly investigates these aspects.

4. Identifiability of Attribute Hierarchy and Model Parameters: Pushing Towards Necessity

In order to close the gap between necessity and sufficiency, in this section we thoroughly investigate the necessity of the identifiability conditions for $(\mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$ under the assumption that \mathbf{Q} is known and fixed. In the following Subsection 4.1, we first investigate the necessity of the conditions proposed in Section 3 individually, to gain insight into how the necessity changes as the attribute hierarchy changes. Then in Subsection 4.2, we further establish the general necessary and sufficient conditions for identifying the attribute hierarchy and other parameters under an arbitrary hierarchy graph \mathcal{E} .

4.1 Investigating the Necessity of Conditions A, B, C Individually

Our first result establishes the necessity of Condition A in Theorem 1.

Proposition 3. *Consider a DINA-based HLAM. Condition A that the sparsified $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$*

contains an I_K is necessary for identifiability of $(\Gamma(\mathbf{Q}, \mathcal{E}), \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$.

Proposition 3 shows that Condition A can not be relaxed under any attribute hierarchy. On the other hand, Condition B and Condition C are more “local” in the sense that they regard individual attributes (equivalently, individual columns of the \mathbf{Q} -matrix). Interestingly, it turns out that the necessity of these two conditions highly depends on the role of each attribute in the attribute hierarchy graph. We next characterize the fine boundary between sufficiency and necessity of identifiability conditions for various types of attributes. Given any attribute hierarchy graph \mathcal{E} , we define the following four types of attributes.

Definition 3 (Singleton Attribute). *An attribute k is a “singleton attribute” if there **neither exists** any attribute h such that $k \rightarrow h$ **nor exists** any attribute ℓ such that $\ell \rightarrow k$.*

Definition 4 (Root Attribute). *An attribute k is an “**root attribute**” if there **exists** some attribute h such that $k \rightarrow h$ but **does not exist** any attribute ℓ such that $\ell \rightarrow k$.*

Definition 5 (Leaf Attribute). *An attribute k is a “leaf attribute” if there **exists** some attribute ℓ such that $\ell \rightarrow k$ but **does not exist** any attribute h such that $k \rightarrow h$.*

Definition 6 (Intermediate Attribute). *An attribute k is an “intermediate attribute” if there **exists** some attribute ℓ with $\ell \rightarrow k$ and also **exists** some attribute h with $k \rightarrow h$.*

The above four definitions together describe a full categorization of attributes given any attribute hierarchy. In other words, given any \mathcal{E} , an attribute is either a singleton, or a root, or a leaf, or an intermediate attribute. As a special case, when the attribute pattern space $\mathcal{A} = \{0, 1\}^K$ is saturated, all the K attributes are singleton attributes.

Example 4. Leighton *et al.* (Leighton et al., 2004) is among the first works that considered the attribute hierarchy method for the purpose of cognitive diagnosis. In particular, they

presented and named the four different types of hierarchies among $K = 6$ attributes, as shown in our Fig 3. In our terminology, in plot (a), attribute 1 is a root attribute, attribute 6 is a leaf attribute, and the remaining attributes 2, 3, 4, 5 are intermediate attributes; in plot (b), the roles of the six attributes are the same as those in plot (a); in plot (c), attribute 1 is a root attribute, attribute 2 and 3 are intermediate attributes, attributes 4, 5, 6 are leaf attributes; in plot (d), attribute 1 is a root, and the remaining 2, 3, 4, 5, 6 are leaves.

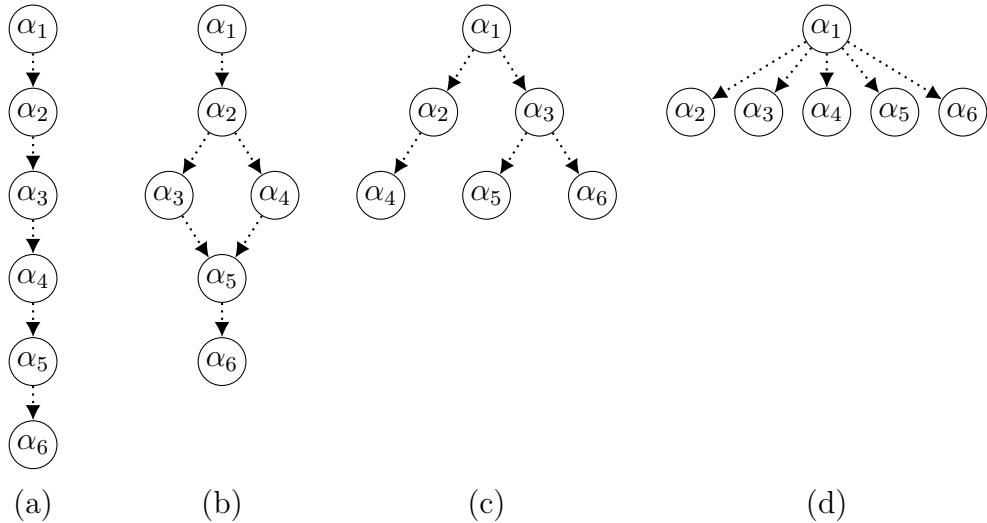


Figure 3: Four attribute hierarchies presented in Leighton et al. (2004), named as: (a) *linear*, (b) *convergent*, (c) *divergent*, and (d) *unstructured*. For example, in (b), α_1 is a root attribute, $\alpha_2, \dots, \alpha_6$ are intermediate attributes, α_7 is a leaf attribute, and there are no singleton attributes.

For ease of discussion, in the following conclusions regarding necessity of the identifiability conditions, we shall focus on the **Q**-matrices that satisfy the restricted **Q**-matrix design. Recall that a **Q**-matrix is said to satisfy the restricted **Q**-matrix design if each of its row vectors is a permissible attribute pattern under the hierarchy \mathcal{E} . In the literature of cognitive diagnostic modeling, the restricted **Q**-matrix design is shown empirically to be useful in improving clustering accuracy of diagnostic test takers (Tu et al., 2019). Our theoretical findings in the rest of this subsection reveal that in addition to the restricted **Q**-matrix

design, what other requirements are necessary to ensure identifiability.

Before presenting the next identifiability result, we first introduce a new notion of identifiability of the attribute hierarchy \mathcal{E} and proportion parameters \mathbf{p} ; that is, identifiability up to equivalence classes $[\mathcal{E}]$ and $[\mathbf{p}]$. Under an unknown nonempty hierarchy $\mathcal{E} \neq \emptyset$, if all row vectors of \mathbf{Q} respect the attribute hierarchy, then there exists a trivial nonidentifiability issue that can be resolved by introducing an equivalence relation, similar in spirit to that in Gu and Xu (2020). To see this, consider $K = 2$ and $\mathcal{E} = \{1 \rightarrow 2\}$, then a \mathbf{Q} -matrix $\mathbf{Q} = \mathbf{E} = (1, 0; 1, 1)$ has both rows respecting the attribute hierarchy. Further, consider the simplest special case without any item-level noise, $1 - \theta_1^+ = 1 - \theta_2^+ = \theta_1^- = \theta_2^- = 0$. Now if \mathcal{E} is unknown, then it is not hard to see that any alternative proportion parameters $\bar{\mathbf{p}}$ satisfying the following equations will be nondistinguishable from the true parameters \mathbf{p} :

$$p_{(00)} = \bar{p}_{(00)} + \bar{p}_{(01)}; \quad p_{(10)} = \bar{p}_{(10)}; \quad p_{(11)} = \bar{p}_{(11)}. \quad (4.10)$$

The above phenomenon is closely related to the \mathbf{p} -partial identifiability defined in Gu and Xu (2020), which means when \mathbf{Q} does not contain an identity submatrix I_K (often called “incomplete” in cognitive diagnosis models), the proportion parameters can at best be identified up to the equivalence classes induced by \mathbf{Q} . In the toy example in the last paragraph, the attribute patterns (00) and (01) are equivalent under the incomplete $\mathbf{Q} = (1, 0; 1, 1)$ because $\Gamma_{\mathbf{Q},(00)} = \Gamma_{\mathbf{Q},(01)}$; so $\bar{p}_{(00)}$ and $\bar{p}_{(01)}$ can be identified up to their sum at best, as illustrated in (4.10). Therefore, we will say $(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, [\mathcal{E}], [\mathbf{p}])$ are identifiable, if continuous parameters $(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-)$ are identifiable in the usual sense, and the \mathcal{E} and \mathbf{p} are identifiable up to the equivalence classes each of them belongs to; namely, the only nonidentifiability issue regarding \mathcal{E} (and hence \mathbf{p}) is due to the equivalence relation induced by \mathbf{Q} like the example in (4.10). We give more formal definitions of $[\mathcal{E}]$ and $[\mathbf{p}]$ as follows: given an attribute hierarchy \mathcal{E} and a \mathbf{Q} -matrix with all row vectors respecting the hierarchy \mathcal{E} (i.e., satisfying

the restricted \mathbf{Q} -matrix design), define the equivalence class of attribute hierarchies $[\mathcal{E}]$ and that of proportion parameters $[\mathbf{p}]$ as

$$[\mathcal{E}] = \{\bar{\mathcal{E}} : \Gamma(\mathbf{Q}, \mathcal{E}) = \Gamma(\mathbf{Q}, \bar{\mathcal{E}})\};$$

$$[\mathbf{p}] = \{\bar{\mathbf{p}} : \bar{\mathbf{p}} \text{ is associated with some } \bar{\mathcal{E}} \in [\mathcal{E}], \text{ with } (\boldsymbol{\theta}^+ = \mathbf{1}_{J \times 1}, \boldsymbol{\theta}^- = \mathbf{0}_{J \times 1}, \mathcal{E}, \mathbf{p}) \text{ and}$$

$$(\boldsymbol{\theta}^+ = \mathbf{1}_{J \times 1}, \boldsymbol{\theta}^- = \mathbf{0}_{J \times 1}, \bar{\mathcal{E}}, \bar{\mathbf{p}}) \text{ giving rise to the same distribution of } \mathbf{R}\}$$

$$= \{\bar{\mathbf{p}} = (\bar{p}_{\alpha}, \alpha \in \{0, 1\}^K) : \forall \alpha \text{ that respects the hierarchy } \mathcal{E}, \sum_{\alpha' : \Gamma_{\mathbf{Q}, \alpha'} = \Gamma_{\mathbf{Q}, \alpha}} \bar{p}_{\alpha'} = p_{\alpha}\}.$$

We point out that the above nonidentifiability of a specific \mathcal{E} within its equivalence class $[\mathcal{E}]$ is somewhat trivial and can be easily resolved, by simply defining the final $\bar{\mathcal{E}}^*$ to be the hierarchy with the most directed edges among all the possible hierarchies in the equivalence class $[\bar{\mathcal{E}}]$. It is easy to see that such $\bar{\mathcal{E}}^*$ equals the true \mathcal{E} in the toy example, because in order for $\bar{\mathcal{E}}$ to have the most directed edges, one needs to set $\bar{p}_{(01)} = 0$ under (4.10) and that exactly makes the resulting $\bar{\mathbf{p}} = \mathbf{p}$ and $\bar{\mathcal{E}}^* = \mathcal{E} = \{1 \rightarrow 2\}$. By a similar reasoning, this procedure also works more generally for any hierarchy \mathcal{E} . Therefore, when a fixed \mathbf{Q} -matrix has all rows respecting the hierarchy, it is still very meaningful and useful to study the identifiability of $(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, [\mathcal{E}], [\mathbf{p}])$ and to investigate the minimal identifiability conditions. Our results in this section will establish the necessary and sufficient identifiability conditions in this regard.

In the following Propositions 4–6, we show how Condition B can be generally relaxed, depending on whether the attribute is root, leaf, or intermediate.

Proposition 4 (Necessary Condition for Singleton Attribute). *Consider a DINA-based HLAM with a fixed \mathbf{Q} -matrix whose row vectors respect the hierarchy \mathcal{E} . The following hold for a singleton attribute k in any attribute hierarchy.*

(a) $\sum_{j=1}^J q_{j,k} \geq 3$ is necessary for the identifiability of $(\mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$.

(b) *There exists scenarios where the equality in part (a) is achieved with $\sum_{j=1}^J q_{j,k} = 3$ and the identifiability of $(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, [\mathcal{E}], [\mathbf{p}])$ is guaranteed.*

Proposition 5 (Necessary Condition for Root or Leaf Attribute). *Consider a DINA-based HLAM with a fixed \mathbf{Q} -matrix whose row vectors respect the hierarchy \mathcal{E} . Denote the (j, k) th entry of $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$ by $q_{j,k}^{\text{sparse}}$. The following conclusions hold for k if attribute k is either a root attribute or a leaf attribute.*

(a) *$\sum_{j=1}^J q_{j,k}^{\text{sparse}} \geq 2$ is necessary for the identifiability of $(\mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$.*

(b) *There exist scenarios where the equality in part (a) is achieved with $\sum_{j=1}^J q_{j,k}^{\text{sparse}} = 2$ and the identifiability of $(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, [\mathcal{E}], [\mathbf{p}])$ is guaranteed.*

Proposition 6 (Necessary Condition for Intermediate Attribute). *Consider a DINA-based HLAM with a fixed \mathbf{Q} -matrix whose row vectors respect the hierarchy \mathcal{E} . Denote the (j, k) th entry of $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$ by $q_{j,k}^{\text{sparse}}$. The following statements hold for an intermediate attribute k .*

(a) *$\sum_{j=1}^J q_{j,k}^{\text{sparse}} \geq 1$ is necessary for the identifiability of $(\mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$.*

(b) *There exist scenarios where the equality in part (a) is achieved with $\sum_{j=1}^J q_{j,k}^{\text{sparse}} = 1$ and the identifiability of $(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, [\mathcal{E}], [\mathbf{p}])$ is guaranteed.*

Propositions 4–6 together characterize the different identifiability phenomena caused by different types of attributes in the attribute hierarchy graph. An intuitive explanation behind these conclusions is as follows. For a singleton attribute k that is not connected to any other attribute in the attribute hierarchy graph, no additional information is provided by the other attributes. Therefore the requirement of k being measured by ≥ 3 items in the \mathbf{Q} -matrix is necessary. This aligns well with the conclusion for a latent attribute model without any hierarchy established in Xu and Zhang (2016) and Gu and Xu (2019b), where

all the attributes are singletons and each needs to be measured by ≥ 3 items. However, this requirement can be relaxed for any other type of attribute which is somewhat connected in the attribute hierarchy graph. In particular, fewer measurements are needed for k in the \mathbf{Q} -matrix as more information is available for this attribute in the attribute hierarchy graph. For a root attribute k with some “child” or a leaf attribute with some “parent” as one-sided information, the requirement is relaxed to k being measured by ≥ 2 items in $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$; while for an intermediate attribute k with both some child and some parent as two-sided information, the requirement is further relaxed to k being measured by ≥ 1 items in $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$.

We next discuss the necessity of Condition C. Given a \mathbf{Q} , we denote by $\mathbf{Q}_{1:K,:}$ the submatrix consisting of its first K rows and by $\mathbf{Q}_{(K+1):J,:}$ the submatrix consisting of its last $J - K$ rows. For a \mathbf{Q} with rows respecting the attribute hierarchy, Condition C requires $\mathbf{Q}_{(K+1):J,k} \neq \mathbf{Q}_{(K+1):J,\ell}$ for any $k \neq \ell$ when $\mathbf{Q}_{1:K,:} = \mathbf{E}$. We have the following result.

Proposition 7 (Discussing Necessity of Condition C). *Consider a DINA-based HLAM with a fixed \mathbf{Q} whose row vectors respect the hierarchy \mathcal{E} . The condition that $\mathbf{Q}_{(K+1):J,k} \neq \mathbf{Q}_{(K+1):J,\ell}$ (when $\mathbf{Q}_{1:K,:} = \mathbf{E}$) is **necessary** for identifiability if both α_k and α_ℓ are singleton attributes.*

4.2 Bridging the Necessity and Sufficiency of the Identifiability Conditions

Still under a fixed and known \mathbf{Q} -matrix as in Section 4.1, we next investigate how the sufficient identifiability conditions for $(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$ can meet the necessary identifiability conditions proposed earlier in Propositions 5–7. In the next theorem, we establish that the individual necessary conditions established in Section 4.1 combined are actually sufficient to guarantee the identifiability in fully general scenarios. This result therefore establishes the general necessary and sufficient condition on the \mathbf{Q} -matrix for identifiability under an

arbitrary attribute structure.

Theorem 2 (Necessary and Sufficient Conditions under a Fully General \mathcal{E}). *Consider a DINA-based HLAM with a fixed \mathbf{Q} -matrix whose row vectors respect the hierarchy \mathcal{E} . Then Condition A and the following Condition B^* and C^* are necessary and sufficient for the identifiability of $(\boldsymbol{\theta}^+, \boldsymbol{\theta}^-, [\mathcal{E}], [\mathbf{p}])$.*

B^* . *In $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$, any intermediate attribute is each measured by ≥ 1 items, any root attribute and any leaf attribute is each measured by ≥ 2 items, and any singleton attribute is each measured by ≥ 3 items.*

C^* . *For any two singleton attributes α_k and α_ℓ , there is $\mathbf{Q}_{(K+1):J,k} \neq \mathbf{Q}_{(K+1):J,\ell}$. (Assume $\mathbf{Q}_{1:K,:} = \mathbf{E}$ under Condition A.)*

Theorem 2 covers any type of attribute structure and allows for any type of attributes in the attribute hierarchy graph. In the special case where there are no singleton attributes in the attribute hierarchy graph, the necessary and sufficient identifiability conditions in Theorem 2 can be simplified. We term such a family of hierarchies without any singleton attributes the *connected-graph hierarchy*.

Corollary 1 (Necessary and Sufficient Condition under a Connected Graph Hierarchy). *Consider a DINA-based HLAM with fixed \mathbf{Q} -matrix whose row vectors respect the hierarchy \mathcal{E} . Suppose the K attributes form a connected graph. Then Condition A and the following Condition D are necessary and sufficient for the identifiability of $(\mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, [\mathbf{p}])$.*

D. In $\mathcal{S}^{\mathcal{E}}(\mathbf{Q})$, any root attribute and any leaf attribute is each measured by ≥ 2 items, and any intermediate attribute is each measured by ≥ 1 items.

Remark 4. In the first extreme case, if $\mathcal{E} = \emptyset$ without any true hierarchy among attributes, then Conditions A, B^* , and C^* in Theorem 2 exactly become Conditions A, B, C in Theorem 1 in Section 3. In the second extreme case, if there does not exist any singleton attribute in the attribute hierarchy graph, then Condition B^* in Theorem 2 reduces to Condition D in the above Corollary 1; and Condition C^* in Theorem 2 should be understood as always satisfied and hence can be omitted. Namely, under a connected-graph hierarchy without any singleton attributes, the Conditions A, B^* , and C^* in Theorem 2 exactly reduce to Conditions A and D in Corollary 1. Therefore, Theorem 2 covers Corollary 1 as a special case and is indeed fully general. We state these two results separately to highlight both the most general form of the result, and also how the necessary and sufficient conditions simplify under the popular family of connected-graph hierarchy as depicted in Corollary 1.

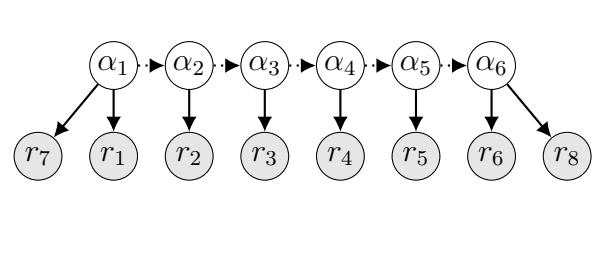
The following example illustrates the minimal requirements on \mathbf{Q} under those attribute hierarchies considered in Leighton et al. (2004).

Example 5. Under the linear hierarchy $\mathcal{E} = \mathcal{E}^{\text{linear}}$ in Fig 4(b), the 8×6 matrix $\mathbf{Q}_{8 \times 6}^{\text{linear}}$ shown in Fig 4(a) encodes the minimal requirement for identifiability. Fig 4(b) visualizes the sparsified version of $\mathbf{Q}_{8 \times 6}^{\text{linear}}$ as the directed solid edges from the latent attributes to the observed item responses. Under the so-called convergent hierarchy and divergent hierarchy presented earlier in Fig 3, the minimal requirement on \mathbf{Q} for model identifiability are presented in parts (c)-(d) and parts (e)-(f) of Figure 4, repectively. For the divergent hierarchy $\mathcal{E} = \mathcal{E}^{\text{div}}$ in Fig 4(f), the $\mathbf{Q}_{10 \times 6}^{\text{div}}$ in Fig 4(c) gives an identifiable model under minimal conditions.

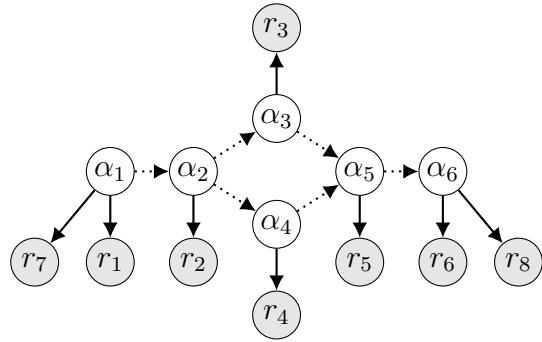
5. Identifiability of other HLAMs different from the DINA-based HLAMs

We also study identifiability of some other HLAMs in addition to the DINA-based HLAMs.

$$\mathbf{Q}_{8 \times 6}^{\text{linear}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

 (a) $\mathbf{Q}_{8 \times 6}^{\text{linear}}$

 (b) visualization of the sparsified $\mathcal{S}^{\mathcal{E}}(\mathbf{Q}_{8 \times 6}^{\text{linear}})$

$$\mathbf{Q}_{8 \times 6}^{\text{conv}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

 (c) $\mathbf{Q}_{8 \times 6}^{\text{conv}}$

 (d) visualization of the sparsified $\mathcal{S}^{\mathcal{E}}(\mathbf{Q}_{8 \times 6}^{\text{conv}})$

$$\mathbf{Q}_{10 \times 6}^{\text{div}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

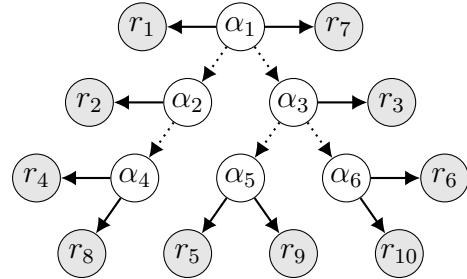
 (e) $\mathbf{Q}_{10 \times 6}^{\text{div}}$

 (f) visualization of the sparsified $\mathcal{S}^{\mathcal{E}}(\mathbf{Q}_{10 \times 6}^{\text{div}})$

 Figure 4: Minimally sufficient requirements on \mathbf{Q} for identifiability under the *linear* hierarchy, *convergent* hierarchy, and *divergent* hierarchy proposed in Leighton et al. (2004), respectively.

5.1 DINO-based HLAMs

As introduced earlier in Section 2, the DINO model is also a popular type of latent attribute model often used for psychiatric and clinical measurement of mental disorders (Templin

and Henson, 2006; de la Torre et al., 2018). A careful examination of the definitions of ideal responses Γ^{AND} and Γ^{OR} in (2.1) and (2.5) reveals the following relationship $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}^{\text{OR}} = 1 - \Gamma_{\mathbf{q}_j, \mathbf{1}_K - \boldsymbol{\alpha}}^{\text{AND}}$, where $\mathbf{1}_K - \boldsymbol{\alpha} = (1 - \alpha_1, \dots, 1 - \alpha_K)^\top$ also denotes an attribute pattern. Building upon such duality between DINA and DINO, the following proposition characterizes how the identifiability results obtained under a DINA-based HLAM can be translated into those under a DINO-based HLAM.

Proposition 8. *Consider a DINO-based HLAM with a fixed \mathbf{Q} -matrix and an unknown attribute hierarchy \mathcal{E} . Define the reversed attribute hierarchy $\mathcal{E}^{\text{reverse}}$ as*

$$\mathcal{E}^{\text{reverse}} = \{\ell \rightarrow k : \text{if } k \rightarrow \ell \text{ under the original hierarchy } \mathcal{E}\}. \quad (5.11)$$

(a) *For any $\boldsymbol{\alpha} \in \{0, 1\}^K$, $\boldsymbol{\alpha} \in \mathcal{A}(\mathcal{E})$ if and only if $\mathbf{1}_K - \boldsymbol{\alpha} \in \mathcal{A}(\mathcal{E}^{\text{reverse}})$. That is, any attribute pattern $\boldsymbol{\alpha}$ that is allowable under the original hierarchy \mathcal{E} if and only if another attribute pattern $\boldsymbol{\alpha}' = \mathbf{1} - \boldsymbol{\alpha}$ is allowable under the reversed hierarchy $\mathcal{E}^{\text{reverse}}$.*

(b) *The attribute hierarchy \mathcal{E} and model parameters under the DINO-based HLAM are identifiable if and only if the reversed attribute hierarchy $\mathcal{E}^{\text{reverse}}$ and model parameters are identifiable under a DINA-based HLAM with the same \mathbf{Q} -matrix.*

For any attribute hierarchy graph \mathcal{E} , the reversed hierarchy $\mathcal{E}^{\text{reverse}}$ in (5.11) is another directed graph among attributes, where the direction of each arrow in \mathcal{E} is reversed. Therefore, for the same set of K attributes, any root attribute in \mathcal{E} becomes a leaf attribute in $\mathcal{E}^{\text{reverse}}$, and any leaf in \mathcal{E} in turn becomes a root in $\mathcal{E}^{\text{reverse}}$. Any intermediate attribute or singleton attribute remain the same type when \mathcal{E} is reversed to be $\mathcal{E}^{\text{reverse}}$. Proposition 8 provides guidelines on how to check identifiability for a DINO-based HLAM using the identifiability results established earlier for DINA-based HLAMs. In particular, we have

the following necessary and sufficient conditions for identifiability of $(\mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$ under a DINO-based HLAM with a fixed \mathbf{Q} -matrix.

Corollary 2 (Necessary and Sufficient Conditions under a General \mathcal{E} for a DINO-based HLAM). *Consider a DINO-based HLAM with an attribute hierarchy \mathcal{E} and a fixed \mathbf{Q} -matrix whose rows respect the reversed hierarchy $\mathcal{E}^{\text{reverse}}$. Consider the following condition.*

Condition A^{}. The $\mathcal{E}^{\text{reverse}}$ -densified matrix $\mathcal{D}^{\mathcal{E}^{\text{reverse}}}(\mathbf{Q})$ contains a submatrix which is the reachability matrix under the reversed hierarchy $\mathcal{E}^{\text{reverse}}$.*

Then this Condition A^{}, and the earlier Conditions B^{*}–C^{*} given in Theorem 2 are necessary and sufficient for the identifiability of $(\mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p})$.*

5.2 Main-effect-based HLAMs

Another family of HLAMs in the literature (e.g., DiBello et al., 1995; von Davier, 2008; Henson et al., 2009) incorporate the main effects of latent attributes into the model. We next review these main-effect-based HLAMs in the following Example 6 and then provide the identifiability result for them.

Example 6 (HLAMs which Model the Main Effects of Attributes). The *main-effect HLAMs* assume the main effects of the attributes measured by each item indicated by \mathbf{q}_j play a role in distinguishing the item parameters. Under a main-effect HLAM the Bernoulli parameter $\theta_{j,\alpha}$ can be written as

$$\theta_{j,\alpha}^{\text{main-eff}} = f\left(\beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k\right), \quad (5.12)$$

where $f(\cdot)$ is a link function. Note not all the β -coefficients in the above display are needed in the model specification; instead, only when $q_{j,k} = 1$ will $\beta_{j,k}$ be needed and truly incorporated

in the model. Different link functions $f(\cdot)$ in (5.12) lead to different models, including the Linear Logistic Model (LLM; Maris, 1999) with $f(\cdot)$ being the sigmoid function, and the Additive Cognitive Diagnosis Model (ACDM; de la Torre, 2011) with $f(\cdot)$ being the identity. When $f(\cdot)$ is a monotonically increasing function, it is usually assumed in practice that each $\beta_{j,k} > 0$ wherever $q_{j,k} = 1$ for interpretability.

There are also *all-effect HLAMs* that model not only the main effects but also all the interaction effects of attributes. The Bernoulli parameter $\theta_{j,\alpha}$ of an all-effect model is

$$\theta_{j,\alpha}^{\text{all-eff}} = f\left(\beta_{j,0} + \sum_{k=1}^K \beta_{j,k} (q_{j,k} \alpha_k) + \sum_{1 \leq k_1 < k_2 \leq K} \beta_{j,k_1 k_2} (q_{j,k_1} \alpha_{k_1}) (q_{j,k_2} \alpha_{k_2}) + \cdots + \beta_{j,12\cdots K} \prod_{k=1}^K (q_{j,k} \alpha_k)\right). \quad (5.13)$$

Similarly as in (5.12), not all the β -coefficients above are needed in the model specification. When $f(\cdot)$ in (5.13) is the identity function, (5.13) gives the Generalized DINA (GDINA) model in de la Torre (2011); and when $f(\cdot)$ is the sigmoid function, (5.13) gives the Log-linear Cognitive Diagnosis Models (LCDMs) in Henson et al. (2009); see also the General Diagnostic Models (GDMs) in von Davier (2008). We generally call the main-effect HLAMs in (5.12) and the all-effect HLAMs in (5.13) the main-effect-based HLAMs, because they both incorporate the main effects of the latent attributes in to the model.

Under the main-effect-based HLAMs, the probability mass function of the J -dimensional random response vector \mathbf{R} can be generally written as

$$P(\mathbf{R} = \mathbf{r} \mid \mathbf{Q}, \mathcal{E}, \boldsymbol{\theta}^+, \boldsymbol{\theta}^-, \mathbf{p}) = \sum_{\alpha \in \mathcal{A}(\mathcal{E})} p_{\alpha} \prod_{j=1}^J \theta_{j,\alpha}^{r_j} \times (1 - \theta_{j,\alpha})^{1-r_j},$$

where $\mathbf{r} \in \{0, 1\}^J$ is an arbitrary response pattern. Notably, these main-effect-based HLAMs generally have quite different algebraic structures from the family of two-parameter HLAMs, the DINA and the DINO models. The key structure of any two-parameter HLAM is captured

by the ideal response $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}}$ in (2.1) or (2.4), under the “AND” or “OR” operations, respectively. Intuitively, the two-parameter HLAMs are characterized by a probabilistic version of the *Boolean product* of two groups of binary vectors, the group of \mathbf{q}_j ’s and the group of $\boldsymbol{\alpha}$ ’s; however, this is not the case for any HLAM in Example 6 due to the incorporation of the main effects of attributes. Indeed, incorporating main effects in the form of $\sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k$ in (5.12) or (5.13) is taking a *inner product* of vectors \mathbf{q}_j , $\boldsymbol{\alpha}$ and an additional β -coefficient vector, rather than the Boolean product. Because of such distinction, the necessary and sufficient identifiability conditions derived carefully for the two-parameter HLAMs in Sections 3-4 are not applicable to main-effect-based HLAMs.

Next we give a set of sufficient conditions for the identifiability of main-effect-based HLAMs. The technical concept of $\Gamma(\mathbf{Q}, \mathcal{E})$ (specifically, with $\Gamma = \Gamma^{\text{AND}}$ defined in (2.1)) introduced earlier in Section 3 is still useful here. Denote the collection of all the per-item Bernoulli parameters by $\boldsymbol{\Theta} = (\theta_{j,\boldsymbol{\alpha}})$. We have the following theorem.

Theorem 3 (Identifiability of HLAMs which Model the Main Effects of Attributes). *Consider an HLAM that incorporates the main effects of the attributes with \mathbf{Q} and \mathcal{E} both unknown. Suppose $\boldsymbol{\Theta}$ satisfies a natural inequality constraint $\theta_{j,\boldsymbol{\alpha}} \neq \theta_{j,\boldsymbol{\alpha}'} \text{ if } \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} \neq \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}'}$. If $\Gamma(\mathbf{Q}, \mathcal{E})$ satisfies the following conditions with the number of columns known, then the $(\boldsymbol{\Theta}, \mathbf{p})$ and $\Gamma(\mathbf{Q}, \mathcal{E})$ are identifiable.*

E. There exist two disjoint sets of items $S_1, S_2 \subseteq [J]$, such that $\Gamma(\mathbf{Q}_{S_1,:}, \mathcal{E})$ and $\Gamma(\mathbf{Q}_{S_2,:}, \mathcal{E})$

each has distinct column vectors.

F. For any $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}' \in \mathcal{A}(\mathcal{E})$, there exists some $j \notin S_1 \cup S_2$ such that $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} \neq \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}'}$.

G. For any $\boldsymbol{\alpha} \in \mathcal{A}(\mathcal{E})$, $\boldsymbol{\alpha}' \in \{0, 1\} \setminus \mathcal{A}(\mathcal{E})$, there exists some $j \in [J]$ such that $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} \neq \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}'}$.

In addition to the above three conditions, if \mathbf{Q} is known in part to contain an identity submatrix I_K , then the attribute hierarchy \mathcal{E} is identifiable from $\Gamma(\mathbf{Q}, \mathcal{E})$.

For the main-effect-based HLAMs, the ideal response matrix $\Gamma(\mathbf{Q}, \mathcal{E})$ may not sharply characterize the entire latent structure due to the incorporation of the main effects, which is in contrary to the DINA-based HLAMs. To see this, considering two latent patterns $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ with $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} = \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}'} = 0$, then the specification in (5.12) or (5.13) implies it is possible that $\theta_{j, \boldsymbol{\alpha}} \neq \theta_{j, \boldsymbol{\alpha}'}$. Therefore it is hard, if at all possible, to explicitly characterize the necessary identifiability conditions in terms of $\Gamma(\mathbf{Q}, \mathcal{E})$ for main-effect-based HLAMs. However, the $\Gamma(\mathbf{Q}, \mathcal{E})$ is still useful to derive sufficient conditions for identifiability, as revealed in the above Theorem 3. This is because if $\Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}} = \Gamma_{\mathbf{q}_j, \boldsymbol{\alpha}'} = 1$, the two attribute patterns $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ both satisfy $\boldsymbol{\alpha} \succeq \mathbf{q}_j$ and $\boldsymbol{\alpha}' \succeq \mathbf{q}_j$ by the definition in (2.1). This implies both patterns $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ possess all the attributes measured by the vector \mathbf{q}_j . As a result, the definition of main-effect-based models in (5.12) or in (5.13) shows that there must be $\theta_{j, \boldsymbol{\alpha}} = \theta_{j, \boldsymbol{\alpha}'}$ for these two patterns. This intuitively explains why $\Gamma(\mathbf{Q}, \mathcal{E})$ can be used to describe a set of sufficient identifiability conditions for the main-effect-based HLAMs.

We make a remark on the relationship between the main-effect-based HLAMs and the DINA-based HLAMs studied in the previous Sections 3–4. On the one hand, the main-effect-based HLAMs are more general than DINA-based HLAMs in the sense that the formulation of $\theta_{j, \boldsymbol{\alpha}}^{\text{main-eff}}$ in (5.12) or $\theta_{j, \boldsymbol{\alpha}}^{\text{all-eff}}$ in (5.13) can generally allow for more than two Bernoulli parameters for each j , while DINA-based HLAMs always have two parameters θ_j^+ and θ_j^- for each j . On the other hand, however, we would like to point out that in this work we still put the main focus on the DINA-based two-parameter HLAMs, which are widely used in the motivating applications of cognitive diagnosis in educational settings. Indeed, these edu-

tional settings are where the attribute hierarchy receives the most attention in modeling the sequential acquisition of skill attributes (e.g., Leighton et al., 2004; Gierl et al., 2007; Wang and Lu, 2020). On the practical side, assuming the conjunctive relationship among the attributes as in DINA is often believed to be suitable for modeling the response mechanism of diagnostic test items in such settings (e.g., Junker and Sijtsma, 2001; de la Torre and Douglas, 2004). On the theoretical side, the identifiability of two-parameter DINA-based HLAMs is also more intriguing to study because of the Boolean product involved. The rich combinatorial nature of such models gives the opportunity to close the gap between the necessity and sufficiency of identifiability requirements; interestingly, these minimal requirements are explicit conditions on the discrete structure: the **Q**-matrix and attribute types, as depicted in Section 4. Therefore, we believe that closely examining the DINA-based two-parameter HLAMs and establishing the minimal identifiability conditions for them (as done in Sections 3–4) are highly desirable, due to their theoretical interest and practical relevance.

6. Discussion

In this paper, we provide a first study on identifiability of the hierarchical latent attribute model, a complex-structured latent variable model popular in modeling modern assessment data. We propose sufficient identifiability conditions that explicitly depend on the attribute hierarchy graph and the structural **Q**-matrix. We also discuss the necessity of the identifiability conditions and sharply characterize the different impacts on identifiability cast by different types of attributes in the attribute hierarchy graph. In this paper we mainly focus on the basic and popular HLAMs, the DINA-based HLAMs, where each item is modeled using two parameters. We also extend the theory to other types of HLAMs in Section 5.

One nice implication of identifiability is the estimability of both the latent structure and the parameters that define the probabilistic model. When the proposed conditions are satisfied, all the components of the HLAM can be uniquely and consistently estimated from data based on maximum likelihood. In practical data analysis under the HLAM framework, if the \mathbf{Q} and \mathcal{E} are specified by domain experts or applied researchers, then before seeing any data, one can check whether \mathbf{Q} and \mathcal{E} satisfy our proposed conditions to assess model identifiability. On the other hand, if \mathbf{Q} and \mathcal{E} are not known and one hopes to estimate them exploratorily from data, our identifiability results can also be useful. In such scenarios, one can check whether the estimated $\hat{\mathbf{Q}}$ and $\hat{\mathcal{E}}$ satisfy necessary identifiability conditions; if not, then more careful investigation of the diagnostic test design may be needed. Therefore, this study provides useful insights into designing valid diagnostic tests and drawing valid scientific conclusions from assessment data under a potentially complicated attribute hierarchy.

Acknowledgements. This research was partially supported by NSF CAREER SES-1846747, DMS-1712717, SES-1659328; and NIH R01ES027498, R01ES028804. The authors are grateful to the editor, associate editor, and two reviewers for constructive and helpful comments.

Supplementary Materials. The supplementary material contains proofs of the theorems and some illustrative examples.

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*, 37:3099–3132.

Anandkumar, A., Ge, R., Hsu, D., Kakade, S. M., and Telgarsky, M. (2014). Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 15(1):2773–2832.

Anderson, T. W. (2009). *An introduction to multivariate statistical analysis*. John Wiley & Sons, New York.

Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics*, 40(1):436–465.

Cai, Y., Tu, D., and Ding, S. (2018). Theorems and methods of a complete Q matrix with attribute hierarchies under restricted Q-matrix design. *Frontiers in Psychology*, 9:1413.

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

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

Choi, M. J., Tan, V. Y., Anandkumar, A., and Willsky, A. S. (2011). Learning latent tree graphical models. *Journal of Machine Learning Research*, 12(May):1771–1812.

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

de la Torre, J. and Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3):333–353.

de la Torre, J., van der Ark, L. A., and Rossi, G. (2018). Analysis of clinical data from a cognitive diagnosis modeling framework. *Measurement and Evaluation in Counseling and Development*, 51(4):281–296.

DiBello, L. V., Stout, W. F., and Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. *Cognitively diagnostic assessment*, pages 361–389.

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

Fraley, C. and Raftery, A. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631.

Ghahramani, Z. and Griffiths, T. (2006). Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems*, pages 475–482.

Gierl, M. J., Leighton, J. P., and Hunka, S. M. (2007). Using the attribute hierarchy method to make diagnostic inferences about respondents’ cognitive skills. *Cognitive diagnostic assessment for education: Theory and applications*, Cambridge, UK: Cambridge University Press, pages 242 – 274.

Gu, Y. and Xu, G. (2019a). Learning attribute patterns in high-dimensional structured latent attribute models. *Journal of Machine Learning Research*, 20(115):1–58.

Gu, Y. and Xu, G. (2019b). The sufficient and necessary condition for the identifiability and estimability of the DINA model. *Psychometrika*, 84(2):468–483.

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.

Hamada, M. and Wu, C. J. (1992). Analysis of designed experiments with complex aliasing. *Journal of Quality Technology*, 24(3):130–137.

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.

Hernán, M. A. and Robins, J. M. (2006). Instruments for causal inference: an epidemiologist’s dream? *Epidemiology*, pages 360–372.

Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800.

Junker, B. W. and Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3):258–272.

Köhn, H.-F. and Chiu, C.-Y. (2019). Attribute hierarchy models in cognitive diagnosis: identifiability of the latent attribute space and conditions for completeness of the Q -matrix. *Journal of Classification*, 36(3):541–565.

Leighton, J. P., Gierl, M. J., and Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: a variation on Tatsuoka’s rule-space approach. *Journal of Educational Measurement*, 41(3):205–237.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64(2):187–212.

McLachlan, G. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

O’Brien, K. L., Baggett, H. C., Brooks, W. A., et al. (2019). Causes of severe pneumonia requiring hospital admission in children without hiv infection from africa and asia: the perch multi-country case-control study. *The Lancet*.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial intelligence*, 29(3):241–288.

Ravanbakhsh, S., Póczos, B., and Greiner, R. (2016). Boolean matrix factorization and noisy completion via message passing. In *Proceedings of the 33rd International Conference on Machine Learning-Volume 48*, pages 945–954.

Rukat, T., Holmes, C. C., Titsias, M. K., and Yau, C. (2017). Bayesian Boolean matrix factorisation. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2969–2978. JMLR. org.

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

Swanson, S. A., Miller, M., Robins, J. M., and Hernán, M. A. (2015). Definition and evaluation of the monotonicity condition for preference-based instruments. *Epidemiology (Cambridge, Mass.)*, 26(3):414.

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

Tatsuoka, M. M. (1986). Graph theory and its applications in educational research: A review and integration. *Review of Educational Research*, 56(3):291–329.

Templin, J. and Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, 79(2):317–339.

Templin, J. L. and Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods*, 11(3):287.

Tu, D., Wang, S., Cai, Y., Douglas, J., and Chang, H.-H. (2019). Cognitive diagnostic models with attribute hierarchies: Model estimation with a restricted q-matrix design. *Applied Psychological Measurement*, 43(4):255–271.

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 DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology*, 67(1):49–71.

Wang, C. and Lu, J. (2020). Learning attribute hierarchies from data: two exploratory approaches. *Journal of Educational and Behavioral Statistics*.

Wang, S., Yang, Y., Culpepper, S. A., and Douglas, J. A. (2018). Tracking skill acquisition with cognitive diagnosis models: a higher-order, hidden markov model with covariates. *Journal of Educational and Behavioral Statistics*, 43(1):57–87.

Wu, Z., Deloria-Knoll, M., Hammitt, L. L., Zeger, S. L., and for Child Health Core Team, P. E. R. (2016). Partially latent class models for case–control studies of childhood pneumonia aetiology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(1):97–114.

Wu, Z., Deloria-Knoll, M., and Zeger, S. L. (2017). Nested partially latent class models for dependent binary data; estimating disease etiology. *Biostatistics*, 18(2):200–213.

Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *The Annals of Statistics*, 45:675–707.

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*, 81:625–649.

Yuan, M., Joseph, V. R., and Zou, H. (2009). Structured variable selection and estimation. *The Annals of Applied Statistics*, pages 1738–1757.