Statistical Inference and Experimental Design for Q-matrix Based Cognitive Diagnosis Models

Stephanie S. Zhang

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

©2014
Stephanie S. Zhang
All Rights Reserved

ABSTRACT

Statistical Inference and Experimental Design for Q-matrix Based Cognitive Diagnosis Models

Stephanie S. Zhang

There has been growing interest in recent years in using cognitive diagnosis models for diagnostic measurement, i.e., classification according to multiple discrete latent traits. The Q-matrix, an incidence matrix specifying the presence or absence of a relationship between each item in the assessment and each latent attribute, is central to many of these models. Important applications include educational and psychological testing; demand in education, for example, has been driven by recent focus on skills-based evaluation. However, compared to more traditional models coming from classical test theory and item response theory, cognitive diagnosis models are relatively undeveloped and suffer from several issues limiting their applicability. This thesis exams several issues related to statistical inference and experimental design for Q-matrix based cognitive diagnosis models.

We begin by considering one of the main statistical issues affecting the practical use of Q-matrix based cognitive diagnosis models, the identifiability issue. In sta-

tistical models, identifiability is prerequisite for most common statistical inferences, including parameter estimation and hypothesis testing. With Q-matrix based cognitive diagnosis models, identifiability also affects the classification of respondents according to their latent traits. We begin by examining the identifiability of model parameters, presenting necessary and sufficient conditions for identifiability in several settings.

Depending on the area of application and the researchers degree of control over the experiment design, fulfilling these identifiability conditions may be difficult. The second part of this thesis proposes new methods for parameter estimation and respondent classification for use with non-identifiable models. In addition, our framework allows consistent estimation of the severity of the non-identifiability problem, in terms of the proportion of the population affected by it. The implications of this measure for the design of diagnostic assessments are also discussed.

Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, I thank my advisors, Professors Zhiliang Ying and Jingchen Liu, for their support and guidance through these past few years. They have been essential to my intellectual development as a researcher, and this work would never have been without their help.

I would also like to thank my committee members, Professors Michael Sobel, Lawrence DeCarlo, and Melanie Wall, for their advice and helpful insights, and also the time and attention they were willing to give my work. Much thanks also goes to Professor DeCarlo for inspiring and working together with me to develop my work on nonidentifiable models. Thank you also to the faculty of the Department of Statistics and to Dood Kalicharan for all their help over the years.

For their support and friendship, I would like to thank my fellow students in the department. I would especially like to thank my fellow students in the educational statistics research group, including my collaborator, Gongjun Xu. I have learned a lot from all of you and our shared experiences have greatly enriched my time as a student here.

Finally, many thanks to my family: to my mother for her unfailing patience and support, to my father for his wisdom, advice, and inspiring example, and to my sister, Kimberly, for always brightening my day. I love you.

To my father and mother, for all their love and care.

Chapter 1

Introduction

According to the Oxford English Dictionary, to diagnose a disease is "to distinguish and determine its nature from its symptoms; to recognize and identify by careful observation" (OED Online, 2014). In essence, based on the observable symptoms, the medical practitioner is attempting to arrive at a classification-based decision as to the underlying cause. In psychometrics, diagnosis follows a similar vein. Diagnostic measurement is the process of arriving at a classification-based decision about an individual's latent traits, based on the observed responses to the items in a diagnostic assessment. The idea of conducting diagnostic measurement has been gaining increasing traction, thanks in part to demand from the fields of education and psychology. Both are traditional psychometric fields, with strong interest in measurement techniques for psychological constructs such as skills, knowledge, personality traits, or psychological disorders. While diagnosis has strong medical roots, educational interest in the field is more recent. A diagnostic, skills-based focus is a key part of recent government initiatives such as the United States Department of Education's Race to the Top, which included "building data systems that measure student growth and success, and inform teachers and principals about how they can improve instruction" as one of its four main goals (U.S. Department of Education 2009). Measuring students' growth and success means obtaining diagnostic information about their skill set; this is very important for constructing efficient, focused remedial strategies for improving student and teacher results. In the private sector, consider the College Board's feedback on the Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT®). Traditionally the SAT has been a way to rank students according to mathematical and verbal ability; it was not developed for diagnostic purposes. Diagnostic feedback has recently become increasingly prevalent, though, and it is especially important for the PSAT/NMSQT®, since diagnostic feedback from the preliminary exam can help students target further preparation for the SAT. Thus, beyond the well-know writing, math, and verbal scores resulting from any such examination, PSAT/NMSQT® score reports also provide diagnostic skills-based feedback on specific areas such as numbers and operations, algebra and functions, geometry and measurement, and data, statistics, and probability; in fact, the PSAT/NMSQT is the first nationally standardized test to give diagnostic skills-based feedback ® (Roussos, Templin, & Henson 2007).

This thesis focuses on statistical inference and experimental design for Q-matrix based cognitive diagnosis models. Cognitive diagnosis models (CDMs) are a modern tool for conduction psychometric diagnostic measurement. Though they exhibit much promise, and have had some success in specific applications, the field is still young and there are several issues that need to be addressed before techniques involving CDMs can be reliably applied on a broader scale.

We begin by considering one of the main statistical issues affecting the practical use of Q-matrixe based CDMs, the identifiability issue. In statistical models, identifiability is prerequisite for most common statistical inferences, including parameter estimation and hypothesis testing. With Q-matrix based CDMs, identifiability also affects the classification of respondents according to their latent traits. Yet a long-standing problem with Q-matrix based CDMs is that the models are often not identifiable (DeCarlo 2011; DiBello, Stout, & Roussos 1995; G. Maris & Bechger 2009; C. Tatsuoka 2009; K. K. Tatsuoka 1991). In the first part of this thesis, we provide the necessary and sufficient conditions for the identifiability of two well-known CDMs. We also study how these identifiability conditions change under attribute hierarchy, which limits the appearance of certain latent trait patterns in the population.

Depending on the area of application and the researchers degree of control over the experiment design, fulfilling these identifiability conditions may be difficult. The second part of this thesis proposes new methods for parameter estimation and respondent classification for use with non-identifiable models. In addition, our framework allows the quantification of the severity of the identifiability issue in terms of the proportion of the population affected by it. A version of this work can be found in Zhang, DeCarlo, and Ying (2013). We conclude with some guidelines for assessment design, including suggestions on balancing the identifiability issue against other design concerns.

Chapter 2

Introduction to Q-matrix based cognitive diagnostic models

To measure something is to assign it numbers in a systematic way so as to represent some intrinsic quantitative property (Allen & Yen 1979/2002, p. 2). In the physical world, measurement can generally be done rather directly: lengths can be measured by rulers, weights by scales, and time by clocks. On the other hand, psychological constructs such as knowledge, skills or personality traits are latent, unobservable traits. They can be measured indirectly via assesments such as tests or questionnaires, and the field concerning itself with psychological measurement is known as psychometrics. Psychometrics, like other branches of measurement theory, concerns itself with the evaluation of measurement quality, the improvement of measurement accuracy and interpretability, and the development of newer, better measurement tools for measurement (Allen & Yen 1979/2002, p. 2). These tools include early developments such as classifcal test theory (CTT) and item response theory (IRT) and more recent developments such as cognitive diagnosis modeling.

2.1 Traditional psychometric techniques

2.1.1 Classical test theory

Early developments in measurement theory include the recognition of the presence of measurement error and its characterization as a random variable, along with the establishment of correlation as an important statistical concept and recognition of the effect of measurement error on its estimation (Traub 2005). One of the first formalizations of measurement theory for psychometrics was CTT, born from the pioneering work of Spearman (1904) on correcting correlation coefficients for attenuation due to measurement error; see also Lord, Novick, and Birnbaum (1968); Novick (1966) for more details.

In CTT, individuals are assumed to have a true test score, T, that would be obtained if there were no errors from measurement. The observed score, X, is the result of incorporating some independent additive error E, i.e.,

$$X = T + E.$$

Disentangling the variability in subject ability from other external sources of variability is a central concern of CTT; in particular, the reliability

$$\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

of a test, defined as the ratio between the variance of the true score and the observed score, can be considered CTT's most important concept.

One of the shortcomings of CTT is that it in no way separates the characteristics of the test from the characteristics of the individual; they come together in one quantity of interest, the 'true score' T. This leaves psychometricians with what is still quite an indirect measure of any psychological constructs of interest (Hambleton, Swaminathan, & Rogers 1991). Later developments in psychometrics, such as IRT and cognitive diagnosis modeling, have been able to overcome this shortcoming, at the cost of stronger model assumptions.

2.1.2 Item response theory

Item response theory (IRT) was developed in the 1950s and 1960s by researchers such as Frederic M. Lord of the Educational Testing Service and Danish mathematician Georg Rasch. In IRT, the binary responses to the questions, or items, in an assessment depend on a unidimensional continuous latent trait θ representing, for instance, the respondent's ability. Given the latent trait θ, responses are assumed to be conditionally independent following some distribution $p(\theta) := \Pr(X_j = 1|\theta)$; this property is known as *local independence* (Lord 1980). The function $p(\theta)$ is known as the item response function, item characteristic curve (Lord 1952), or item curve (Tucker 1946). The item response function is generally taken to be monotonically increasing in θ, so that increasing ability levels correspond to larger probabilities of correct responses. In particular, logistic and normal ogive models are popular. A popular IRT model is the Rasch model, also known as the one-parameter logistic IRT model (Rasch 1960; 1961). The item response function for the i-th item is

$$p_i(\theta) = \frac{1}{1 + \exp[-(\theta - b_i)]},$$

where b_i is a location parameter representing the item's difficult; it is also the nominal 'one parameter.' A high ability θ increases the probability of correct response $p_i(\theta)$, but a high item difficulty b_i decreases that same probability. The Rasch model is mathematically convenient to work with, but imposes strong assumptions. In particular, it assumes that all items discriminate between subjects in a similar way.

Birnbaum's two parameter logistic (2-PL) model generalizes the Rasch model by including an item discrimination parameter that affects the slope of the logistic curve (Lord et al. 1968). Under this model, the i-th item has item response function

$$p_i(\theta) = \frac{1}{1 + \exp[-a_i(\theta - b_i)]},$$

where a_i is the new item discrimination parameter and b_i still represents the item's difficulty. The 2-PL model can be further generalized to the three-parameter logistic (3-PL) model, which includes item guessing parameters affecting the lower asymptotic limit of the item response function, i.e., the 'guessing' probability of correct response for an individual with no ability $\theta = -\infty$ (Lord 1980; Lord et al. 1968).

2.2 Cognitive diagnosis modeling

Traditional psychometric approaches such those mentioned above generally focus on scaling and ranking individuals along some latent continuum. However, in diagnostic classification, the aim is instead to detect the presence or absence of multiple fine-grained skills or attributes. This provides more informative feedback on, for example, student skillsets, and allows for the design of more effective intervention strategies (Rupp, Templin, & Henson 2010).

Researchers have brought a number of tools to bear on the problem of diagnostic classification, including multidimensional IRT, factor analysis, the rule-space method, the attribute hierarchy method, clustering methods, and CDMs; for a recent review, see Rupp et al. (2010). The last of these approaches, CDMs, shares characteristics such as local independence with the IRT approach, but in CDMs the latent variables are multidimensional and discrete. These latent variables may represent mastery of a finite set of skills in an educational setting, or the state of having a particular psychiatric disorder in a psychological testing setting. Statistically speaking, CDMs can be categorized as latent structure models; in particular, they are restricted latent class models within the broader family of generalized linear and nonlinear mixed models (von Davier 2009). Analysis of CDMs generally results in a probabilistic attribute profile, which can be parsed into a classification decision for the respondents; this makes them well-suited to diagnostic classification. Well known models include the Deterministic Input, Noisy "And" Gate (DINA) model, the Deterministic Input, Noisy "Or" Gate (DINO) model, the Noisy Inputs, Deterministic "And" Gate (NIDA) model, the Noisy Inputs, Deterministic "Or" Gate (NIDO) model, and the Conjunctive Reparameterized Unified Model (C-RUM), among others (de la Torre 2008; de la Torre & Douglas 2004; Haertel 1989; Junker & Sijtsma 2001; E. Maris 1999; Rupp et al. 2010; Templin 2006; Templin & Henson 2006).

Most diagnostic classification models begin from the same basic setting, in which subjects known as respondents provide observed responses to the items which make up the assessment. These responses depend in some way on unobserved latent attributes. We consider tests consisting of a pre-specified number of items J depending on a known number of attributes K, given to N subjects.

Some specific terminology and notations are listed below.

Attributes are conceptualizations the respondent's (unobserved) states of mastery of certain skills. If we suppose that there are N respondents and K attributes, let the matrix of attributes be $A = (\alpha_{i,k})$, where where $\alpha_{i,k} \in \{0, 1\}$ indicates the presence or absence of the k-th attribute in the i-th respondent. An *attribute profile* $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^\top$ is the vector of all attributes; note that the superscript $\top$ denotes the matrix transpose operation. An individual respondent i will have attribute profile $\boldsymbol{\alpha}^i$ such that $\alpha^i_k = \alpha_{i,k}$.

Responses are the respondent's binary responses to items. Given N respondents and J items, the responses can be written as a $N \times J$ matrix $X = (X_{i,j})$, where $X_{i,j} \in \{0, 1\}$ is the response of the i-th respondent to the j-th item. The i-th respondent's responses will be denoted by the vector $\boldsymbol{X}^i$, where the j-th element $X^i_j = X_{i,j}$ for all i, j.

The Q-matrix , developed by K. K. Tatsuoka (1983), describes the underlying cognitive structure of the assessment; in other words, it provides the link between the items and attribute. It is a $J \times K$ matrix $Q = (q_{j,k})$, where for each j, k, $q_{j,k} \in \{0, 1\}$ indicates whether the j-th item requires the k-th attribute. The set of all attributes linked to a particular item can be considered that item's *attribute requirement.* From the Q-matrix we can extract the attribute requirements of an item j as the vector $\boldsymbol{q}^j$, where the k-th element $q^j_k = q_{j,k}$ for all j, k.

Note that when unspecified, vectors indexed by $\boldsymbol{\alpha}$ or $\boldsymbol{x}$ will presume the natural lexicographic (alphabetic) ordering on the indices. For example, in the case of $\{0, 1\}^2$, $(0, 0) < (0, 1) < (1, 0) < (1, 1)$.

2.2.1 Types of CDMs

Depending on the parameterization of the model and the interpretation of the Q-matrix, many different families of models may arise. Generally, CDMs can written as restricted logistic models with latent classes. In the unrestricted model, $p_{j,\boldsymbol{\alpha}} = \Pr(X_{i,j} = 1|\boldsymbol{\alpha}^{(i)} = \boldsymbol{\alpha})$ is a function of the full kernel of 2^K terms incorporating all interactions of the K latent attributes:

$$\begin{aligned}\log\left(\frac{p_{j,\boldsymbol{\alpha}}}{1-p_{j,\boldsymbol{\alpha}}}\right) &= \lambda_0 + \sum_{k=1}^{K}\lambda_k\alpha_k + \sum_{1\le k_1<k_2\le K}\lambda_{k_1,k_2}\alpha_{k_1}\alpha_{k_2} + \cdots + \lambda_{1,\ldots,K}\prod_{k=1}^{K}\alpha_k \\ &= \sum_{L=0}^{K}\sum_{1\le k_1<\cdots<k_L\le K}\lambda_{k_1,\ldots,k_L}\prod_{\ell=1}^{L}\alpha_{k_\ell}.\end{aligned}$$

The idea of Q-matrix based CDMs being encompassed by the unrestricted latent class model dates back to von Davier (2005), who coined the term "general diagnostic model;" Rupp et al. (2010) refers to this as the log-linear cognitive diagnosis model framework.

Since the Q-matrix indicates the lack of a relationship between certain attributes and items, at the most general level a Q-matrix based CDM should restrict terms incorporating attributes that have no relation to the j-th item to zero; mathematically, $\lambda_{k_1,\ldots,k_L} = 0$ when $q_{jk} = 0$ for some $k \in \{k_1, \ldots, k_L\}$. When this is the only restriction, and the model is written with the identity link, we have the G-DINA model of de la Torre (2008; 2011).

Further restrictions can produce many of the wide variety of CDMs that have arisen from different applications. Q-matrix based CDMs are often divided into two major groups, depending how the latent variables interact. In compensatory models, high values on one trait may compensate for low values on another. For example, in psychological screening, suppose that the items are symptoms and the attributes are psychological disorders. If a particular symptom can be the result of one of multiple disorders, having one disorder will 'compensate' for not having another in terms of the respondent's chance of expressing that symptom. Other models are non-compensatory in nature; low values on a particular attribute linked to an item cannot by compensated by high values in other linked attributes. Non-compensatory models are often appropriate in educational settings where a combination of skills are needed to solve a particular problem, and missing one of the necessary skills cannot be compensated by having another skill. We now examine two popular CDMs, one compensatory and one not, that shall be the focus of our work.

2.2.2 The DINA model and its variants

The DINA model is one of the most widely used CDMs, and is particularly popular in the context of educational testing. Underlying the model is the assumption that, before randomness in response come into play, a respondent must have mastered all necessary (as specified by a loading matrix known as the Q-matrix) attributes required by a particular item in order to answer that item correctly. Missing any attribute linked to an item by the Q-matrix results in an incorrect answer; thus, the absence of one attribute cannot be compensated by the presence of another and the DINA model is non-compensatory. The DINA model is well-suited to educational assessments in areas such as mathematics where correct answers are obtained by correctly employing all of an item's required skills together. It has been frequently employed in the analysis of assessments, including the widely analyzed fraction subtraction data set of K. K. Tatsuoka (1990). See de la Torre (2009); de la Torre and Douglas (2004; 2008); DeCarlo (2011); Henson, Templin, and Willse (2009); Templin, Henson, and Douglas (2006) for examples.

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

$$\xi_j(Q, \boldsymbol{\alpha}) = \prod_{k=1}^{K} (\alpha_k)^{q_{j,k}} = I(\alpha_k \geq q_{j,k} : k = 1, \ldots, K), \tag{2.1}$$

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

In the DINA model, uncertainty is incorporated at the item level, using the slipping and guessing parameters $\boldsymbol{s}$ and $\boldsymbol{g}$; the names "slipping" and "guessing" arise from the educational applications. For each item $j = 1, \ldots, J$, the slipping parameter $s_j = \Pr(X_j = 0 | \xi_j = 1)$ denotes the probabilitiy of the respondent making an incorrect response despite mastering all necessary skills and having a correct ideal response; similarly, the guessing parameter $g_j = \Pr(X_j = 1 | \xi_j = 0)$ denotes the probability of a correct response despite an incorrect ideal response. In the technical development, it is more convenient to work with the complement of the slipping parameter, referred to as $\boldsymbol{c} = 1 - \boldsymbol{s}$; this alternative parameterization is used throughout the paper.

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

$$\Pr(X_j = 1 | Q, \boldsymbol{\alpha}, \boldsymbol{c}, \boldsymbol{g}) = c_j^{\xi_j(Q,\boldsymbol{\alpha})} g_j^{1-\xi_j(Q,\boldsymbol{\alpha})}. \tag{2.2}$$

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

$$p(\boldsymbol{x} | Q, \boldsymbol{\alpha}, \boldsymbol{c}, \boldsymbol{g}) = \prod_{j=1}^{J} c_j^{\xi_j x_j} g_j^{(1-\xi_j) x_j} (1 - c_j)^{\xi_j (1 - x_j)} (1 - g_j)^{(1-\xi_j)(1-x_j)}, \tag{2.3}$$

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

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

$$\Pr(\boldsymbol{\alpha}^i = \boldsymbol{\alpha}) = \pi_{\boldsymbol{\alpha}},$$

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

$$p(\boldsymbol{x}|Q,\boldsymbol{c},\boldsymbol{g},\boldsymbol{\pi}) = \sum_{\boldsymbol{\alpha}\in\{0,1\}^K} p(\boldsymbol{x}|Q,\boldsymbol{c},\boldsymbol{g},\boldsymbol{\alpha})\pi_{\boldsymbol{\alpha}} \tag{2.4}$$

for each $\boldsymbol{x} \in \{0,1\}^J$. The conditional probabilities $p(\boldsymbol{x}|Q,\boldsymbol{c},\boldsymbol{g},\boldsymbol{\alpha})$ are calculated as in (2.3).

Several variants of the DINA can be constructed by either restricting $\boldsymbol{\pi}$ to some lower-dimensional subspace or putting a prior on it. For example, assuming independence among the attributes so that

$$\pi_{\boldsymbol{\alpha}} = \prod_{k=1}^{K} p(\alpha_k)$$

reduces the $2^K - 1$-dimensional parameter space of all $p \in [0,1]^{2^K}$ such that $\sum \pi_{\boldsymbol{\alpha}} = 1$ to the K-dimensional one of $(p(\alpha_1), \ldots, p(\alpha_K)) \in [0,1]^K$. We refer to this restriction as the independent DINA (ind-DINA) from hereon. It is convenient to model each α_k with a logistic link, so that

$$p(\alpha_k) = \exp(\alpha_k b_k)/[1 + \exp(b_k)],$$

where b_k denotes the attribute's 'difficulty.'

Another alternative is the higher-order DINA (HO-DINA) model (de la Torre & Douglas 2004; Templin, Henson, Templin, & Roussos 2008). This model assumes that the probability of possessing a skill is dependent on a continuous skill factor θ

following the standard normal distribution, so that

$$\pi_{\boldsymbol{\alpha}} = \int_{\theta} p(\boldsymbol{\alpha}|\theta)p(\theta)d\theta.$$

Each individual attribute is assumed to be conditionally independent given θ, so that

$$p(\boldsymbol{\alpha}|\theta) = \prod_{k=1}^{K} p(\alpha_k|\theta).$$

Finally the individual probabilities $p(\alpha_k|\theta)$ can be modeled with a logistic link,

$$p(\alpha_k|\theta) = \exp(\alpha_k(b_k + a_k\theta))/[1 + \exp(b_k + a_k\theta)],$$

where b_k denotes the attribute's 'difficulty,' and a_k is the attribute discrimination parameter. It is also possible to fit a restricted version of this model, for which all the a_k must be equal, as in de la Torre and Douglas (2004). This is referred to as the restricted higher order DINA (RHO-DINA) model (DeCarlo 2011).

2.2.3 The DINO Model

The DINO model also specifies item and attribute relationships using a Q-matrix, but it is the compensatory analog of the DINA model. Instead of an "and" relationship between the required attributes, the DINO model interprets attribute requirements with an "or" relationship. Consider applications in psychological clinical screening, where items may be symptoms of certain psychogical disorders and attributes are the disorders themselves. When observed symptoms may be the result of either (or both) disorders, the compensatory "or" relationship of the DINO model may be appropriate for modeling the item-attribute relationships. In the DINO model, the ideal responses are calculated as

$$\xi_j(Q, \boldsymbol{\alpha}) = 1 - \prod_{k=1}^{K}(1 - \alpha_k)^{q_{j,k}} = \mathbf{1}(\alpha_k = q_{j,k} = 1 \text{ for some } k).$$

As in the DINA model, the response probabilities are functions of item parameters $s_j = \Pr(X_j = 0|\xi_j = 1)$ and $g_j = \Pr(X_j = 1|\xi_j = 0)$, and the marginal response probabilities depend on the prevalence parameter $\boldsymbol{\pi}$. It is intereting to note that under the DINA model, ideal responses are *correct* when the respondent possesses all required attributes; under the DINO model, ideal responses are *incorrect* when the respondent *does not* possess all required attributes. Thus, for responses X following the DINO model, the reversed responses $1-X$ follow the DINA model, with a reversed interpretation of the attribute profile vectors.

Chapter 3

The Identifiability of Q-matrix based CDMs

3.1 Introduction

The study of identifiably dates back to Koopmans (1950); Koopmans and Reiersøl (1950). The key issue is the feasibility of recovering the model parameters based on the observed data. Identifiability is a prerequisite for statistical inferences such as parameter estimation and hypothesis testing. Moreover, it is also essential to the correct interpretation of model parameters. In this chapter, we focus on the identifiability of the parameters in cognitive diagnosis models. In particular, we propose sufficient and necessary conditions under which the slipping, guessing, and population parameters are estimable from the data under the DINA model assumption. The analysis here is based on the theoretical framework in Liu, Xu, and Ying (2012; 2013), and is generic in the sense that it can be employed for the analysis of other diagnostic classification or latent class models.

3.2 Identifiability of CDMs

Although CDMs have many attractive traits for practitioners looking to perform diagnostic classification and are a fertile area of active research, several persistent statistical issues limit their practical use. One of the most troubling is model identifiability. Identifiability is based on the idea that, in order to obtain reasonable statistical inferences, different models should correspond to different response distributions. With parametric models such as the DINA, we say that a set of parameters θ for a family of distributions $\{f(x|\theta) : \theta \in \Theta\}$ is identifiable if distinct values of the parameter θ correspond to distinct probability density functions, i.e., for any θ there is no $\tilde{\theta} \in \Theta\backslash\{\theta\}$ such that $f(x|\theta) = f(x|\tilde{\theta})$. In addition, we say that a set of parameters θ is locally identifiable if there exists a neighborhood of θ, $\mathcal{N}_\theta \in \Theta$, such that there is no $\tilde{\theta} \in \mathcal{N}_\theta\backslash\{\theta\}$ such that $f(x|\theta) = f(x|\tilde{\theta})$. Local identifiability is a weaker form of identifiability, which ensures that the model parameters are identifiable in a neighborhood of the true parameter values. Both the identifiability and the local identifiability of latent class models are well-established concepts in latent class analysis (e.g. Goodman 1974; McHugh 1956).

Definition 1 (identifiability) *A a set of parameters θ for a family of distributions $\{f(x|\theta) : \theta \in \Theta\}$ is identifiable if distinct values of θ correspond to distinct pdfs, i.e., for any θ there is no $\tilde{\theta} \in \Theta\backslash\{\theta\}$ for which $f(x|\theta) \equiv f(x|\tilde{\theta})$.*

Definition 2 (local identifiability) *We say a set of parameters θ is locally identifiable if there exists a neighborhood $\mathcal{N} \in \Theta$ such that there is no $\tilde{\theta} \in \mathcal{N}\backslash\{\theta\}$ for which $f(x|\theta) \equiv f(x|\tilde{\theta})$.*

Researchers have long been aware of the fact that Q-matrix based CDMs are generally not identifiable (DeCarlo 2011; DiBello et al. 1995; G. Maris & Bechger 2009;

C. Tatsuoka 2009; K. K. Tatsuoka 1991), though there is a tendency to gloss over the issue in practice due to a lack of theoretical development on the topic (de la Torre & Douglas 2004). Identifiability issues lead to problems in estimation and classification, and unprincipled use of standard CDMs may lead to misleading conclusions about the respondents' latent traits (G. Maris & Bechger 2009; C. Tatsuoka 2009). Let us consider the DINA model as an example. Estimation of the parameters has been studied extensively in the literature and different estimation procedures have been proposed. de la Torre (2009) uses the EM algorithm and the MCMC method to estimate the slipping and guessing parameters in the DINA model. Identifiability is important no matter what estimation method is used; it is necessary for the consistency of the EM algorithm and for the interpretability and convergence of the estimates generated by the MCMC method. However, the identifiability of the parameters in the DINA model is quite difficult to address, and in fact the necessary and sufficient conditions for identifiability are as of yet unknown.

The earliest work on the identifiability of the DINA model concerns the identifiability of ideal response vectors. This is related to the identifiability of the parameters in that, when multiple attribute profiles lead to identical ideal response vectors, it is impossible to tell responses from one profile from those from another. Then the prevalence parameter $\boldsymbol{\pi}$ is not identfiable; transfering weight between indistinguishable attribute profiles gives a set of distinct parameters associated with only one response distribution. This type of identifiability depends solely on theQ-matrix. Chiu, Douglas, and Li (2009) calls Q-matrices under which $\boldsymbol{\xi}(Q, \boldsymbol{\alpha}) \neq \boldsymbol{\xi}(Q, \boldsymbol{\alpha}')$ for all $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}'$ *complete*. The mathematical requirements on the Q-matrix for completeness are well known (Chiu et al. 2009; DeCarlo 2011; DiBello et al. 1995; K. K. Tatsuoka 1991), and we use these requirements to create a mathematical definition of completeness:

Definition 3 *A Q-matrix is said to be* complete *if* $\{\boldsymbol{e}_j^\top : j = 1, ..., K\} \subset \mathcal{R}_Q$*; otherwise, we say that* Q *is* incomplete*.*

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

3.3 The identifiability of the DINA model

Here we consider the identifiability of the DINA model under several different settings, depending on which parameters are considered known or unknown. We derive necessary and sufficient conditions for identifiability in all cases except the most difficult one, where all the parameters $\boldsymbol{s}$, $\boldsymbol{g}$, and $\boldsymbol{c}$ are unknown; in this case a small gap between the necessary and sufficient conditions yet remains, and is potential avenue of future research.

Before we begin, please note that for the rest of the chapter, we will be employing the following notation. The following list summarizes some general notation used throughout the rest of the discussion.

- For a matrix M, let M_i denote the i-th row of M. This may be extended to $M_{1:n}$, which denotes the sub-matrix containing the first n rows of M.

- Let $\mathcal{R}_M = \{M_i^\top : i = 1, \ldots, d_1\}$ denote the set of row vectors of a $d_1 \times d_2$ matrix M.

- The matrix $\mathcal{I}_d$ is the $d \times d$ identity matrix.

- The column vector $\boldsymbol{e}_i$ is a standard basis vector; its i-th element is one and the rest are zero.

- The symbols $\mathbf{0}$ and $\mathbf{1}$ denote the zero and one column vectors, i.e., $(0, ..., 0)^\top$ and $(1, ..., 1)^\top$, respectively.

- Given d-dimensional vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, let $\boldsymbol{u} \succ \boldsymbol{v}$ if the entries $u_i > v_i$ for all $i \in \{1, \ldots, d\}$. Similarly define the operations $\prec$, $\succeq$, and $\preceq$.

3.3.1 Conditions

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

(C1) Q is complete. When this holds, we assume WLOG that the Q-matrix takes the following form:

$$Q = \begin{pmatrix} \mathcal{I}_K \\ Q' \end{pmatrix}. \tag{3.1}$$

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

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

(C4) Suppose Q has the structure defined in (3.1). For each $k \in \{1, \ldots, K\}$, there must exist subsets S_k^+, S_k^- of the items in Q' such that the attribute requirements of S_k^+ and S_k^- are identical except in the k-th attribute, which is required by at least one item in S_k^+ but not by any in S_k^-. Here, the 'attribute requirements' of a set of items is the set of attributes required to have correct ideal responses to all items in that set.

Note that the null set is also a valid subset, with no required attributes. Then Condition C4 is satisfied if, for example, the Q-matrix contains two copies of the identity matrix. When

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

for every $k \in \{1, \ldots, K\}$, $S_k^+ = \{K + k\}$ requires solely the k-th attribute, while $S_k^- = \emptyset$ requires no attributes. Alternatively, letting $Q' = 1 - \mathcal{I}_K$ would also fulfill Condition C4. For the k-th attribute, let $S_k^+ = \{K + 1, \ldots, 2K\}$ be the set of all items in Q'; it requires all attributes. Let $S_k^- = \{K + k\}$ include only the k-th item in Q'; it requires all attributes except the k-th one. Thus, the attribute requirements of S_k^+ and S_k^- are identical except in the k-th dimension and Condition C4 has been fulfilled.

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

3.3.2 Theorems

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

Theorem 1 *Prevalence parameters* $\boldsymbol{\pi}$ *are identifiable only if Condition C1 is satisfied. Moreover, Condition C1 is sufficient when both the slipping and the guessing parameters are known.*

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

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

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

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

Theorem 2 *Under the DINA model with known guessing parameter* $\boldsymbol{g}$*, the slipping parameter* $\boldsymbol{s}$ *and the prevalence parameter* $\boldsymbol{\pi}$ *are identifiable if and only if Conditions C1 and C2 hold.*

Consider the Q-matrices

$$Q^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Q^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}. \tag{3.2}$$

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

In the most difficult setting, neither the slipping nor the guessing parameters are known. Then we have the following two results.

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

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

According to Theorem 3, neither Q^1 nor Q^2 from (3.2) describe identifiable DINA models when $\boldsymbol{s}$, $\boldsymbol{g}$, and $\boldsymbol{\pi}$ are all unknown. In order for the model to be identifiable, at the very least Conditions C1 and C3 must hold. Consider the following four Q-matrices where both conditions hold:

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

$$Q^5 = \begin{pmatrix} & \mathcal{I}_4 & & \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad Q^6 = \begin{pmatrix} \mathcal{I}_3 \\ \mathcal{I}_3 \\ \mathbf{1}^\top \\ \mathbf{1}^\top \end{pmatrix}.$$

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

3.4 Identifiability under attribute hierarchy

Attribute hierarchies are an active area of research in diagnostic assessment where cognitive theory is used to specify attribute dependencies within the population (Leighton, Gierl, & Hunka 2004; Rupp et al. 2010; Su, Choi, Lee, Choi, & McAninch 2013). These dependencies translate into hypotheses about the prevalence of certain profiles; to be specific, certain attribute profiles $\boldsymbol{\alpha}$ will be disallowed and $\pi_{\boldsymbol{\alpha}} = 0$. For example, suppose an educator hypothesizes that no one learns multiplication without first knowing addition. We say that addition is a *prerequisite* for multiplication, and an attribute profile denoting mastery of multiplication without mastery of addition would be excluded from the population.

The structure of attribute hierarchies can vary. The simplest structure is the linear hierarchy, where the skills are fully ordered and can be put in a sequence such that each skill cannot be mastered without mastery of the previous skill. In Figure 3.1, the linear hierarchy on the left shows that Attribute D requires Attribute C, which requires Attribute B, which requires Attribute A. A reasonable set of skills for which such a hierarchy may be constructed is the set of four simple operations of addition, subtraction, multiplication, and division. The linear hierarchy imposes a strong structure on the attributes; all attributes must be learned in the given order. In contrast, the rightmost hierarchy in Figure 3.1 represents a relatively unstructured hierarchy; although all other attributes require the first one, there are no structural requirements amongst the later attributes. Hierarchies may also be described as convergent, when multiple attributes are required by another, or divergent, when one attribute serves as prerequisite for multiple others. Of course, mixtures of these types are also possible (Leighton et al. 2004).

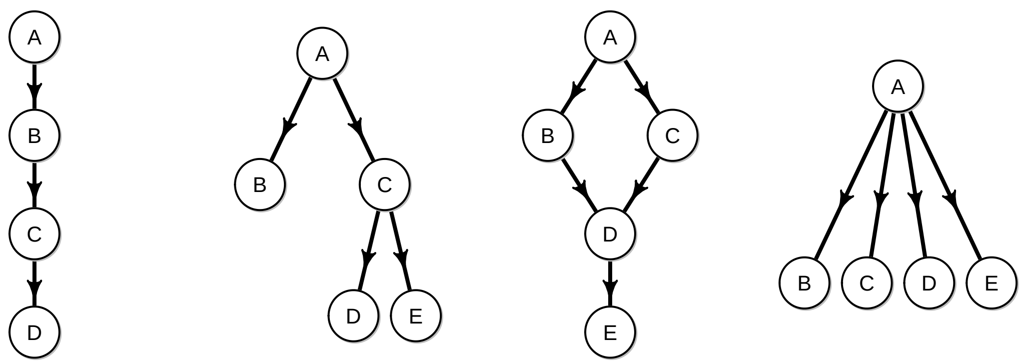


Figure 3.1: Examples of attribute hierarchies

Mathematically, an attribute hierarchy can be characterized as a list of prerequisites describing a strict partial order. When the k_1-th attribute is a prerequisite for the k_2-th attribute, we denote the relation with by $k_1 \lhd k_2$. This relation is

- irreflexive: $k \ntriangleleft k$. An attribute cannot be its own prerequisite.
- transitive: $k_1 \lhd k_2, k_2 \lhd k_3 \Rightarrow k_1 \lhd k_3$. If one attribute is a prerequisite for a second, and that second attribute is a prerequisite for a third, then the first attribute is also a prerequisite for the third attribute.

- asymmetric: $k_1 \lhd k_2 \Rightarrow k_2 \ntriangleleft k_1$. If one attribute is a prerequisite for another, the latter attribute cannot be a prerequisite for the first. Note this property can be derived from the first two properties.

Thus, it describes a strict partial order. For convenience, we also define the non-strict partial order induced by the strict ordering. Let $k_1 \unlhd k_2$ if $k_1 \lhd k_2$ or $k_1 = k_2$, i.e., if k_1 is a prerequisite for or equal to k_2.

3.4.1 Conditions

Since attribute hierarchies reduce the parameter space by disallowing certain profiles, and identifiability in Q-matrix based CDMs is often an issue of making sure that distinct attribute profiles result in distinct responses, one may suspect that having an attribute hierarchy may allow a relaxation of the identifiability conditions. This is indeed the case. We have necessary and sufficient conditions for identifiability both when the item parameters are known and when only $\boldsymbol{g}$ is known. We will use the following two conditions:

(D1) Every attribute is measured by at least one item that measures only that attribute and, possibly, a subset of its prerequisites. Mathematically, for each $k \in \{1, \ldots, K\}$, there exists $j \in \{1, \ldots, J\}$ s.t. $q_{j,k} = 1$ and $q_{j,k'} = 1 \Rightarrow k' \unlhd k$. When this holds, we assume WLOG that the k-th item fulfills the condition for the k-th attribute.

(D2) Each terminal attribute in the hierarchy is required by at least two items. Here, a *terminal* attribute is one which does not serve as a prerequisite for any others. Mathematically, for every $k \in \{1, \ldots, K\}$ s.t. $\{k' : k \lhd k'\} = \emptyset$, $\sum_{j=1}^{J} q_{j,k} \geq 2$.

Conditions D1 and D2 are analogs of Conditions C1 and C2, respectively. We are able to relax both of the original conditions. In the case of Condition D1, it is no longer necessary to have items measuring each attribute in isolation. When attributes have prerequisites, these attributes can be measured together with their prerequisites in an item that helps fulfill Condition D1. In the case of Condition D2, only terminal attributes need to be measured by two items. In a linear hierarchy, for example, that would be only the final attribute. Note that when there is no hierarchy, the new conditions are equivalent ot the old ones. Since no attributes have prerequisites, Condition D1 is equivalent to completeness, and since every attribute is terminal, Condition D2 is equivalent to Condition C2.

3.4.2 Theorems

The corresponding identifiability theorems are as follows, and are analogs of the identifiability theorems in Section 3.3.2. Recall that the default assumptions now are that $\pi_{\boldsymbol{\alpha}} > 0$ for all attribute profiles *allowed by the attribute hierarchy* and that $\boldsymbol{c} > \boldsymbol{g}$.

Theorem 5 *Consider the DINA model with an attribute hierachy limiting the set of possible attribute profiles. Prevalence parameters* $\boldsymbol{\pi}$ *are identifiable only if Condition D1 is satisfied. Moreover, Condition D1 is sufficient when both the slipping and the guessing parameters are known.*

Theorem 6 *Consider the DINA model with an attribute hierarchy and known guessing parameter* $\boldsymbol{g}$*. The slipping parameter* $\boldsymbol{s}$ *and the prevalence parameter* $\boldsymbol{\pi}$ *are identifiable if and only if Conditions D1 and D2 hold.*

These relaxed requirements can be very useful in assessment design. For instance, it can sometimes be difficult to create items measuring advanced skills without measuring any basic skills.

3.5 Simulations

In this section, we conduct simulation studies to illustrate the results in Section 3.3. We generate data from the DINA model under different Q-matrices and check the parameter estimators $\hat{\boldsymbol{s}}$, $\hat{\boldsymbol{g}}$, and $\hat{\boldsymbol{\pi}}$. All Q-matrices are designs for $K = 3$ attributes. In total there are three different simulation settings, each of which is detailed below:

Setting A We begin with the complete Q-matrix $Q^A = \mathcal{I}_5$. This is the minimal Q-matrix necessary for completeness. The item parameters are

$$\boldsymbol{s} = \begin{pmatrix} 0.08, & 0.15, & 0.23, & 0.25, & 0.20 \end{pmatrix}^\top$$

and

$$\boldsymbol{g} = \begin{pmatrix} 0.08, & 0.22, & 0.16, & 0.19, & 0.14 \end{pmatrix}^\top .$$

The population parameter is

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

Setting B The second Q matrix is

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

It fulfills both Condition C1 and Condition C2. The population parameter is the same as in Setting 1, as are the item parameters for the first five items. The sixth item has item parameters $s_6 = 0.18$ and $g_6 = 0.11$.

Setting C The third Q-matrix is

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

It fulfills Conditions C1, C3, and C4. The population parameter is the same as in Setting 1, as are the item parameters for the first five items. The last five items have slipping and guessing parameters

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

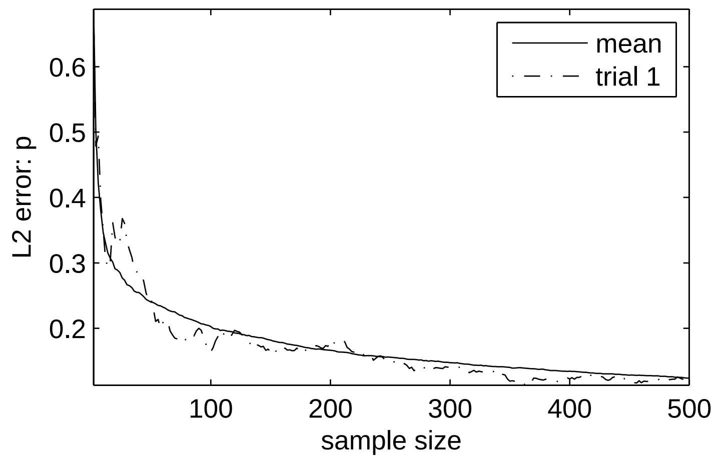
and

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

For examples of the non-identifiability of $\boldsymbol{\pi}$ when $\boldsymbol{s}$ and $\boldsymbol{g}$ are known, we refer to the simulations conducted on incomplete Q-matrices in Zhang et al. (2013). By Theorem 1, the model in Setting 1 is identifiable if $\boldsymbol{s}$ and $\boldsymbol{g}$ are known. Figure 3.2 plots the L^2 error of the maximum likelihood estimates of $\boldsymbol{\pi}$ as the sample size N grows. The L^2 error is the Euclidean distance between the estimate and the true value. The consistency of the estimator $\hat{\boldsymbol{\pi}}$ is an important consequence of the identifiability of the model.

As stated in Theorem 2, Q^A is associated with a model that is non-identifiable when only $\boldsymbol{g}$ is known. Then, even for very large sample sizes parameter estimates will converge to the truth. This occurs because multiple sets of parameters maximize the same marginal likelihood. Shown below are two distinct sets of parameters that both maximize the likelihood for a random response matrix generated under Setting A for 500,000 individuals. They were obtained by running multiple repetitions of the

Figure 3.2: L^2 error for estimates of $\boldsymbol{\pi}$ under Setting A, when $\boldsymbol{s}$ and $\boldsymbol{g}$ are known.



EM-algorithm in order to obtain true global maximizers. For the first set of estimates,

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

and

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

In the second set,

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

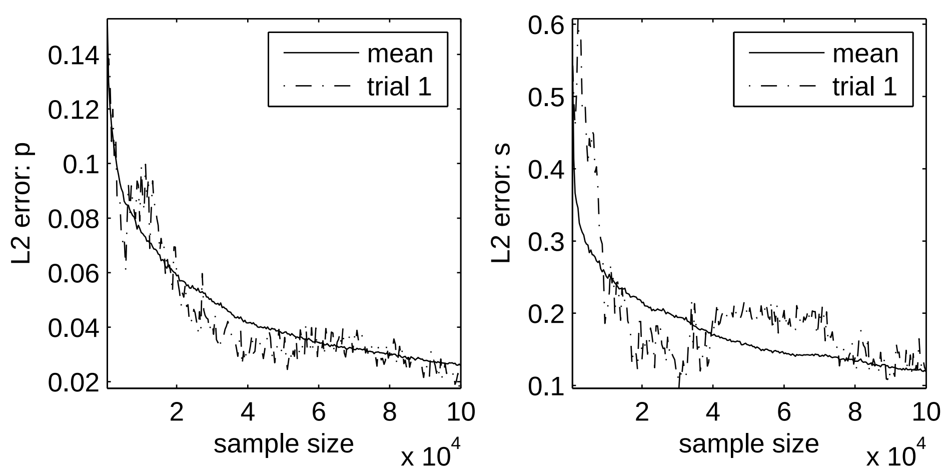
and

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

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

The DINA model is identifiable when only $\boldsymbol{g}$ is known only if Condition C2 is filled, in addition to Condition C1. The Q-matrix Q^B from Setting B fulfills both conditions, and when $\boldsymbol{g}$ is known, the maximum likelihood estimates of both $\boldsymbol{s}$ and $\boldsymbol{\pi}$ will be consistent. The L^2 error of both estimates is plotted in Figure 3.3. When

Figure 3.3: L^2 error for estimates of $\boldsymbol{\pi}$ (left) and $\boldsymbol{s}$ (right) under Setting B, with $\boldsymbol{g}$ known.



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

$$\begin{aligned}\hat{\boldsymbol{\pi}} \;=\; (&0.017, & 0.019, & 0.017, & 0.015, & 0.048, & 0.019, & 0.068, & 0.006,\\ &0.060, & 0.014, & 0.045, & 0.015, & 0.067, & 0.057, & 0.124, & 0.019,\\ &0.022, & 0.039, & 0.012, & 0.010, & 0.048, & 0.006, & 0.011, & 0.018,\\ &0.025, & 0.037, & 0.016, & 0.011, & 0.114, & 0.008, & 0.001, & 0.013)^\top\end{aligned}$$

$$\hat{\boldsymbol{s}} = \Big(0.15,\ \ 0.01,\ \ 0.08,\ \ 0.28,\ \ 0.28,\ \ 0.11\Big)^\top.$$

and

$$\hat{\boldsymbol{g}} = \Big(0.09,\ \ 0.17,\ \ 0.25,\ \ 0.26,\ \ 0.20,\ \ 0.22\Big)^\top.$$

In the second set,

$$\begin{aligned}\hat{\boldsymbol{\pi}} \;=\; (&0.016, & 0.034, & 0.032, & 0.061, & 0.026, & 0.020, & 0.074, & 0.035,\\ &0.029, & 0.014, & 0.045, & 0.028, & 0.004, & 0.028, & 0.041, & 0.043,\\ &0.026, & 0.079, & 0.038, & 0.083, & 0.032, & 0.000, & 0.032, & 0.028,\\ &0.012, & 0.017, & 0.019, & 0.023, & 0.037, & 0.019, & 0.018, & 0.010)^\top\end{aligned}$$

$$\hat{\boldsymbol{s}} = \Big(0.04,\ \ 0.33,\ \ 0.30,\ \ 0.04,\ \ 0.04,\ \ 0.11\Big)^\top.$$

and

$$\hat{\boldsymbol{g}} = \Big(0.09,\ \ 0.17,\ \ 0.26,\ \ 0.26,\ \ 0.21,\ \ 0.00\Big)^\top.$$

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

When the Q-matrix fulfills Condition C4, parameters can be estimated consistently, as shown in the last set of simulations. The Q-matrix Q^C from Setting C fulfills Condition C4, in addition to Conditions C1 and C3. The plots in Figure 3.4 show the convergence of the maximum likelihood estimates of $\boldsymbol{\pi}$, $\boldsymbol{s}$, and $\boldsymbol{g}$ under Setting C.

3.6 Discussion

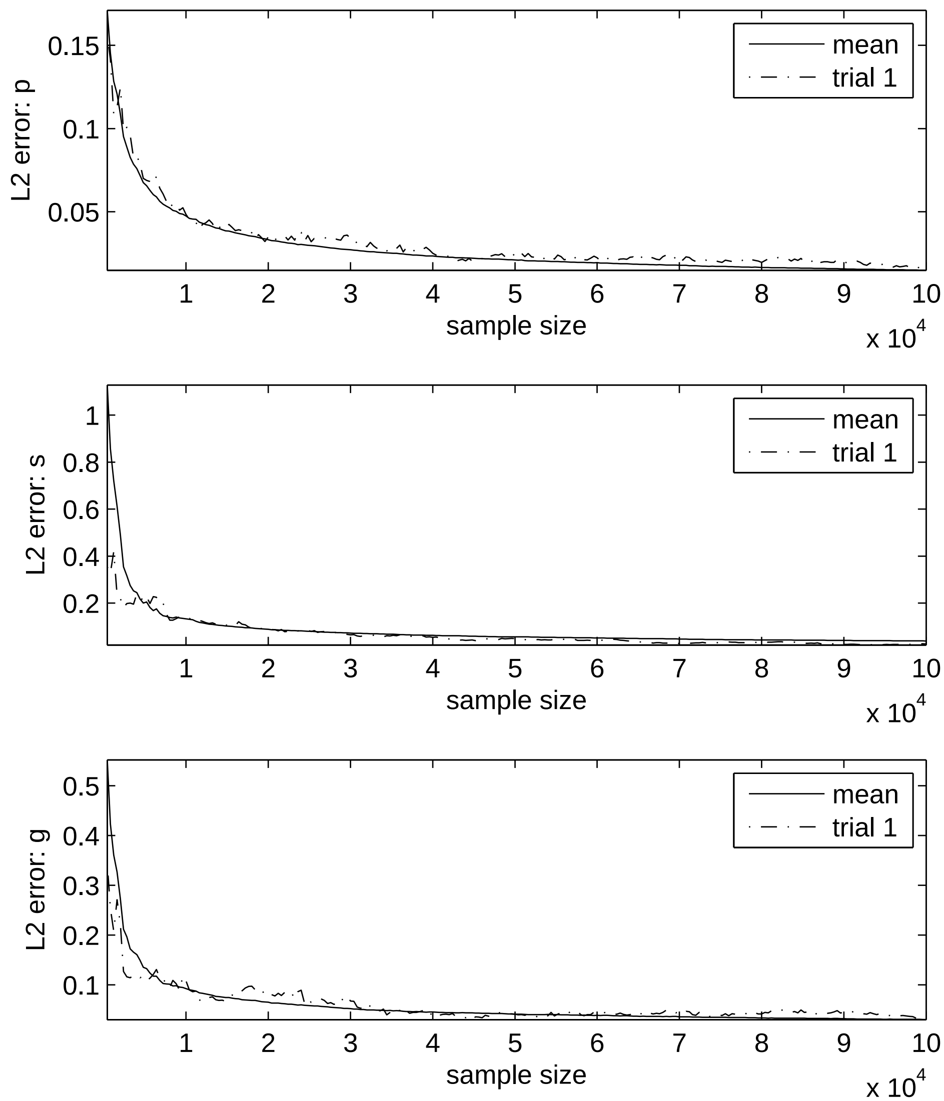
Identification is a serious problem in cognitive diagnosis modeling; in almost all statistical analyses,identifiability is essential in ensuring proper statistical inference and interpretable parameters. The work here lays out some relatively simple conditions to check the identifiability of the DINA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

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

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

Work on identifiability conditions is helpful in pointing out potentially unknown problems with statistical inference. It is also useful as a guide for designers of diagnostic assessments, who should attempt to fulfill identifiability conditions if possible. However, non-identifiability in Q-matrix based CDMs is a common issue not simply because of a lack of precise identifiability conditions. Obtaining an assessment that fulfills even the basic completeness condition may be difficult, either because of lack of control over the design or practical issues in creating items that fulfill the condition. Although attribute hierarchies may relax the identifiability conditions, a well-developed cognitive theory upon which to base the hierarchy may not be available, or the researcher may be reluctant to make such assumptions. In the next chapter, we move on to a somewhat unusual problem in statistics, as we consider the issue of statistical inference for non-identifiable models.

Figure 3.4: L^2 error for estimates of $\boldsymbol{\pi}$ (top), $\boldsymbol{s}$ (middle), and $\boldsymbol{g}$ (bottom) under Setting C, when all parameters are unknown.



Chapter 4

Methods for non-identifiable models

As seen in the previous chapter, under the DINA model, Q-matrices where some attributes are required by items solely in conjunction with other attributes lead to identifiability issues (DeCarlo 2011; C. Tatsuoka 2009). Unfortunately, the issue is quite well-spread. Even well-studied datasets like Tatsuoka's fraction subtraction dataset lack identifiability, and the consequences include analyses where respondents who answer all items incorrectly are classified as having most of the skills (DeCarlo 2011). Some of this is a result of a lack of awareness of the precise conditions needed for the identifiability of the DINA model. Furthermore, researchers analyzing datasets oftentimes have no control over the design of the assessment and must make do with what they have. This issue becomes especially severe when diagnostic models are applied to assessments that were not meant to be diagnostic in the first place; see Su et al. (2013)'s Q-matrix for the 2003 Trends in International Mathematics and Science Study (TIMSS) for an example. Even when those involved in assessment design wish to follow identifiability conditions, this may be practically infeasible. For instance, it is often difficult to measure mathematical skills without simultaneously measuring some basic ones. In addition, items measuring only one attribute at the time, though desirable from an identifiability standpoint, require sacrifices in other areas of concern in assessment design, such as efficiency and test length.

In this chapter we present contributions to the field of statistical inference under conditions of non-identifiability for Q-matrix based CDMs. We propose methods for both parameter estimation and respondent classification. In addition, we are able to quantify the severity of the identifiability issue in terms of the proportion of the population affected by it. We conclude with some more nuanced guidelines for assessment design. Together, the ability to measure the severity of non-identifiability and a full understanding of the mechanism behind the problem suggest that, rather than insist on fulfilling identifiability conditions, it may be possible to balance identifiability concerns with other design issues.

4.1 Completeness and identifiability

We begin with a deeper examination of the intuition behind the why a complete Q-matrix is necessary for identifiability. Diagnostic assessments are meant to provide detailed information about respondents' possession of a variety of traits. Preferably, a well-designed exam will be able to provide information about each trait for every respondent. However, recovering information about the latent variables from a '0' response may be difficult when an item measures too many attributes simultaneously; in comparison to a '1' response, which suggests that a respondent is more likely to possess each attribute associated with that item, a '0' response may indicate the failure to master any one or several of the required attributes. Consider the following two simple Q-matrices for the DINA model:

$$Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; Q_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}. \tag{4.1}$$

In assessments based on the Q-matrix Q_1, a correct response to each item generally indicates a higher probability that the respondent possesses the corresponding attribute, while an incorrect response indicates a lower probability of the same. However, with Q_2, an incorrect response to the second item only implies that at least one of the attributes is probably missing. In fact, given that a student does not possess Attribute 1, Item 2 provides no information about his or her mastery of Attribute 2, and so respondents with attribute profiles $(0, 0)$ and $(0, 1)$ have statistically identical responses. Thus, the assessment as a whole is incapable of differentiating between the two profiles, and any classification decision between them will solely be a reflection of the prior information.

A slightly more complicated situation appears if we add a third attribute to the example above. Consider an assessment following the DINA model with Q-matrix Q_3, where

$$Q_3 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}. \tag{4.2}$$

The attribute requirements of the first two items match those of the items corresponding to Q_2. Now, however, the proportion of individuals for whom Attribute 2 is not identifiable is smaller. Of those who do not possess Attribute 1, some will possess Attribute 3. Then Attribute 2 is identifiable because of differing response distributions on Item 3. However, response distributions for those with attribute profiles $(0, 1, 0)$ and $(0, 0, 0)$ are still indistinguishable. Thus, although the assessment provides no information about Attribute 2 for a smaller part of the population, the issue has not been completely resolved.

4.1.1 Partitioning the Attribute Profile Space

We begin with an intuitive criterion for deciding whether an assessment has the ability to differentiate between two attribute profiles.

Definition 4 *Two attribute profiles are* separable *if they lead to different response distributions.*

The differing response distributions of separable attribute profiles imply that the data will favor one profile or the other; there is some differential effect on the likelihood and thus the posterior. Profiles that are not separable are statistically identical, with the same likelihood functions, making any differences in their posteriors simply artifacts of the prior.

Determining whether attributes are separable can be done without the full response distribution;in fact, only the ideal responses $\boldsymbol{\xi}(Q, \boldsymbol{\alpha})$ are necessary.

Proposition 7 *Given a Q-matrix Q and slipping and guessing parameters $\boldsymbol{s}$ and $\boldsymbol{g}$, two attribute profiles $\boldsymbol{\alpha}^1$ and $\boldsymbol{\alpha}^2$ can be separated if and only if they produce ideal response vectors $\boldsymbol{\xi}^1 = \boldsymbol{\xi}(Q, \boldsymbol{\alpha}^1)$ and $\boldsymbol{\xi}^2 = (Q, \boldsymbol{\alpha}^2)$ such that for some $j \in \{1, \ldots, J\}$, $\boldsymbol{\xi}_j^1 \neq \boldsymbol{\xi}_j^2$ and $1 - s_j \neq g_j$.*

Throughout the rest of this paper we assume that $1 - s_j \neq g_j$ for each $j = 1, \ldots, J$, which simplifies Proposition 7 into Corollary 8. Should such an item indeed be present, then it has no discriminating power and may be omitted.

Corollary 8 *If every item j has different success probabilities given $\xi_j = 1$ or given $\xi_j = 0$, i.e. $1 - s_j \neq g_j$ for $j = 1, \ldots, J$, then two attribute profiles can be separated if and only if they produce different ideal response vectors.*

Lastly, it is also of interest whether an attribute profile can be separated from all other attribute profiles, and is thus identifiable. This definition of identifiability will be tied to the general statistical concept in Section 4.2.

Definition 5 *An attribute profile $\boldsymbol{\alpha}$ is* identifiable *when it can be separated from any other attribute profile $\boldsymbol{\alpha}' \neq \boldsymbol{\alpha}$.*

4.1.1.1 Complete Separation of Attribute Profiles

The first step in understanding the identifiability issue is determining under what circumstances all attribute profiles are identifiable. This depends on the Q-matrix, which is called complete when it each attribute profile produces a distinct ideal response vector Chiu et al. (2009). Formally, we have the following definition:

Definition 6 *(Chiu et al. 2009, Definition 1) Under a* complete *Q-matrix, all attribute profiles are identifiable, i.e. $\boldsymbol{\xi}(Q, \boldsymbol{\alpha}) \neq \boldsymbol{\xi}(Q, \boldsymbol{\alpha}')$ iff $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}'$.*

The requirements for completeness have long been known Chiu et al. (2009); DiBello et al. (1995); C. Tatsuoka (2009); K. K. Tatsuoka (1991). In essence, the assessment must contain at least one item devoted solely to each attribute. In terms of the Q-matrix, this means that for each $k \in \{1, \ldots, K\}$, there should be at least one row with an entry of '1' solely in the k-th position.

Proposition 9 *(Chiu et al. 2009, Lemma 1) Let X_Q be the set of row vectors of Q-matrix Q. Then Q is* complete *iff $\{e_k : k = 1, \ldots, K\} \subset X_Q$, where e_k is a vector such that the k-th element is one and all other elements are zero.*

4.1.1.2 Partial Separation of Attribute Profiles

While a complete Q-matrix is necessary for identifiability, the requirement can be quite onerous, and many of the Q-matrices used in practice are unfortunately incomplete. This is especially true under circumstances where the researcher has no control over the design of the assessment; consider the typical Q-matrix that results from attempts to retrofit cognitive diagnosis models onto data from the usual standardized tests Su et al. (2013). Furthermore, in certain cases designing items measuring only one attribute may be infeasible; this will generally occur for more 'advanced' skills, which may be difficult to dissociate from the basic skills they build on.

The partition is a standard mathematical construct that separates a set of objects into groups of 'equivalent' objects. It is a natural tool for exploring latent structures in which multiple latent classes produce the same response distribution; Goodman (1974); C. Tatsuoka (1996; 2009); C. Tatsuoka and Ferguson (2003) have also discussed the use of partitions in similar settings.

Before forming a partition, one must define the notion of equivalence. Recall the standard mathematical notion of the equivalence relation, a relation '$\sim$' that is reflexive ($a \sim a$), symmetric ($a \sim b$ iff $b \sim a$), and transitive ($a \sim b$ and $b \sim c \Rightarrow a \sim c$).

Proposition 10 *Let '$\sim$' denote the binary relation 'cannot be separated,' where $\boldsymbol{\alpha}^1 \sim \boldsymbol{\alpha}^2$ when $\boldsymbol{\xi}(Q, \boldsymbol{\alpha}^1) = \boldsymbol{\xi}(Q, \boldsymbol{\alpha}^2)$. Then '$\sim$' is an equivalence relation.*

Putting profiles into groups, commonly known as equivalence classes, based on an equivalence relation results in a partition; in this case, any two attribute profiles in the same equivalence class cannot be separated, while any two in different classes can be. We denote a particular equivalence class by $[\boldsymbol{\alpha}]$, where $\boldsymbol{\alpha}$ may be any attribute profile in the class; literally, $[\boldsymbol{\alpha}]$ can be read as "the set of attribute profiles equivalent to $\boldsymbol{\alpha}$."

The most direct way of determining the partition would be to calculate the ideal response vector of each of the 2^K attribute profiles and then sort them lexicographically. Both steps are simple and run quickly; see Table 4.1 for the detailed algorithm. C. Tatsuoka's algorithm performs the same task in an iterative fashion, refining the partition as it considers the ideal responses to each item C. Tatsuoka (1996); C. Tatsuoka and Ferguson (2003). For another alternative based in Boolean algebra, see K. K. Tatsuoka (1991).

Table 4.1: Algorithm for partitioning an attribute profile space

Step	Procedure
Input:	A $J \times K$ Q-matrix Q.
(0)	(optional) Remove items with duplicate attribute requirements
(1)	List all 2^K attribute profiles $\boldsymbol{\alpha}$.
(2)	Find the ideal response vector $\boldsymbol{\xi}(Q, \boldsymbol{\alpha})$ for each $\boldsymbol{\alpha}$.
(3)	Do a lexicographic (alphabetic) sort of the ideal response vectors.
(4)	Check whether each successive profile has the same ideal response vector as the previous profile. If so, $\boldsymbol{\alpha}$ is the first member of a new equivalence class $[\boldsymbol{\alpha}]$. Else, $\boldsymbol{\alpha}$ is part of the current equivalence class.
Output:	A list of equivalence classes $[\boldsymbol{\alpha}]$ and their members.

Note that our algorithm results in equivalence classes labeled by their smallest member, which shall be called the *minimal representative*. The minimal representative has additional meaning as the attribute requirements of the corresponding ideal response vector and is therefore convenient label for each equivalence class.

As seen in Table 4.2, performing the algorithm on the 3×3 Q-matrix Q_3 from

(4.2) results in five different equivalence classes, each of which is labeled with by its minimal representative: $[000] = \{000, 010, 001\}$, $[011] = \{011\}$, $[100] = \{100, 101\}$, $[110] = \{110\}$, and $[111] = \{111\}$. Note that since the bracket notation may be read as 'the equivalence class containing,' it is possible to change the labeling of each equivalence class by choosing any other member as the titular profile: [000], [010], and [001] all refer to the same equivalence class, for example.

Table 4.2: Generating the partition associated with the Q-matrix Q_3.

Q_3		$\boldsymbol{\alpha}$	$\boldsymbol{\xi}(Q_3, \boldsymbol{\alpha})$		$\boldsymbol{\alpha}$	$\boldsymbol{\xi}(Q_3, \boldsymbol{\alpha})$		$[\boldsymbol{\alpha}]$
		000	000		000	000		[000]
		100	100		010	000		
		010	000		001	000		
$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$	$\xrightarrow{(1),(2)}$	001	000	$\xrightarrow{(3)}$	011	001	$\xrightarrow{(4)}$	[011]
		110	110		100	100		[100]
		101	100		101	100		
		011	001		110	110		[110]
		111	111		111	111		[111]

Steps from Table 4.1 labeled (1), (2), (3), and (4).

4.2 Parameter estimation

We now consider the problem of parameter estimation, specifically that of $\pi_{\boldsymbol{\alpha}}$, the proportion of the population possessing each attribute profile $\boldsymbol{\alpha}$. Unless $\boldsymbol{\pi}$ is assumed known, its consistent estimation has important consequences for respondent classification and exam validity. Unfortunately, when an assessment's Q-matrix is incomplete, it is impossible to consistently estimate $\boldsymbol{\pi}$.

Essentially, non-identifiability is the problem of distinct parameters mapping to identical likelihoods. A natural solution to the problem of a many-to-one mapping is to turn each "many" into a "one" by way of equivalence relation. Suppose two distinct parameters are equivalent $\theta \sim \theta'$ if their likelihoods are equal $L(\theta) \equiv L(\theta')$. The equivalence classes $[\theta]$ map to likelihoods $L([\theta])$ in a one-to-one fashion, solving the identifiability issue. However, in most problems, the mapping from the parameter space to the likelihood is very complicated, and the reverse mapping is quite intractable. The equivalence classes $[\theta]$ can be difficult to calculate and impossible to interpret. Fortunately, this is not the case for our specific problem. For each equivalence class $[\boldsymbol{\alpha}]$, let $\pi_{[\boldsymbol{\alpha}]}$ be the proportion of the population possessing an attribute profile within that equivalence class. Then,

$$\pi_{[\boldsymbol{\alpha}]} = \sum_{\boldsymbol{\alpha}' \in [\boldsymbol{\alpha}]} \pi_{\boldsymbol{\alpha}'}, \tag{4.3}$$

i.e., the probability of a respondent belonging to a particular equivalence class is the sum of probabilities for each attribute profile within that class. We will discover that the desired equivalence classes on the prevalence parameter have a natural interpretation:

$$[\boldsymbol{\pi}] = \{\boldsymbol{\pi}' : \boldsymbol{\pi}_{[\boldsymbol{\alpha}]} = \boldsymbol{\pi}'_{[\boldsymbol{\alpha}]} \forall [\boldsymbol{\alpha}]\}.$$

Then estimating the problem of estimating the $\pi_{\boldsymbol{\alpha}}$, which is impossible, has become the problem of estimating $[\boldsymbol{\pi}]$, which is intractable, and that has become the problem of estimating the $\pi_{[\boldsymbol{\alpha}]}$, which is very simple.

It is very easy to see that one can do no better than estimating the $\pi_{[\boldsymbol{\alpha}]}$. The probability of observing any particular set of data depends only on $\pi_{[\boldsymbol{\alpha}]}$, since the probability of any response depends only on equivalence class membership, not on the respondent's possession of a specific profile. With an incomplete Q-matrix it is possible to observe populations with different distributions $\boldsymbol{\pi}^1 \not\equiv \boldsymbol{\pi}^2$ over the attribute profile space that have identical distributions over the equivalence classes $[\boldsymbol{\alpha}]$, i.e., $p^1_{[\boldsymbol{\alpha}]} = p^2_{[\boldsymbol{\alpha}]}$ for all $\boldsymbol{\alpha} \in \{0,1\}^K$, and thus identical response distributions. The phenomenon where different parameter values lead to identical response distributions is generally known as non-identifiability, and it destroys the ability of likelihood-based estimation methods to achieve consistency.

While consistent estimation of $\pi_{\boldsymbol{\alpha}}$ cannot be achieved, it is possible to consistently estimate the proportion of individuals within each equivalence class $[\boldsymbol{\alpha}]$.

Theorem 11 *Suppose an assessment follows the DINA model, with known Q-matrix Q and item parameters $\boldsymbol{s}$ and $\boldsymbol{g}$. Let $\pi_{[\boldsymbol{\alpha}]}$, representing the proportion of the population possessing an attribute profile $\boldsymbol{\alpha}' \in [\boldsymbol{\alpha}]$, be defined as in (4.3), and let the population parameter $\boldsymbol{\pi}$ be the vector of all $\pi_{[\boldsymbol{\alpha}]}$. We may write its likelihood as*

$$L(\boldsymbol{\pi}) = p(X|\boldsymbol{\pi}) = \prod_{i=1}^{N} p(\boldsymbol{x}^i|\boldsymbol{\pi}) = \prod_{i=1}^{N} \sum_{[\boldsymbol{\alpha}]} p(\boldsymbol{x}^i|[\boldsymbol{\alpha}])\pi_{[\boldsymbol{\alpha}]}.$$

Then the maximum likelihood estimate $\hat{\boldsymbol{\pi}}$ of $\boldsymbol{\pi}$ is consistent as $N \to \infty$.

Consistent estimation of the $\pi_{[\boldsymbol{\alpha}]}$ is an important result, justifying the results of the following sections. To emphasize the differences in parameter space and procedure, work based on equivalence classes $[\boldsymbol{\alpha}]$ rather than profiles $\boldsymbol{\alpha}$ will from hereon be referred to under the name of the Non-Identifiability Adjusted DINA (NIAD-DINA) model.

4.3 Respondent classification

Non-identifiability has potentially serious effects on respondent classification. Classification is generally conducted based on the posterior distribution using the Bayes rule, $p(\boldsymbol{\alpha}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{\alpha})\pi_{\boldsymbol{\alpha}}$. Under L_2 loss, where we sum the squared errors in our estimate $\hat{\boldsymbol{\alpha}}$ across the K dimensions, the optimal estimator is the posterior mean $E[\alpha_k|\boldsymbol{x}]$.

When $\pi_{\boldsymbol{\alpha}}$ is unknown and must be estimated, the classification algorithm becomes empirical Bayes. The effect of non-identifiability then becomes two-fold. First, flatness in the likelihood increases dependence on the prior. In the DINA model, profiles in the same equivalence class have the same likelihood. Thus, within the equivalence class, the posterior will simply be a reflection of the prior, without any added information from the data DeCarlo (2011). Given a prior $p(\boldsymbol{\alpha})$, for any $\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2 \in [\boldsymbol{\alpha}]$,

$$\frac{p(\boldsymbol{\alpha}^1|\boldsymbol{x})}{p(\boldsymbol{\alpha}^2|\boldsymbol{x})} = \frac{p(\boldsymbol{x}|\boldsymbol{\alpha}^1|)p(\boldsymbol{\alpha}^1)}{p(\boldsymbol{x}|\boldsymbol{\alpha}^2)p(\boldsymbol{\alpha}^2)} = \frac{p(\boldsymbol{x}|[\boldsymbol{\alpha}])p(\boldsymbol{\alpha}^1)}{p(\boldsymbol{x}|[\boldsymbol{\alpha}])p(\boldsymbol{\alpha}^2)} = \frac{p(\boldsymbol{\alpha}^1)}{p(\boldsymbol{\alpha}^2)},$$

and the posterior ratio between attribute profiles is identical to the prior one. Second, the prior $\boldsymbol{\pi}$ that has now become extremely important can no longer be consistently estimated. In combination, the double effect of non-identifiability on empirical Bayes classification completely destroys the reliability of our classifier.

Despite the serious effects of non-identifiability on our ability to classify respondents under the DINA model, it is still possible to classify respondents when the Q-matrix describes a non-identifiable model. We describe such a classification method below.

4.3.1 Marginal Separability

Before arriving at the classification algorithm, let us first extend the concept of identifiability, which has heretofore been focused on the multidimensional attribute profiles, to each attribute individually. This is motivated by the marginal nature of the L_2 loss function that penalizes each incorrect attribute separately, and the fact that, though the presence of multiple profiles in the same equivalence class signals non-identifiability, some individual attributes may still be identifiable within the class. To illustrate, consider the the Q-matrix Q_3 from (4.2) and one of its equivalence classes, $[000] = \{000, 010, 001\}$. If a profile $\boldsymbol{\alpha} \in [000]$, then its first component $\alpha_1 = 0$, but the values of α_2 and α_3 are uncertain. Thus, posterior weight $p([000]|x)$ on this class counts as positive evidence that $\alpha_1 = 0$, but does not help in deciding α_2 or α_3. This observation motivates the following definition:

Definition 7 *An attribute is* marginally separable *within an equivalence class when either all members of that class possess that attribute or none of them do.*

Define the marginal separability indicator $\delta_{[\boldsymbol{\alpha}],k}$ as follows:

$$\delta_{[\alpha],k} = \prod_{\boldsymbol{\alpha}' \in [\boldsymbol{\alpha}]} \alpha'_k + \prod_{\boldsymbol{\alpha}' \in [\boldsymbol{\alpha}]} (1 - \alpha'_k). \tag{4.4}$$

Then, $\delta_{[\boldsymbol{\alpha}],k} = 1$ when Attribute k is marginally separable within equivalence class $[\boldsymbol{\alpha}]$. Posterior weight on a class $[\boldsymbol{\alpha}]$ only provides information about the k-th attribute when $\delta_{[\boldsymbol{\alpha}],k} = 1$; otherwise, there is no information beyond the prior.

4.3.2 Classification algorithm

In empirical Bayes for the DINA model, posteriors are often calculated by maximizing the marginal maximum likelihood $L(\boldsymbol{\pi}, \boldsymbol{s}, \boldsymbol{g})$ via the E-M algorithm de la Torre (2009); Haertel (1989); Rupp et al. (2010). Then, since all vectors $\boldsymbol{\pi}$ with identical weights on each class $\pi_{[\boldsymbol{\alpha}]}$ have identical likelihoods, any $\boldsymbol{\pi}$ achieving the maximizing $\pi_{[\boldsymbol{\alpha}]}$ may result. The values chosen are determined by the starting values, which have little validity for classification.

Since the posterior is sensitive to the prior, it is important to work with $p([\boldsymbol{\alpha}])$, which can be estimated consistently, rather than $p(\boldsymbol{\alpha})$. Thus classification here will be conducted based on $p([\boldsymbol{\alpha}]|\boldsymbol{x} \propto p(\boldsymbol{x}|[\boldsymbol{\alpha}])p([\boldsymbol{\alpha}])$ instead of the usual posterior. This calculation does not require a separate fitting of the model, since

$$p([\boldsymbol{\alpha}]|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|[\boldsymbol{\alpha}])\pi_{[\boldsymbol{\alpha}]}}{p(\boldsymbol{x})} = \frac{p(\boldsymbol{x}|[\boldsymbol{\alpha}])\sum_{\boldsymbol{\alpha}'\in[\boldsymbol{\alpha}]}\pi_{\boldsymbol{\alpha}'}}{p(\boldsymbol{x})} = \sum_{\boldsymbol{\alpha}'\in[\boldsymbol{\alpha}]} p(\boldsymbol{\alpha}'|\boldsymbol{x}) \tag{4.5}$$

From this posterior, we then define

$$\pi_k^{\min}(\boldsymbol{x}) = \sum_{[\boldsymbol{\alpha}]:\alpha_k=1, d_{[\boldsymbol{\alpha}],k}=1} p([\boldsymbol{\alpha}]|\boldsymbol{x}), \tag{4.6}$$

$$\pi_k^{\max}(\boldsymbol{x}) = \pi_k^{\min}(\boldsymbol{x}) + \sum_{[\boldsymbol{\alpha}]:d_{[\boldsymbol{\alpha}],k}=0} p([\boldsymbol{\alpha}]|\boldsymbol{x}), \tag{4.7}$$

where $\delta_{[\boldsymbol{\alpha}],k}$ is the marginal separability indicator defined in (4.4). Classification follows from the fact that, depending upon the specific hyperprior on $\boldsymbol{\pi}$ or starting point of the E-M algorithm, the DINA model may produce marginal posterior probabilities of mastery $\Pr(\alpha_k = 1|\boldsymbol{x})$ anywhere in the range $[\pi_k^{\min}(\boldsymbol{x}), \pi_k^{\max}(\boldsymbol{x})]$. Thus, it is only appropriate to conclude that $\alpha_k = 1$ when $\pi_k^{\min}(\boldsymbol{x})$ is high, or that $\alpha_k = 0$ when $\pi_k^{\max}(\boldsymbol{x})$ is low. A natural cutoff for both is 0.5, but it may be adjusted as necessary, depending on the amount of classification error that can be tolerated. This classification method, from hereon referred to as the NIAD-DINA classification algorithm, accounts for both uncertainty in the prior and uncertainty caused by slipping and guessing. It is a conservative algorithm that identifies for which attributes a respondent may be misclassified due to non-identifiability by the usual classification methods and instead leaves the respondent unclassified for those attributes. The algorithm is summarized in Table 4.3.

Table 4.3: NIAD-DINA classification algorithm

Step	Procedure	q.v.
Input:	Q-matrix $Q = (q_{j,k})_{J\times K}$, data $X = (x_{i,k})_{N\times J}$.	
(1)	Fit the model to produce $p(\boldsymbol{\alpha}\|\boldsymbol{x})$.	
(2)	Partition the attribute profile space.	Table 4.1
(3)	Calculate the marginal separability vector $\delta_{[\boldsymbol{\alpha}]}$.	(4.4)
(4)	Sum posteriors $p(\boldsymbol{\alpha}\|x)$ for $p([\boldsymbol{\alpha}]\|\boldsymbol{x})$.	(4.5)
(5)	Calculate $\pi_k^{\min}(\boldsymbol{x})$ and $\pi_k^{\max}(\boldsymbol{x})$ for every $k, \boldsymbol{x}$.	(4.6), (4.7)
(6)	Classify: If $\pi_k^{\min} > 0.5$, then $\hat{\alpha}_k = 1$. If $\pi_k^{\max} < 0.5$, then $\hat{\alpha}_k = 0$. Else, $\hat{\alpha}_k = *$ (unclassified).	
Output:	Classifications $\hat{\alpha}_k^i \in \{0, 1, *\}$ for all i, k.	

4.4 Measuring the impact of nonidentifiability: the marginal separability rate

Since non-identifiability is frequently unavoidable with Q-matrix based CDMs, it is important to measure its extent. For a more nuanced view, this is done on a marginal, basis.

Given the prevalence $\pi_{\boldsymbol{\alpha}}$ of each attribute profile $\boldsymbol{\alpha}$, the proportion of the population for which the k-th attribute is marginally separable can be quantified by ζ_k, as follows:

$$\zeta_k = \sum_{\{\boldsymbol{\alpha}:\delta_{[\boldsymbol{\alpha}],k}=1\}} \pi_{\boldsymbol{\alpha}}. \tag{4.8}$$

Let $\boldsymbol{\zeta}$ be the vector of all ζ_k. Then $\boldsymbol{\zeta}$ is the proportion of the population for which each attribute is marginally separable, i.e., the marginal separability rate.

Oftentimes $\pi_{\boldsymbol{\alpha}}$, and thus $\boldsymbol{\zeta}$, is unknown. Under the conditions of Theorem 11, $\boldsymbol{\zeta}$

can be consistently estimated by its maximum likelihood estimator $\hat{\boldsymbol{\zeta}}$.

Proposition 12 *Suppose an assessment follows the DINA model, with known Q-matrix Q and item parameters* $\boldsymbol{s}$ *and* $\boldsymbol{g}$*. Let* $\hat{\pi}_{[\boldsymbol{\alpha}]}$ *be the MLE estimate of* $\pi_{[\boldsymbol{\alpha}]}$*. Then*

$$\hat{\zeta}_k = \sum_{\{[\boldsymbol{\alpha}]:\delta_{[\boldsymbol{\alpha}],k}=1\}} \hat{\pi}_{[\boldsymbol{\alpha}]}, \; k = 1, \ldots, K. \tag{4.9}$$

is consistent as $N \to \infty$*.*

The consistency of $\hat{\boldsymbol{\zeta}}$ is a direct consequence of the consistency of $\hat{\pi}_{[\boldsymbol{\alpha}]}$ in Theorem 11. We thus obtain a very reasonable measure of exam quality, in terms of the proportion of the population for which each attribute is marginally separable.

4.5 Results

4.5.1 Simulation Results

We first demonstrate the procedures on simulated data. Responses are generated for $N = 5000$ resondents taking an assessment with $J = 6$ items measuring $K = 3$ distinct attributes. The respondents' mastery or nonmastery of the measured attributes is randomly generated according to the probability $\boldsymbol{\pi}^{sim}(\boldsymbol{\alpha})$ of each profile $\boldsymbol{\alpha} \in \{0,1\}^3$, as listed in Table 4.4. The population vector itself was generated randomly from the uniform distribution over the set of all $\boldsymbol{\pi} \in [0,1]^{2^K}$ summing to one.

Table 4.4: Prevalence of each attribute profile

	$\boldsymbol{\alpha}$							
	000	001	010	011	100	101	110	111
$\pi^{sim}_{\boldsymbol{\alpha}}$	0.27	0.00	0.01	0.04	0.10	0.16	0.20	0.21

The responses themselves follow the DINA model according to the Q-matrix Q^{sim} with slipping $\boldsymbol{s}^{sim}$ and guessing $\boldsymbol{g}^{sim}$ as shown in Table 4.5.

Table 4.5: Q-matrix, slipping, and guessing for simulated data.

Item (j)	Attribute vector ($\boldsymbol{q}^j$)	Slipping (s_j)	Guessing (g_j)
1.	100	0.14	0.10
2.	110	0.12	0.15
3.	011	0.18	0.18
4.	100	0.17	0.18
5.	110	0.08	0.06
6.	011	0.05	0.06

The Q-matrix Q^{sim} is incomplete, and the resulting instability in the posterior becomes clear once the data is fitted multiple times. As an example, the posterior probabilities of each attribute profile given the zero response vector $\mathbf{0} = (0, 0, \ldots, 0)$ are summarized in Table 4.6. Here, the DINA, HO-DINA, and RHO-DINA all have identifiability issues; as a result, marginal maximum likelihood via the EM algorithm may randomly produce a wide range of results for Profiles [000], [001], and [010]. The slight variability in the ind-DINA estimates is a numerical artifact. While the ind-DINA does not suffer from non-identifiability, it still does not give accurate estimates in this case since the model assumptions are incorrect.

Partitioning the attribute profile space as directed by Table 4.1 produces the five equivalence classes listed in Table 4.7, two of which have multiple members. The table also reports the marginal separability vector $\delta_{[\boldsymbol{\alpha}]}$ for each class. Note that since Items 1 and 4 are devoted to Attribute 1, it is always marginally separable and $\delta_{[\boldsymbol{\alpha}],1} \equiv 1$. It is also clear that non-identifiability most seriously affects Attribute 3, which is marginally non-identifiable for members of both [000] and [100]. Finally, Table 4.7 also reports E-M estimates of the proportion of respondents in each class under the DINA and several variants, along with the true prevalence. Note the accuracy of the

Table 4.6: Posterior probabilities given zero correct responses, $p(\boldsymbol{\alpha}|\boldsymbol{x} = \mathbf{0})$

	α							
	000	001	010	011	100	101	110	111
truth								
	0.91	0.02	0.05	0.00	0.01	0.01	0.00	0.00
minimums								
DINA	0.01	0.02	0.03	0.00	0.00	0.00	0.00	0.00
HO-DINA	0.13	0.09	0.03	0.00	0.01	0.01	0.00	0.00
RHO-DINA	0.55	0.11	0.29	0.00	0.02	0.01	0.00	0.00
ind-DINA	0.29	0.24	0.37	0.00	0.02	0.02	0.00	0.00
maximums								
DINA	0.71	0.86	0.56	0.00	0.02	0.03	0.00	0.00
HO-DINA	0.62	0.81	0.71	0.00	0.02	0.02	0.00	0.00
RHO-DINA	0.58	0.13	0.30	0.00	0.02	0.01	0.00	0.00
ind-DINA	0.31	0.28	0.40	0.01	0.03	0.02	0.00	0.00

Note: Minimum and maximum values of the posterior $p(\boldsymbol{\alpha}|\boldsymbol{x} = \mathbf{0})$, as generated over ten runs of the (random start) E-M algorithm.

DINA estimates, which are consistent, and the inaccuracy of the ind-DINA estimates due to model misfit.

We now consider variability in the marginal posterior probabilities $\Pr(\alpha_k = 1|\boldsymbol{x})$. Table 4.8 gives the sample range of $\Pr(\alpha_k = 1|\boldsymbol{x})$ after ten runs of the E-M algorithm, in addition to the theoretical range. Note the large theoretical ranges for $\Pr(\alpha_k = 1|\boldsymbol{x} = \mathbf{0}), k = 2, 3$.

Classification was conducted on a marginal basis, based on $\Pr(\alpha_k = 1|\boldsymbol{x})$, under each of the models. In addition, NIAD-DINA classification was performed (see Table 4.3). Marginal misclassification rates $\Pr(\hat{\alpha}_k \neq \alpha_k)$ are compared in Table 4.9. Note that NIAD-DINA classification results in unclassified individuals; for example, $\hat{\boldsymbol{\alpha}} = (0{*}{*})$ for those with the zero response vector. This proportion of respondents left

Table 4.7: Equivalence classes, along with their class sizes, true and maximum likelihood probabilities, and marginal separability vectors.

		$\pi_{[\boldsymbol{\alpha}]}$					
$[\boldsymbol{\alpha}]$	Size	True	DINA	HO-DINA	RHO-DINA	ind-DINA	$\delta_{[\boldsymbol{\alpha}]}$
[000]	3	0.29	0.30	0.29	0.29	0.22	100
[100]	2	0.26	0.26	0.27	0.26	0.31	110
[011]	1	0.04	0.04	0.04	0.04	0.08	111
[110]	1	0.20	0.20	0.19	0.20	0.20	111
[111]	1	0.21	0.21	0.21	0.21	0.18	111

Table 4.8: Variability in $\Pr(\alpha_k = 1|\boldsymbol{x} = \mathbf{0})$, the marginal posterior given the zero response vector.

	k		
	1	2	3
sample min	0.03	0.03	0.04
$\pi_k^{min}(\mathbf{0})$	0.03	0.00	0.00
sample max	0.03	0.56	0.89
$\pi_k^{max}(\mathbf{0})$	0.03	0.97	1.00

Note: Probabilities calculated by fitting the DINA model over ten runs of E-M algorithm with random starts.

unclassified by the NIAD-DINA algorithm is listed within parentheses in Table 4.9. The DINA and HO-DINA are overparameterized and the misclassification rate for Attribute 3 may reach over 40% in both models. The ind-DINA also performs poorly, but due to an overly restricted parameter space rather than nonidentifiability; the set of population parameters $\boldsymbol{\pi}$ allowed under the ind-DINA does not contain the true population parameter, or even something close to it. Adjusting classification under the DINA to account for nonidentifiability according to the method described in Section 4.3.2 solves both these issues. It may leave a large proportion of individuals unclassified, but this is a necessary consequence of the assessment design. Classifying these individuals would require further assumptions beyond the model.

Table 4.9: Marginal misclassification rates under a variety of models.

	Model				
k	DINA	HO-DINA	RHO-DINA	ind-DINA	NIAD-DINA
1	0.07	0.07	0.07	0.09	0.07 (0.00)
2	0.07-0.32	0.07-0.32	0.07	0.26	0.04 (0.32)
3	0.19-0.44	0.19-0.43	0.20	0.21	0.04 (0.56)

Note: Range over 10 runs reported for overparameterized models. All cut-offs equal to 0.5. The proportion of respondents left unclassified under the NIAD-DINA is displayed within parentheses.

In addition to controlling misclassification errors, we may also evaluate the quality of the assessment by measuring the marginal separability rate $\boldsymbol{\zeta}$ defined in (4.8). Table 4.10 shows both true and estimated values for $\boldsymbol{\zeta}$. Note once again that non-identifiability affects Attribute 3 more severely than it does Attribute 2. In addition, estimates are generally accurate, except in the case of the ind-DINA, which suffers from lack of fit.

Table 4.10: True and estimated values for $\boldsymbol{\zeta}$, marginal separability rate.

		Model			
k	true	DINA	HO-DINA	RHO-DINA	ind-DINA
1	1.00	1.00	1.00	1.00	1.00
2	0.71	0.70	0.71	0.71	0.78
3	0.44	0.45	0.45	0.45	0.47

In terms of model selection, reducing the number of parameters for the DINA model to $2M + L$ from the original $2M + 2^K$ reduces the comparative advantage of the restricted models. In Table 4.11, the AIC value of the RHO-DINA barely edges out that of the DINA with identifiability adjustment.

Table 4.11: AIC and BIC for the DINA, RHO-DINA, and ind-DINA.

	parameters	AIC	BIC
NIAD-DINA	17	32862.8	32973.6
RHO-DINA	16	32861.2	32965.5
ind-DINA	15	32995.9	33093.6

4.5.2 A mixed fraction subtraction dataset example

We now turn to the widely analyzed fraction subtraction data set of K. K. Tatsuoka (1990). It is composed of the twenty items listed in Table 4.12. The Q-matrix in

Table 4.12: Items from the fraction subtraction data set K. K. Tatsuoka (1990).

No.	Item	No.	Item	No.	Item	No.	Item
1.	$5/3 - 3/4$	6.	$6/7 - 4/7$	11.	$4\,1/3 - 2\,4/3$	16.	$4\,5/7 - 1\,4/7$
2.	$3/4 - 3/8$	7.	$3 - 2\,1/5$	12.	$11/8 - 1/8$	17.	$7\,3/5 - 4/5$
3.	$5/6 - 1/9$	8.	$2/3 - 2/3$	13.	$3\,3/8 - 2\,5/6$	18.	$4\,1/10 - 2\,8/10$
4.	$3\,1/2 - 2\,3/2$	9.	$3\,7/8 - 2$	14.	$3\,4/5 - 3\,2/5$	19.	$4 - 1\,4/3$
5.	$4\,3/5 - 3\,4/10$	10.	$4\,4/12 - 2\,7/12$	15.	$2 - 1/3$	20.	$4\,1/3 - 1\,5/3$

Table 4.13 comes from de la Torre and Douglas (2004), and specifies the following eight attributes: α_1 = convert a whole number to a fraction; α_2 = separate a whole number from a fraction; α_3 = simplify before subtracting; α_4 = find a common denominator; α_5 = borrow from whole number part; α_6 = column borrow to subtract the second numerator from the first; α_7 = subtract numerators; and α_8 = reduce answers to simplest form.

Table 4.13: Q-matrix from de la Torre and Douglas (2004).

Item	Attributes ($\boldsymbol{q}^j$)	Item	Attributes ($\boldsymbol{q}^j$)	Item	Attributes ($\boldsymbol{q}^j$)
1.	00010110	8.	00000010	15.	10000010
2.	00010010	9.	01000000	16.	01000010
3.	00010010	10.	01001011	17.	01001010
4.	01101010	11.	01001010	18.	01001110
5.	01010011	12.	00000011	19.	11101010
6.	00000010	13.	01011010	20.	01101010
7.	11000010	14.	01000010		

As pointed out by DeCarlo (2011), this assessment exemplifies the identifiability issues of the DINA model. While Attributes 2 and 7 have items dedicated solely to them, all other attributes appear only in combination. In fact, Attribute 3 only appears in Item 4, in conjunction with Attributes 2, 5, and 7. Attribute 7 is required for all items except one, making it difficult to draw conclusions about other attributes when it has not been mastered. Table 4.14 displays the marginal posterior probabilities of mastery for each attribute, given the zero response vector. The posterior displayed for the DINA is just one possible output of the E-M algorithm for this data; meanwhile, note the high probabilities of mastery under the ind-DINA model. Common sense dictates that something is out of place when the analysis states that students with a score of zero cannot subtract numerators, but can do everything else, from finding a common denominator to borrowing to reducing to simplest form.

With eight attributes in the Q-matrix, there are a total of 256 possible attribute profiles. They can be divided into just 58 different equivalence classes by the partitioning algorithm, 32 of them containing a single identifiable element. The 26 multiple-profile equivalence classes are listed in Table 4.15, which also displays their class sizes, maximum likelihood probabilities, and marginal separability vectors. Within these multiple-profile equivalence classes, Attributes 2 and 7 are always marginally separable, while Attribute 3 is never so; this is natural considering our previous obser-

Table 4.14: Marginal posterior probabilities of mastery given the zero response vector, $\Pr(\alpha_k = 1|\boldsymbol{x} = \mathbf{0})$

	k							
	1	2	3	4	5	6	7	8
DINA	0.50	0.08	0.50	0.52	0.53	0.41	0.00	0.59
HO-DINA	0.00	0.13	0.31	0.05	0.02	0.30	0.00	0.25
RHO-DINA	0.02	0.13	0.12	0.05	0.02	0.25	0.00	0.18
ind-DINA	0.74	0.86	0.96	0.86	0.75	0.98	0.00	0.94

vations about the Q-matrix. Profiles within the largest classes contain many zeroes, since under the DINA model attributes are not marginally separable for respondents who do not possess the other attributes that are measured simultaneously. For example, given mastery of Attribute 7, it becomes much easier to determine mastery of other attributes. When Attribute 7 has not been mastered, only mastery of Attribute 2 can be determined; this results in the two largest equivalence classes of attribute profiles, with sixty-four members each, [00000000] and [01000000]. Also note that the ind-DINA shows signs of model misspecification, since its estimates $\hat{\pi}_{[\boldsymbol{\alpha}]}$ deviate strongly from the estimates derived from the other models.

Table 4.16 shows the estimated marginal separability rates, $\hat{\boldsymbol{\zeta}}$. At the low end, $\hat{\zeta}_3 = 0.48$, bringing into question the ability of this assessment to measure mastery of Attribute 3 for a large proportion of the population. Attribute 6 does only slightly better, with $\hat{\zeta}_6 = 0.64$. Note that, like Attribute 3, Attribute 6 is seldom measured. It is only utilized in Items 1 and 18; in both cases it appears in conjunction with at least two other attributes.

The NIAD-DINA classification algorithm corrects for non-identifiability as a source of classification error and leaves individuals unclassified on certain attributes if not

Table 4.15: Multiple-member equivalence classes, along with their class sizes, maximum likelihood probabilities, and marginal separability vectors.

$[\boldsymbol{\alpha}]$	$\#[\boldsymbol{\alpha}]$	$\pi_{[\boldsymbol{\alpha}]}$				$\delta_{[\boldsymbol{\alpha}]}$
		DINA	HO-DINA	RHO-DINA	ind-DINA	
[00000000]	64	0.15	0.12	0.12	0.02	01000010
[01000000]	64	0.04	0.06	0.06	0.31	01000010
[00000010]	8	0.01	0.02	0.03	0.00	11010011
[10000010]	8	0.00	0.00	0.00	0.00	11010011
[00000011]	8	0.02	0.03	0.02	0.00	11010011
[10000011]	8	0.00	0.00	0.00	0.00	11010011
[01000010]	4	0.03	0.04	0.04	0.00	11011011
[01000011]	4	0.11	0.09	0.08	0.00	11011011
[11000010]	4	0.00	0.00	0.00	0.00	11011011
[11000011]	4	0.01	0.01	0.02	0.01	11011011
[00010010]	4	0.00	0.00	0.00	0.00	11010111
[10010010]	4	0.00	0.00	0.00	0.00	11010111
[00010011]	4	0.00	0.00	0.00	0.00	11010111
[10010011]	4	0.00	0.00	0.00	0.00	11010111
[00010110]	4	0.02	0.00	0.00	0.00	11010111
[10010110]	4	0.00	0.00	0.00	0.00	11010111
[00010111]	4	0.00	0.01	0.01	0.01	11010111
[10010111]	4	0.00	0.00	0.00	0.03	11010111
[01010010]	2	0.00	0.00	0.00	0.00	11011111
[11010010]	2	0.00	0.00	0.00	0.00	11011111
[01010011]	2	0.01	0.01	0.00	0.00	11011111
[11010011]	2	0.00	0.00	0.00	0.00	11011111
[01010110]	2	0.00	0.01	0.01	0.00	11011111
[11010110]	2	0.00	0.00	0.00	0.01	11011111
[01010111]	2	0.05	0.06	0.06	0.03	11011111
[11010111]	2	0.06	0.06	0.06	0.09	11011111

enough information can be found in the assessment to classify them. Consider the results of NIAD-DINA classification displayed in Table 4.17. Individuals with zero response patterns $\boldsymbol{x} = \mathbf{0}$, who may be erroneously classified as possessing any of the attributes other than Attributes 2 and 7 by traditional methods, are left unclassified in those dimensions by the NIAD-DINA since the range $[\pi_k^{\min}, \pi_k^{\max}]$ stretches across the boundary value of 0.5 for those k. Note that uncertainty due to non-identifiability does not have to result in leaving an individual unclassified. In the second example

Table 4.16: Estimated marginal separability rates ζ_k.

	k							
	1	2	3	4	5	6	7	8
DINA	0.81	1.00	0.48	0.81	0.75	0.64	1.00	0.81
HO-DINA	0.82	1.00	0.47	0.82	0.75	0.64	1.00	0.82
RHO-DINA	0.82	1.00	0.48	0.82	0.75	0.63	1.00	0.82
ind-DINA	0.66	1.00	0.47	0.66	0.62	0.64	1.00	0.66

in that table, the posterior expectation of α_3 has a range of $[0.18, 0.29]$. Since $\pi_3^{\max}$ is low, we can classify individuals with the second response pattern as needing work on Attribute 3 with some confidence, despite not being able to obtain a precise posterior expectation. Note that answering more questions correctly has put more posterior weight on classes $[\boldsymbol{\alpha}]$ where the respondent possesses more skills. These classes tend to have fewer members, improving our ability to classify.

Table 4.17: NIAD-DINA classification for two response patterns

	k							
	1	2	3	4	5	6	7	8
Items correct: none								
$\pi_k^{\min}$	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00
$\pi_k^{\max}$	1.00	0.08	1.00	1.00	1.00	1.00	0.00	1.00
$\hat{\alpha}_k$	$*$	0	$*$	$*$	$*$	$*$	0	$*$
Items correct: 6, 11, 12, 14, 17, 18								
$\pi_k^{\min}$	0.00	0.97	0.18	0.00	0.89	0.88	0.99	0.98
$\pi_k^{\max}$	0.01	0.97	0.29	0.00	0.92	0.99	0.99	0.99
$\hat{\alpha}_k$	0	1	0	0	1	1	1	1

'$*$' denotes that the respondent remains unclassified for that k.

Table 4.18 compares the NIAD-DINA classification decisions for three different response patterns. The first and last patterns represent the two extremes, when respondents answer either none or all of the questions correctly. The second pattern was chosen as one of the most common response patterns amongst those sufficiently different from the none correct or all correct patterns to be useful for illustration. We use this table to show that the classification decisions reached by the NIAD-DINA classification algorithm make sense, unlike those potentially reached by traditional approaches. We have already mentioned that respondents answering all items incorrectly are no longer classified as having any skills. Because most of the posterior weight is in large classes where $\alpha_7 = 0$, however, the algorithm is unable to make decisions for most of the skills. Increasing the number of correct responses in the second pattern allows us to decide the respondent has mastered Attributes 2, 7, and 8. These are relatively easy skills such as separating a whole number from a fraction, simple subtraction of numerators, and reducing answers to simplest form. We are also able to decide that the respondent has not mastered several difficult skills; Attributes 1, 4, and 5 concern converting whole numbers into fractions, finding common denominators, and borrowing from the whole number part. The respondent remains unclassified for fewer attributes with the increasing number of correct responses; he or she remains unclassified for the third and sixth attributes. This fits with the low marginal separability rates for those attributes. Finally, the response pattern where all items have been answered correctly allows for the easiest classification, and the respondent is classified as having mastered all the skills. The good news from this comparison of NIAD-DINA classifications for different response patterns is that the results make sense; the bad news is that, depending on the prevalence on the various attribute profiles, potentially many individuals may remain unclassified for several attributes at the end of the day. No further improvements can be made solely based

Table 4.18: Comparison of NIAD-DINA classifications for three response patterns

Items answered correctly $j : r_j = 1$	Skills mastered $\hat{\alpha}_k = 1$	Need work $\hat{\alpha}_k = 0$	No decision $\hat{\alpha}_k = *$
none	none	2,7	1, 3-6, 8,9
6,8,9,12,14,16	2,7,8	1,4,5	3,6
all	all	none	none

on data collected during the assessment; however, a researcher willing to make further assumptions may be able to further classify individuals. For example, an attribute hierarchy that disallows mastery of any skill without mastery of simple subtraction would allow a classification of non-mastery on all skills for individuals with the zero response pattern.

On a final, minor note, we take a look at how the reduction of the parameter space for the DINA model, based on identifiability, affects model selection by AIC and BIC. The nomial number of parameters in the DINA model is incorrect if there is non-identifiability; the number of parameters when the $\pi_{\boldsymbol{\alpha}}$ has been collapsed into $\pi_{[\boldsymbol{\alpha}]}$ counts the parameters correctly. The adjusted values for the AIC and BIC in Table 4.19 show that, in particular, the AIC no longer prefers the ind-DINA to the full DINA model once the reduced parameter space has been applied. The BIC, which will generally choose sparser models than the AIC, still reports lower values for the ind-DINA, but the comparison is much tighter.

4.6 Discussion

The issue of identifiability is a serious one for cognitive diagnosis models, with consequences for both parameter estimation and respondent classification. The impos-

Table 4.19: AIC and BIC for the DINA, RHO-DINA, and ind-DINA.

	parameters	AIC	BIC
DINA	296	9397.0	10665.2
NIAD-DINA	98	9001.0	9420.9
HO-DINA	56	8959.7	9199.6
RHO-DINA	49	8961.9	9171.9
ind-DINA	48	9208.3	9413.9

sibility of learning certain types of information under non-identifiability leaves both procedures highly sensitive to the prior placed on the data; in fact, diagnostic assessments can easily contain no information differentiating certain types of profiles. This is especially true for large-scale assessments that were not created with cognitive diagnosis in mind. Attempting to diagnose overly fine-grained attributes also tends to make avoiding non-identifiability difficult, as does attempting to diagnose too many attributes with too few items. The proposed Q-matrix for the TIMSS 2003 Mathematics in Su et al. (2013), for example, has all of these failings. Diagnostic decisions based on such assessments should be made with great caution and must account for identifiability issues in some way.

Some methods currently in use may resolve identifiability issues by enforcing restrictions on the attribute profile space. Variants of the DINA such as the ind-DINA, HO-DINA, and RHO-DINA accomplish this by specifying a structure and a prior on the probabilities $\pi_{\boldsymbol{\alpha}}$. Although this may eliminate non-identifiability and create a unique global maximum for the likelihood, inappropriate priors and model misspecification becomes a risk. Thus, careful comparison of these variants to the NIAD-DINA becomes important. Similarly, assuming attribute hierarchies may alleviate the issue of identifiability by restricting the parameter space, but misspecification of attribute hierarchies may cause issues.

Using the tools discussed in this paper, it is possible to determine the extent to which non-identifiability affects classification and estimation under the DINA model. Marginal separability rates $\boldsymbol{\zeta}$, which can be estimated consistently, provide an overall measure of the extent of non-identifiability; meanwhile, NIAD-DINA classification should be used to control classification errors that are otherwise quite sensitive to the prior information. The results here suggest that when all items testing a particular attribute also test other attributes, non-identifiability may not be such a serious issue of the other attributes are 'basic' ones mastered by a large proportion of the population. After all, it is only impossible to recover information about a particular attribute when the respondent does not possess one or more of the other attributes tested by the same item. This statement reverses itself for the DINO model; while multi-attribute items are just as problematic, but the attribute profiles that non-identifiability affects the most are those where many skills are in the 'mastered' state.

This observation has important applications to experiment design. Identifiability can often be difficult to achieve; for example, it may be difficult to create items measuring advanced skills without including basic ones. In addition, the design of educational assessment often balances a variety of concerns, and items measuring only one attribute may be less desirable for efficiency reasons, for instance. The results from this paper can guide a practioner in balancing concerns about identifiability and efficiency. The impact of non-identifiability depends on the prevalence of certain attribute profiles in the population. When the proportion of the population possessing certain basic skills is very high, measuring more advanced skills with these basic skills will seldom lead to issues with statistical inference. The prevalence $\pi_{\boldsymbol{\alpha}}$ will have a very small range, and classification on advanced skills will only be difficult for the small number of individuals who do not possess the basic skills. It then becomes reasonable to push for an efficiency gain in measuring these basic skills together, for example, and items can measure advanced skills together with these basic skills with little loss of information. Interestingly enough, relaxing the identifiability conditions of the DINA model through attribute hierarchy brings similar lessons in assessment design. Recall that under attribute hierarchy, an item measuring the k-th attribute in combination with its prerequisites could replace the item measuring solely the k-th attribute in our original requirements for completeness. In other words, when the attribute hierarchy pushes the prevalence of profiles where the basic skills are unknown and the advanced skills are known to zero, identifiability can be achieved even with items requiring a combination of these skills. Thus, again, measuring advanced skills in combination with basic skills does not lead to much of a loss due to non-identifiability.

This thesis addresses the identifiability issue for the DINA and its variants, including its dual, the DINO model, but identifiability is a major concern for most Q-matrix based cognitive diagnosis models. This is especially true for larger models such as the General Diagnostic Model von Davier (2005), the Log-Linear Cognitive Diagnosis Model Henson et al. (2009), and the Generalized DINA Model de la Torre (2011), where the situation is even worse because of the larger number of parameters. The exact identifiability requirements of these models remains a topic of study that must be tackled before they can be rigorously applied to real data. In the most difficult cases, simultaneous parameter estimation and respondent classification may be inadvisable; separate rounds of item calibration and diagnostic testing can be conducted for better results.

References

Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory.* Long Grove, IL: Waveland Press, Inc. (Original work published 1979)

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

de la Torre, J. (2008). *The generalized DINA model.* Paper presented at the annual International Meeting of the Psychometric Society, Durham, NH.

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

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

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

de la Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, *73*, 595-624.

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

diagnose. (2014). In *OED Online.* Oxford University Press. Retrieved April 24, 2014, from `http://www.oed.com/view/Entry/51835?redirectedFrom=diagnose&`

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

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

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*, 301-323.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage Publications, Inc.

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

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

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

Koopmans, T. C., & Reiersøl, O. (1950). The identification of structural characteristics. *Ann. Math. Statist.*, *21*, 165-181.

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

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

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

Lord, F. M. (1952). The relation of the reliability of multiple-choice tests to the distribution of item difficulties. *Psychometrika*, *17*, 181–194.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Erlbaum.

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

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

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

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

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*, 1-18.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Danmarks Paedagogiske Institut.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth berkeley symposium on mathematical statistics and probability* (Vol. 4, pp. 321–333).

Roussos, L. A., Templin, J. L., & Henson, R. A. (2007). Skills diagnosis using IRT-based latent class models. *Journal of Educational Measurement*, *44*, 293-311.

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

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72-101.

Su, Y.-L., Choi, K. M., Lee, W.-C., Choi, T., & McAninch, M. (2013). *Hierarchical* *cognitive diagnostic analysis for timss 2003 mathematics* (CASMA Research report No. 35). Iowa City, Iowa: Center for Advanced Studies in Measurement and Assessment.

Tatsuoka, C. (1996). *Sequential classification on partially ordered sets.* Unpublished doctoral dissertation, Cornell University.

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

Tatsuoka, C., & Ferguson, T. (2003). Sequential classification on partially ordered sets. *Journal of the Royal Statistical Society: Series B*, *65*, 143-157.

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

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

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

Templin, J. L. (2006). CDM: cognitive diagnosis modeling with Mplus [Computer software manual]. Retrieved from `http://jtemplin.myweb.uga.edu/cdm/cdm.html`

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

Templin, J. L., Henson, R. A., & Douglas, J. A. (2006, April). *General theory and estimation of cognitive diagnosis models: Using Mplus to derive model estimates.* Paper presented at the 2007 National Council on Measurement in Education training session. Chicago, IL.

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

Traub, R. E. (2005). Classical test theory in historical perspective. *Measurement: Issues and Practice*, *16*(4), 8-14.

Tucker, L. R. (1946). Maximum validity of a test with equivalent items. *Psychometrika*, *11*, 1–13.

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

von Davier, M. (2005). *A general diagnostic model applied to language testing data* (Research report). Princeton, NJ: Educational Testing Service.

von Davier, M. (2009). Some notes on the reinvention of latent structure models as diagnostic classification models. *Measurement: Interdisciplinary Research and Perspectives*, *7*, 67-74.

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

Appendix A

Proofs

A.1 The T-matrix

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

The T-matrix specifies the probability that an individual with attribute profile $\boldsymbol{\alpha}$ will answer all items in some subset of the items $S \subseteq \{1, \ldots, J\}$ correctly. The subsets S may be indexed by response vectors $\boldsymbol{x} \in \{0,1\}^J$ with exactly the items in S correct, i.e., $x_j = 1$ iff $j \in S$. Then the T-matrix $T(Q, \boldsymbol{c}, \boldsymbol{g})$ is a matrix of conditional survival probabilities $\Pr(\boldsymbol{X} \succeq \boldsymbol{x} | Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\alpha})$ written as a function of Q, $\boldsymbol{c}$, and $\boldsymbol{g}$. Thus its entries, indexed by $\boldsymbol{x}$ and $\boldsymbol{\alpha}$, are

$$\begin{aligned} t_{\boldsymbol{x},\boldsymbol{\alpha}}(Q, \boldsymbol{c}, \boldsymbol{g}) &= \Pr(\boldsymbol{X} \succeq \boldsymbol{x} | Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\alpha}) \\ &= \sum_{\boldsymbol{x}' \succeq \boldsymbol{x}} p(\boldsymbol{x}' | Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\alpha}) = \prod_{j : x_j = 1} c_j^{\xi_j(Q,\boldsymbol{\alpha})} g_j^{1-\xi_j(Q,\boldsymbol{\alpha})} \end{aligned}$$

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

By definition, multiplying the T-matrix by the the distribution of attribute profiles $\boldsymbol{\pi}$ results in a vector containing the marginal survival probabilities $\Pr(\boldsymbol{X} \geq \boldsymbol{x}|Q, \boldsymbol{c}, \boldsymbol{g})$ of successfully answering each subset of items correctly. The $\boldsymbol{x}$-th entry of this vector is

$$\begin{aligned} T_{\boldsymbol{x}}(Q, \boldsymbol{c}, \boldsymbol{g})^\top \boldsymbol{\pi} &= \sum_{\boldsymbol{\alpha}} t_{\boldsymbol{x},\boldsymbol{\alpha}}(Q, \boldsymbol{c}, \boldsymbol{g}) \pi_{\boldsymbol{\alpha}} \\ &= \sum_{\boldsymbol{x}' \succeq \boldsymbol{x}} p(\boldsymbol{x}'|Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi}) = \Pr(\boldsymbol{X} \succeq \boldsymbol{x}|Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi}). \end{aligned}$$

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

$$T(Q, \boldsymbol{c}, \boldsymbol{g})\boldsymbol{\pi} = \begin{pmatrix} 1 \\ \Pr(X_1 = 1|Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi}) \\ \vdots \\ \Pr(X_J = 1|Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi}) \\ \Pr(X_1 = 1, X_2 = 1|Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi}) \\ \vdots \\ \Pr(X_j = 1 \text{ for } j = 1, \ldots, J|Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi}) \end{pmatrix}.$$

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

A.2 Two useful propositions

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

Recall from Definitions 1 and 2 that identifiability and local identifiability depend on the probability density function $f(x;\theta)$, which, when written as a function of the parameters θ becomes the likelihood $L(\theta)$.

Under the DINA model specified in Section 2.2.2, given the full set of observations $X = \{\boldsymbol{X}^i : i = 1, \ldots, N\}$ and a Q-matrix Q, the likelihood of any set of parameters $\boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi}$ can be written as

$$L(\boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi}; R) = \prod_{i=1}^{n} p(\boldsymbol{X}^i | Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi}) = \prod_{\boldsymbol{x} \in \{0,1\}^J} p(\boldsymbol{x} | Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi})^{N_{\boldsymbol{x}}} \tag{A.1}$$

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

$$p(\boldsymbol{x} | Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi}) = \sum_{\boldsymbol{\alpha}} \pi_{\boldsymbol{\alpha}} \prod_{j=1}^{J} \Pr(X_j = x_j | Q, c_j, g_j, \boldsymbol{\alpha}) \tag{A.2}$$

is the probability of observing $\boldsymbol{x}$ given $Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\pi}$. The conditional probability $\Pr(X_j = x_j | Q, c_j, g_j, \boldsymbol{\alpha})$ may be expressed as

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

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

Proposition 13 *For two sets of parameters* $(\hat{\boldsymbol{c}}, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\pi}})$ *and* $(\bar{\boldsymbol{c}}, \bar{\boldsymbol{g}}, \bar{\boldsymbol{\pi}})$*,*

$$L(\hat{\boldsymbol{c}}, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\pi}}; R) = L(\bar{\boldsymbol{c}}, \bar{\boldsymbol{g}}, \bar{\boldsymbol{\pi}}; R)$$

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

$$T(Q, \hat{\boldsymbol{c}}, \hat{\boldsymbol{g}})\hat{\boldsymbol{\pi}} = T(Q, \bar{\boldsymbol{c}}, \bar{\boldsymbol{g}})\bar{\boldsymbol{\pi}}. \tag{A.3}$$

Proof. The observations follow a multinomial distribution over the set of possible responses $\boldsymbol{x} \in \{0,1\}^J$, with probabilities $\pi_{\boldsymbol{x}}$ as defined in (A.2). Consider a P-matrix $P(Q, \boldsymbol{c}, \boldsymbol{g})$ indexed like the T-matrix by response vectors $\boldsymbol{x} \in \{0,1\}^J$ and attribute profiles $\boldsymbol{\alpha} \in \{0,1\}^K$. The entries of $P(Q, \boldsymbol{c}, \boldsymbol{g})$ are

$$p_{\boldsymbol{x},\boldsymbol{\alpha}}(Q, \boldsymbol{c}, \boldsymbol{g}) = \Pr(\boldsymbol{X} = \boldsymbol{x} | Q, \boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\alpha}).$$

For a particular $(\hat{\boldsymbol{c}}, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\pi}})$ and $(\bar{\boldsymbol{c}}, \bar{\boldsymbol{g}}, \bar{\boldsymbol{\pi}})$,

$$\begin{aligned} L(\hat{\boldsymbol{c}}, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\pi}}; R) &= \prod_{\boldsymbol{x} \in \{0,1\}^J} (P_{\boldsymbol{x}}(Q, \hat{\boldsymbol{c}}, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\pi}}))^{N_{\boldsymbol{x}}} \\ &= \prod_{\boldsymbol{x} \in \{0,1\}^J} (P_{\boldsymbol{x}}(Q, \bar{\boldsymbol{c}}, \bar{\boldsymbol{g}}, \bar{\boldsymbol{\pi}}))^{N_{\boldsymbol{x}}} = L(\bar{\boldsymbol{c}}, \bar{\boldsymbol{g}}, \bar{\boldsymbol{\pi}}; R) \end{aligned}$$

for all observation matrices X iff $P_{\boldsymbol{x}}(Q, \hat{\boldsymbol{c}}, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\pi}}) = P_{\boldsymbol{x}}(Q, \bar{\boldsymbol{c}}, \bar{\boldsymbol{g}}, \bar{\boldsymbol{\pi}})$ for all $\boldsymbol{x} \in \{0,1\}^J$. In matrix notation as $P(Q, \hat{\boldsymbol{c}}, \hat{\boldsymbol{g}})\hat{\boldsymbol{\pi}} = P(Q, \bar{\boldsymbol{c}}, \bar{\boldsymbol{g}})\bar{\boldsymbol{\pi}}$. Since

$$t_{\boldsymbol{x},\boldsymbol{\alpha}}(Q, \boldsymbol{c}, \boldsymbol{g}) = \Pr(\boldsymbol{X} \succeq \boldsymbol{x} | Q, \boldsymbol{c}, \boldsymbol{g}) = \sum_{\boldsymbol{x} \succeq \boldsymbol{x}} p_{\boldsymbol{x},\boldsymbol{\alpha}}(Q, \boldsymbol{c}, \boldsymbol{g})$$

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

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

■

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

Proposition 14 *There exists a matrix $D(\boldsymbol{g}^*)$ depending solely on $\boldsymbol{g}^* = (g_1^*, ..., g_J^*)$, such that for any $\boldsymbol{g}^* \in \mathbb{R}^J$,*

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

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

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

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

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

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

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

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

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

$$d_{\boldsymbol{x},\boldsymbol{x}'}(\boldsymbol{g}^*) = \begin{cases} 0 & \boldsymbol{x}' \not\preceq \boldsymbol{x} \\ (-1)^{\sum_{j=1}^J x_j - x_j'} \prod_{j:x_j - x_j' = 1} g_j^* & \boldsymbol{x}' \prec \boldsymbol{x} \\ 1 & \boldsymbol{x}' = \boldsymbol{x}. \end{cases}$$

Then

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

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

A.3 Proofs for Chapter 3

Proof. [Theorem 1] The case where $\boldsymbol{s} = \boldsymbol{g} = \mathbf{0}$ was shown by (Chiu et al. 2009). For general known $\boldsymbol{s}$ and $\boldsymbol{g}$, by Proposition 13, $\boldsymbol{\pi}$ is nonidentifiable when $Q, \boldsymbol{c}, \boldsymbol{g}$ are known iff there exists $\hat{\boldsymbol{\pi}}, \bar{\boldsymbol{\pi}} \in \mathbb{R}_+^{2^K}$ such that

$$T(Q, \boldsymbol{c}, \boldsymbol{g})\hat{\boldsymbol{\pi}} = T(Q, \boldsymbol{c}, \boldsymbol{g})\bar{\boldsymbol{\pi}}.$$

This occurs iff $T(Q, \boldsymbol{c}, \boldsymbol{g})$ is not a full-rank matrix.

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

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

and $\text{rank}(T(Q, \boldsymbol{c}, \boldsymbol{g})) < 2^K$.

When Q is complete, assume WLOG that $Q_{1:K} = I_K$. The matrix $T(Q, \boldsymbol{c}, \boldsymbol{g})$ is full-rank iff $T(Q, \boldsymbol{c} - \boldsymbol{g}, \mathbf{0})$ is full-rank, since, by Proposition 14, $T(Q, \boldsymbol{c} - \boldsymbol{g}, \mathbf{0}) = D(\boldsymbol{g})T(Q, \boldsymbol{c}, \boldsymbol{g})$ and $D(\boldsymbol{g})$ is invertible. Consider the rows of $T(Q, \boldsymbol{c} - \boldsymbol{g}, \mathbf{0})$ corresponding to combinations of the first K items, i.e. $\boldsymbol{x} \in \{0, 1\}^J$ s.t. $x_j = 0$ for all $j > K$. This constitutes an upper-triangular sub-matrix of size $2^K \times 2^K$ with diagonal entries $\prod_{j: x_j = 1}(c_j - g_j) \neq 0$. Thus, $T(Q, \boldsymbol{c} - \boldsymbol{g}, \mathbf{0})$ is full-rank, and $\boldsymbol{\pi}$ is identifiable. ■

Proof. [Theorem 2] When $\boldsymbol{g}$ is known, we may combine Propositions 13 and 14 to show that two sets of parameters $(\hat{\boldsymbol{c}}, \boldsymbol{g}, \hat{\boldsymbol{\pi}})$ and $(\bar{\boldsymbol{c}}, \boldsymbol{g}, \bar{\boldsymbol{\pi}})$ produce equal likelihoods iff

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

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

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

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

To find $\boldsymbol{x}^j$ for each item j, first suppose that $j \in \{1, \ldots, K\}$. Then $Q_j = \boldsymbol{e}_j^\top$ and there is some $j' \in \{K+1, \ldots, J\}$ s.t. $q_{j'j} = 1$. Let $\boldsymbol{x}^j = \boldsymbol{e}_{j'}$. Otherwise, when $j \in \{K+1, \ldots, J\}$, let $\boldsymbol{x}^j = \sum_{\{\ell : q_{j,\ell} = 1\}} \boldsymbol{e}_\ell$.

Then given any two sets of parameters $(\hat{\boldsymbol{c}}, \mathbf{0}, \hat{\boldsymbol{\pi}})$ and $(\bar{\boldsymbol{c}}, \mathbf{0}, \bar{\boldsymbol{\pi}})$ s.t. $T(Q, \hat{\boldsymbol{c}}, \mathbf{0})\hat{\boldsymbol{\pi}} = T(Q, \bar{\boldsymbol{c}}, \mathbf{0})\bar{\boldsymbol{\pi}}$,

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

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

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

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

Then, for any $\boldsymbol{x} \in \{0,1\}^J$ s.t. $x_1 = 0$, $T_{\boldsymbol{x}}(Q, \hat{\boldsymbol{c}}, \boldsymbol{0}) = T_{\boldsymbol{x}}(Q, \bar{\boldsymbol{c}}, \boldsymbol{0})$ and

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

Otherwise, $x_1 = 1$ and

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

Thus we have found distinct sets of parameters satisfying (A.3), and shown that Condition C2 is necessary. ■

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

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

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

As was done for x_1 in the proof of necessity for Theorem 2, consider each possible value of $(x_1, x_2) \in \{0,1\}^2$ to conclude that, for any distinct sets of parameters $(\hat{\boldsymbol{c}}, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\pi}})$

and $(\bar{\boldsymbol{c}}, \bar{\boldsymbol{g}}, \bar{\boldsymbol{\pi}})$, $T(Q, \hat{\boldsymbol{c}}, \hat{\boldsymbol{g}},)\hat{\boldsymbol{\pi}} = T(Q, \bar{\boldsymbol{c}}, \bar{\boldsymbol{g}})\bar{\boldsymbol{\pi}}$ if for every $\boldsymbol{\alpha} \in \{0,1\}^K$ s.t. $\alpha_1 = 0$,

$$\begin{cases} \hat{\pi}_{\boldsymbol{\alpha}} + \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} = \bar{\pi}_{\boldsymbol{\alpha}} + \bar{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} & (x_1, x_2) = (0,0) \\ \hat{c}_1 \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \hat{g}_1 \hat{\pi}_{\boldsymbol{\alpha}} = \bar{c}_1 \bar{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \bar{g}_1 \bar{\pi}_{\boldsymbol{\alpha}} & (x_1, x_2) = (1,0) \\ \hat{c}_2 \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \hat{g}_2 \hat{\pi}_{\boldsymbol{\alpha}} = \bar{c}_2 \bar{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \bar{g}_2 \bar{\pi}_{\boldsymbol{\alpha}} & (x_1, x_2) = (0,1) \\ \hat{c}_1 \hat{c}_2 \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \hat{g}_1 \hat{g}_2 \hat{\pi}_{\boldsymbol{\alpha}} = \bar{c}_1 \bar{c}_2 \bar{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \bar{g}_1 \bar{g}_2 \bar{\pi}_{\boldsymbol{\alpha}} & (x_1, x_2) = (1,1). \end{cases} \tag{A.4}$$

Otherwise, the Q-matrix can be written WLOG as

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

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

$$\begin{cases} \hat{\pi}_{\boldsymbol{\alpha}} + \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} = \bar{\pi}_{\boldsymbol{\alpha}} + \bar{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} & \forall\ \boldsymbol{\alpha} : \alpha_1 = 0 \\ \hat{c}_1 \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \hat{g}_1 \hat{\pi}_{\boldsymbol{\alpha}} = \bar{c}_1 \bar{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \bar{g}_1 \bar{\pi}_{\boldsymbol{\alpha}} & \forall\ \boldsymbol{\alpha} : \alpha_1 = 0 \\ \hat{c}_2 \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \hat{g}_2 \hat{\pi}_{\boldsymbol{\alpha}} = \bar{c}_2 \bar{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \bar{g}_2 \bar{\pi}_{\boldsymbol{\alpha}} & \forall\ \boldsymbol{\alpha} : \alpha_1 = 0, \boldsymbol{\alpha} \succeq (0\ \boldsymbol{v}^\top) \\ \hat{c}_1 \hat{c}_2 \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \hat{g}_1 \hat{g}_2 \hat{\pi}_{\boldsymbol{\alpha}} = \bar{c}_1 \bar{c}_2 \bar{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} + \bar{g}_1 \bar{g}_2 \bar{\pi}_{\boldsymbol{\alpha}} & \forall\ \boldsymbol{\alpha} : \alpha_1 = 0, \boldsymbol{\alpha} \succeq (0\ \boldsymbol{v}^\top). \end{cases} \tag{A.5}$$

Since the equations in (A.5) are a subset of the equations in (A.4), finding sets of parameters $(\hat{\boldsymbol{c}}, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\pi}})$ and $(\bar{\boldsymbol{c}}, \bar{\boldsymbol{g}}, \bar{\boldsymbol{\pi}})$ fulfilling (A.4) completes the proof for both types of Q-matrics.

Choose a valid set of parameters $(\hat{\boldsymbol{c}}, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\pi}})$ s.t. $\hat{\pi}_{\boldsymbol{\alpha}} / \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} = \rho$ is constant over all $\boldsymbol{\alpha} \in \{0,1\}^K$ s.t. $\alpha_1 = 0$. Then, for any $\bar{\boldsymbol{g}} \in \mathbb{R}^J$, setting

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

and setting

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

for every $\boldsymbol{\alpha} \in \{0,1\}^K$ s.t. $\alpha_1 = 0$ results in a solution to (A.4). By continuity, there is $\bar{\boldsymbol{g}}$ sufficiently close to $\hat{\boldsymbol{g}}$ so that $\bar{\boldsymbol{c}}, \bar{\boldsymbol{g}} \in [0,1]^J$, $\boldsymbol{c} \succ \boldsymbol{g}$, and $\bar{\boldsymbol{\pi}} \succ \mathbf{0}$. Thus, the model is non-identifiable when Condition C3 fails, making it a necessary condition. ■

Proof. [Theorem 4] Suppose that Conditions C1 and C3 hold, and let $(\hat{\boldsymbol{c}}, \hat{\boldsymbol{g}}, \hat{\boldsymbol{\pi}})$ and $(\bar{\boldsymbol{c}}, \bar{\boldsymbol{g}}, \bar{\boldsymbol{\pi}})$ be two sets of parameters solving equation (A.3). According to Condition C1, there is an item requiring solely the k-th attribute for each $k \in \{1, \ldots, K\}$. Moreover, by Condition C3, there are also least two additional items requiring the k-th attribute. We begin the proof of sufficiency by showing that for every k, there exists an item j requiring the k-th attribute s.t. $\hat{g}_j = \bar{g}_j$. The case where all these items require solely the k-th attribute and the case where at least one requires multiple attributes are treated separately.

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

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

By Proposition 14, $T(Q, \hat{\boldsymbol{c}}, \hat{\boldsymbol{g}})\hat{\boldsymbol{\pi}} = T(Q, \bar{\boldsymbol{c}}, \bar{\boldsymbol{g}})\bar{\boldsymbol{\pi}}$ iff

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

Then, since

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

we may conclude that

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

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

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

and

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

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

Case 2 At least one item requiring the k-th attribute requires multiple attributes.

WLOG, $k = 1$ and

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

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

Since

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

we know that

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

Thus,

$$\begin{aligned} & \frac{\tilde{g}_1\tilde{g}_2\bar{\pi}_{0,0}+\tilde{c}_1\tilde{g}_2\bar{\pi}_{1,0}+\tilde{g}_1\tilde{g}_2\bar{\pi}_{0,1}+\tilde{c}_1\tilde{c}_2\bar{\pi}_{1,1}}{\tilde{g}_2\bar{\pi}_{0,0}+\tilde{g}_2\bar{\pi}_{1,0}+\tilde{g}_2\bar{\pi}_{0,1}+\tilde{c}_2\bar{\pi}_{1,1}} \\ = \; & \frac{\tilde{g}_1\tilde{g}_3\tilde{g}_2\bar{\pi}_{0,0}+\tilde{c}_1\tilde{g}_3\tilde{g}_2\bar{\pi}_{1,0}+\tilde{g}_1\tilde{c}_3\tilde{g}_2\bar{\pi}_{0,1}+\tilde{c}_1\tilde{c}_3\tilde{c}_2\bar{\pi}_{1,1}}{\tilde{g}_3\tilde{g}_2\bar{\pi}_{0,0}+\tilde{g}_3\tilde{g}_2\bar{\pi}_{1,0}+\tilde{c}_3\tilde{g}_2\bar{\pi}_{0,1}+\tilde{c}_3\tilde{c}_2\bar{\pi}_{1,1}}, \end{aligned}$$

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

$$\tilde{c}_2 = \frac{(\bar{c}_2 - \hat{c}_2)\sum_{\boldsymbol{\alpha}:\boldsymbol{\alpha}\succeq Q_2} \pi_{\boldsymbol{\alpha}} + (\bar{g}_2 - \hat{g}_2)\sum_{\boldsymbol{\alpha}:\alpha_1=\alpha_2=1,\boldsymbol{\alpha}\not\succeq Q_2} \bar{\pi}_{\boldsymbol{\alpha}}}{\bar{\pi}_{1,1}}.$$

Cross-multiply and cancel to obtain that

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

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

$$\bar{\pi}_{1,0}\bar{\pi}_{0,1}(\bar{g}_2 - \hat{g}_2) = \bar{\pi}_{0,0}\bar{\pi}_{1,1}(\bar{c}_2 - \hat{g}_2). \tag{A.6}$$

In addition, by symmetry,

$$\hat{\pi}_{1,0}\hat{\pi}_{0,1}(\hat{g}_2 - \bar{g}_2) = \hat{\pi}_{0,0}\hat{\pi}_{1,1}(\hat{c}_2 - \bar{g}_2), \tag{A.7}$$

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

Taken together, (A.6) and (A.7) imply that either $\hat{c}_3 > \hat{g}_3 > \bar{c}_3 > \bar{g}_3$ or $\bar{c}_3 > \bar{g}_3 > \hat{c}_3 > \hat{g}_3$. However, since $T_{\boldsymbol{e}_2}(\hat{\boldsymbol{c}}, \hat{\boldsymbol{g}})\hat{\boldsymbol{\pi}} = T_{\boldsymbol{e}_2}(\bar{\boldsymbol{c}}, \bar{\boldsymbol{g}})\bar{\boldsymbol{\pi}}$,

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

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

WLOG, the Q-matrix can be written as

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

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

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

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

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

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

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

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

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

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

$$\boldsymbol{x}^k = \sum_{j=K+1}^{J} \boldsymbol{e}_j(1 - q_{j,k})$$

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

$$T_{\boldsymbol{x}}(Q, \boldsymbol{c} - \boldsymbol{g}^*, \boldsymbol{g} - \boldsymbol{g}^*)\boldsymbol{\pi} = \left(\prod_{j=K+1}^{J} (c_j - \hat{g}_j)^{x_j} \right) \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \pi_{\boldsymbol{\alpha}} t_{\boldsymbol{x},\boldsymbol{\alpha}}(Q)$$

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

$$\sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \hat{\pi}_{\boldsymbol{\alpha}} t_{\boldsymbol{x},\boldsymbol{\alpha}}(Q) = \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \bar{\pi}_{\boldsymbol{\alpha}} t_{\boldsymbol{x},\boldsymbol{\alpha}}(Q) \tag{A.8}$$

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

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

When Condition C4 holds, there is some $\boldsymbol{x}$ s.t. $x_j = 0$ for all $j \leq K$ and $T_{\boldsymbol{e}_k + \boldsymbol{x}^k}(Q) = T_{\boldsymbol{x} + \boldsymbol{x}^k}(Q)$. Then, by (A.8)

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

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

Proof. [Theorem 5] By Proposition 13, $\boldsymbol{\pi}$ is nonidentifiable when $Q, \boldsymbol{c}, \boldsymbol{g}$ are known iff there exists $\hat{\boldsymbol{\pi}}, \bar{\boldsymbol{\pi}} \in \mathbb{R}_+^{2^K}$ such that

$$T(Q, \boldsymbol{c}, \boldsymbol{g})\hat{\boldsymbol{\pi}} = T(Q, \boldsymbol{c}, \boldsymbol{g})\bar{\boldsymbol{\pi}}.$$

Attribute hierarchy forces some entries of $\boldsymbol{\pi}$ to zero. By assumption, for all $\boldsymbol{\alpha}$ allowed under the attribute hierarchy, $\boldsymbol{\pi}_{\boldsymbol{\alpha}} > 0$. It is possible to find valid $\hat{\boldsymbol{\pi}} \neq \bar{\boldsymbol{\pi}}$ iff the reduced T-matrix $T^*(Q, \boldsymbol{c}, \boldsymbol{g})$, which only contains columns corresponding to attribute profiles allowed by the attribute hierachy, is not a full column rank matrix.

Suppose that the identifiability condition does not hold. WLOG, we assume that it fails for the first attribute, i.e., all items measuring the first attribute also measure at least one attribute that is not a prerequisite of the first attribute. Consider the attribute profile $\boldsymbol{\alpha}^{(1)}$, defined by $\alpha_k^{(1)} = I(k \trianglelefteq 1),\ k = 1, \ldots, K$; the profile $\boldsymbol{\alpha}^{(1)}$ represents possession of the first attribute and all its prerequisites. Compare to $\boldsymbol{\alpha}^{(2)}$, defined by $\alpha_k^{(2)} = I(k \triangleleft 1),\ k = 1, \ldots, K$, the profile representing possession of all the first attribute's prerequisites, but not the first attribute itself. By the transitivity property of partial orders, $\boldsymbol{\alpha}^{(1)}$ and $\boldsymbol{\alpha}^{(2)}$ are both valid attribute profiles under the attribute hierarchy. To elaborate for $\boldsymbol{\alpha}^{(2)}$, if $\alpha_{k_1}^{(2)} = 1$ for some $k_1 \neq 1$, then for any prerequisite Attribute k_2 of Attribute k_1, we need to show that $\alpha_{k_2}^{(2)} = 1$. By transitivity $k_2 \triangleleft k_1, k_1 \triangleleft 1 \Rightarrow k_2 \triangleleft 1$. Then $\alpha_{k_2}^{(2)} = 1$ by construction.

We show that the ideal response vectors of $\boldsymbol{\alpha}^{(1)}$ and $\boldsymbol{\alpha}^{(2)}$ are identical by contradiction. Suppose there is some item j s.t. $\xi_j(Q, \boldsymbol{\alpha}^{(1)} > \xi_j(Q, \boldsymbol{\alpha}^{(2)})$. Then $q_{j1} = 1$. Since the identifiability condition does not hold, $q_{k1} = 1$ for some $k \in \{2, \ldots, K\}$ s.t. $k \not\trianglelefteq 1$. Then by construction $\xi_j(Q, \boldsymbol{\alpha}^{(1)}) = 0$, which is a contradiction. Now, since the ideal responses are identical, the columns of the reduced T-matrix T^* corresponding to $\boldsymbol{\alpha}^{(1)}$ and $\boldsymbol{\alpha}^{(2)}$ are identical and the T^* is not full column rank.

Suppose Q fulfills the identifiability condition. WLOG, for items $j = 1, \ldots, K$, the

j-th item requires only the j-th attribute and, possibly, a subset of its prerequisites. Mathematically, $q_{jj} = 1$ and $q_{j,k} = 0$ for all k s.t. $k \ntrianglelefteq j$. Now the 'key item' measuring the k-th attribute is precisely the k-th item; we have 'matched' the k-th item and k-th attribute for $k = 1, \ldots, K$. Next, recall that the columns of the reduced T-matrix T^* are lexicographically ordered and correspond to the attribute profiles allowed by the attribute hierarchy. We find a square sub-matrix T^{**} by selecting only the rows of T^* corresponding to combinations of items that 'match' those attribute profiles. If we suppose the ℓ-th column of T^* corresponds to profile $\boldsymbol{\alpha}^{(\ell)}$, then the ℓ-th row of T^{**} corresponds to the set of items $\{j : \alpha_j^{(\ell)} = 1\}$.

We show that $T^{**}(Q, \boldsymbol{c} - \boldsymbol{g}, \boldsymbol{0})$ is upper triangular, and that thus, by way of Proposition 14, $T^*(Q, \boldsymbol{c}, \boldsymbol{g})$ is full column rank. Consider the diagonal entries of $T^{**}(Q, \boldsymbol{c} - \boldsymbol{g}, \boldsymbol{0})$. Since $\boldsymbol{c} \prec \boldsymbol{g}$, the entries of $\text{diag}(T^{**}(Q, \boldsymbol{c} - \boldsymbol{g}, \boldsymbol{0}))$ are nonzero iff $\boldsymbol{\alpha}^{(\ell)}$ fulfills the attribute requirements of the ℓ-th row of T^{**}. Let $S_\ell = \{k : \alpha_k^{(\ell)} = 1\}$. It is both the set of attributes indicated by the ℓ-th column and the set of items indicated by the ℓ-th row. For any item $j \in S_\ell$, $q_{j,k} = 1 \Rightarrow k \trianglelefteq j \Rightarrow \alpha_k^{(\ell)} = 1$. Then $\xi_j(Q, \boldsymbol{\alpha}^{(\ell)}) = \prod_{k \in S_j} \alpha_k^{(\ell)} = 1$ for all $j \in S_\ell$ and $t^{**}_{\ell,\ell}(Q, \boldsymbol{c} - \boldsymbol{g}, \boldsymbol{0}) > 0$. Below the diagonal, consider the entry at the ℓ_1-th column and ℓ_2-th row, with $\ell_1 < \ell_2$. The items indicated by the ℓ_2-th row require all attributes k s.t. $\alpha_k^{(\ell_2)} = 1$. Since the columns are ordered lexicographically, $\exists k \in S_{\ell_2}$ s.t. $1 = \alpha_k^{(\ell_2)} > \alpha_k^{(\ell_1)} = 0$. Then $\xi_k(Q, \boldsymbol{\alpha}_1^{(\ell_1)} = 0 \Rightarrow t^{**}_{\ell_2,\ell_1}(Q, \boldsymbol{c} - \boldsymbol{g}, \boldsymbol{0}) = 0$. So $T^{**}(Q, \boldsymbol{c} - \boldsymbol{g}, \boldsymbol{0})$ is upper triangular and $T^*(Q, \boldsymbol{c}, \boldsymbol{g})$ is full column rank. ■

Proof. [Theorem 6]

Sufficiency. Suppose Conditions D1 and D2 hold. Recall that WLOG, we assume the first K items fulfill Condition D1 for the K attributes, specifically with the k-th item measuring the k-th attribute and perhaps a subset of its prerequisites. For any item $j^* > K$, let $\boldsymbol{x}^{(j^*)}$ be a J-dimensional vector indicating the items amongst the first K corresponding to j^*-th item's required, so that for $j \leq K$, $x_j^{(j^*)} = I(q_{j^*,j} = 1)$. Then, for any two sets of parameters $(\hat{\boldsymbol{c}}, \hat{\boldsymbol{\pi}})$ and $(\bar{\boldsymbol{c}}, \bar{\boldsymbol{\pi}})$ s.t. $T(Q, \hat{\boldsymbol{c}}, \boldsymbol{g})\hat{\boldsymbol{\pi}} = T(Q, \bar{\boldsymbol{c}}, \boldsymbol{g})\bar{\boldsymbol{\pi}}$,

$$\hat{c}_{j^*} - g_{j^*} = \frac{T_{\boldsymbol{x}^{(j^*)}+\boldsymbol{e}_{j^*}}(Q, \hat{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\hat{\boldsymbol{\pi}}}{T_{\boldsymbol{x}^{(j^*)}}(Q, \hat{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\hat{\boldsymbol{\pi}}} = \frac{T_{\boldsymbol{x}^{(j^*)}+\boldsymbol{e}_{j^*}}(Q, \bar{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\bar{\boldsymbol{\pi}}}{T_{\boldsymbol{x}^{(j^*)}}(Q, \bar{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\bar{\boldsymbol{\pi}}} = \bar{c}_{j^*} - g_{j^*}.$$

This holds because for any $(\boldsymbol{x}, \boldsymbol{c}, \boldsymbol{\pi})$,

$$T_{\boldsymbol{x},\boldsymbol{c},\boldsymbol{\pi}}(Q, \boldsymbol{c} - \boldsymbol{g}, \boldsymbol{0})\boldsymbol{\pi} = \prod_{j:x_j=1} (c_j - g_j) \sum_{\boldsymbol{\alpha}:\xi_j(Q,\boldsymbol{\alpha})=1 \forall j:x_j=1} \pi_{\boldsymbol{\alpha}},$$

and by construction of $\boldsymbol{x}^{(j^*)}$, $\xi_j(Q, \boldsymbol{\alpha}) = 1 \ \forall j : x^{(j^*)} = 1 \Rightarrow \xi_{j^*}(Q, \boldsymbol{\alpha}) = 1$. Then,

$$\sum_{\boldsymbol{\alpha}:\xi_j(Q,\boldsymbol{\alpha})=1 \forall j:x_j^{(j^*)}=1} \pi_{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\alpha}:\xi_j(Q,\boldsymbol{\alpha})=1 \forall j:x_j^{(j^*)}+\boldsymbol{e}_j=1} \pi_{\boldsymbol{\alpha}}$$

and the conclusion follows.

For items $j^* \leq K$, there must be some item $j^{**} \neq j^*$ measuring either the j-th attribute or an attribute requiring the j-th attribute. For terminal attributes, this follows from Condition D2. For non-terminal attributes, this is a consequence of Condition D1. Then, for any two sets of parameters $(\hat{\boldsymbol{c}}, \hat{\boldsymbol{\pi}})$ and $(\bar{\boldsymbol{c}}, \bar{\boldsymbol{\pi}})$ s.t. $T(Q, \hat{\boldsymbol{c}}, \boldsymbol{g})\hat{\boldsymbol{\pi}} = T(Q, \bar{\boldsymbol{c}}, \boldsymbol{g})\bar{\boldsymbol{\pi}}$,

$$\hat{c}_{j^*} - g_{j^*} = \frac{T_{\boldsymbol{e}_{j^{**}}+\boldsymbol{e}_{j^*}}(Q, \hat{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\hat{\boldsymbol{\pi}}}{T_{\boldsymbol{e}_{j^{**}}}(Q, \hat{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\hat{\boldsymbol{\pi}}} = \frac{T_{\boldsymbol{e}_{j^{**}}+\boldsymbol{e}_{j^*}}(Q, \bar{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\bar{\boldsymbol{\pi}}}{T_{\boldsymbol{e}_{j^{**}}}(Q, \bar{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\bar{\boldsymbol{\pi}}} = \bar{c}_{j^*} - g_{j^*}.$$

We need to check that prevalence of profiles $\boldsymbol{\alpha}$ fulfilling the attribute requirements of the j^{**}-th item but not the j^*-th item is zero. Suppose that for some $\boldsymbol{\alpha}$, $\xi_{j^*}(Q, \boldsymbol{\alpha}) = 0$ and $\xi_{j^{**}}(Q, \boldsymbol{\alpha}) = 1$. Since $\xi_{j^*}(Q, \boldsymbol{\alpha}) = 0$, $\alpha_{k_1} = 0$ for some k_1 s.t. $k_1 \trianglelefteq j$. Since $\xi_{j^{**}}(Q, \boldsymbol{\alpha}) = 1$, $\alpha_{k_2} = 1$ for some k_2 s.t. $j \trianglelefteq k_2$. By transitivity, $k_1 \trianglelefteq k_2$. Then, by the rules of attribute hierarchy, $0 = \alpha_{k_1} < \alpha_{k_2} = 1 \Rightarrow \pi_{\boldsymbol{\alpha}} = 0$.

To finish, now that $\hat{c}_j = \bar{c}_j$ for all items j, Condition D1 implies that $\hat{\boldsymbol{\pi}} = \bar{\boldsymbol{\pi}}$.

Necessity. We have shown that Condition D1 is necessary. Suppose it is fulfilled, but Condition D2 is not for some terminal attribute. WLOG let this attribute be the first one. Given any $(\hat{\boldsymbol{c}}, \hat{\boldsymbol{\pi}})$ s.t. $\hat{\boldsymbol{c}} > \boldsymbol{g}$ and $\hat{\pi}_{\boldsymbol{\alpha}} = 0$ iff $\boldsymbol{\alpha}$ is disallowed by the attribute hierarchy, it is possible to find $(\bar{\boldsymbol{c}}, \bar{\boldsymbol{\pi}})$ fulfilling the same restrictions s.t. $T(Q, \hat{\boldsymbol{c}}, \boldsymbol{g})\hat{\boldsymbol{\pi}} = T(Q, \bar{\boldsymbol{c}}, \boldsymbol{g})\bar{\boldsymbol{\pi}}$. To show this, we first set $\hat{c}_j = \bar{c}_j$ for all $j \geq 2$. Divide the set of attribute profiles into $A_0 = \{\boldsymbol{\alpha} : \alpha_1 = 0\}$ and $A_1 = \{\boldsymbol{\alpha} : \alpha_1 = 1\}$. Given any $\bar{c}_1 \in (g_1, \hat{c}_1)$, let

$$\bar{\pi}_{\boldsymbol{\alpha}} = \begin{cases} \frac{\hat{c}_1}{\bar{c}_1}\hat{\pi}_{\boldsymbol{\alpha}} & \boldsymbol{\alpha} \in A_1 \\ \hat{\pi}_{\boldsymbol{\alpha}} + \left(1 - \frac{\hat{c}_1}{\bar{c}_1}\right)\hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} & \boldsymbol{\alpha} \in A_0. \end{cases}$$

Note that attribute profiles $\boldsymbol{\alpha}$ disallowed by the hierarchy still have prevalence $\bar{\pi}_{\boldsymbol{\alpha}} = 0$. For $\boldsymbol{\alpha} \in A_1$ disallowed by the hierarchy, this follows easily from $\hat{\pi}_{\boldsymbol{\alpha}} = 0$. For $\boldsymbol{\alpha} \in A_0$ disallowed by the hierarchy, $\boldsymbol{\alpha} + \boldsymbol{e}_1$ must also be disallowed by the hierarchy since the first attribute is terminal and not a prerequisite for any other attributes. Then, $\hat{\pi}_{\boldsymbol{\alpha}} = \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} = 0 \Rightarrow \bar{\pi}_{\boldsymbol{\alpha}} = 0$. As for the profiles allowed under the hierarchy, there is $\bar{c}_1$ close enough to $\hat{c}_1$ so that $\bar{\pi}_{\boldsymbol{\alpha}} \in (0, 1)$ for all such $\boldsymbol{\alpha}$.

We need to check that $T_{\boldsymbol{x}}(Q, \hat{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\hat{\boldsymbol{\pi}} = T_{\boldsymbol{x}}(Q, \bar{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\bar{\boldsymbol{\pi}}$ for all rows $\boldsymbol{x} \in \{0, 1\}^J$. If $x_1 = 0$, then the first attribute is not required by any of the items denoted by $\boldsymbol{x}$. Then, for $\boldsymbol{\alpha} \in A_0$, $\xi_j(Q, \boldsymbol{\alpha}) = 1 \forall j$ s.t. $x_j = 1$ iff $\xi_j(Q, \boldsymbol{\alpha} + \boldsymbol{e}_1) = 1 \forall j$ s.t. $x_j = 1$. In addition,

$$\hat{\pi}_{\boldsymbol{\alpha}} + \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} = \bar{\pi}_{\boldsymbol{\alpha}} + \bar{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1} \ \forall \boldsymbol{\alpha} \in A_0$$

by construction. Thus,

$$\begin{aligned} T_{\boldsymbol{x}}(Q, \hat{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\hat{\boldsymbol{\pi}} &= \prod_{j:x_j=1} (\hat{c}_j - g_j) \sum_{\boldsymbol{\alpha} \in A_0 : \xi_j(Q,\boldsymbol{\alpha})=1 \forall j : x_j = 1} (\hat{\pi}_{\boldsymbol{\alpha}} + \hat{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1}) \\ &= \prod_{j:x_j=1} (\bar{c}_j - g_j) \sum_{\boldsymbol{\alpha} \in A_0 : \xi_j(Q,\boldsymbol{\alpha})=1 \forall j : x_j = 1} (\bar{\pi}_{\boldsymbol{\alpha}} + \bar{\pi}_{\boldsymbol{\alpha}+\boldsymbol{e}_1}) \\ &= T_{\boldsymbol{x}}(Q, \bar{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\bar{\boldsymbol{\pi}}. \end{aligned}$$

For $\boldsymbol{x}$ s.t. $x_1 = 1$, $t_{\boldsymbol{x},\boldsymbol{\alpha}}(Q, \boldsymbol{c} - \boldsymbol{g}, \boldsymbol{0}) > 0$ only for $\boldsymbol{\alpha} \in A_1$.

$$\begin{aligned}
T_{\boldsymbol{x}}(Q, \hat{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\hat{\boldsymbol{\pi}} &= (\hat{c}_1 - g_1) \prod_{j \geq 2: x_j = 1} (\hat{c}_j - g_j) \sum_{\boldsymbol{\alpha} \in A_1: \xi_j(Q,\boldsymbol{\alpha}) = 1 \forall j: x_j = 1} \hat{\pi}_{\boldsymbol{\alpha}} \\
&= (\bar{c}_1 - g_1) \prod_{j \geq 2: x_j = 1} (\hat{c}_j - g_j) \sum_{\boldsymbol{\alpha} \in A_1: \xi_j(Q,\boldsymbol{\alpha}) = 1 \forall j: x_j = 1} \frac{\hat{c}_1}{\bar{c}_1} \hat{\pi}_{\boldsymbol{\alpha}} \\
&= (\bar{c}_1 - g_1) \prod_{j \geq 2: x_j = 1} (\bar{c}_j - g_j) \sum_{\boldsymbol{\alpha} \in A_1: \xi_j(Q,\boldsymbol{\alpha}) = 1 \forall j: x_j = 1} \bar{\pi}_{\boldsymbol{\alpha}} \\
&= T_{\boldsymbol{x}}(Q, \bar{\boldsymbol{c}} - \boldsymbol{g}, \boldsymbol{0})\bar{\boldsymbol{\pi}}.
\end{aligned}$$

This finishes our proof. ■

A.4 Proofs for Chapter 4

Proof. [Proposition 7] Suppose $\xi_j^1 = \xi_j^2$ for all j such that $1 - s_j \neq g_j$. If $1 - s_j = g_j$, then the response distribution for item j does not depend on $\boldsymbol{\xi}$:

$$\begin{aligned}
p(x_j|\boldsymbol{\xi}, s, g) =& (1 - s_j)^{\xi_j x_j} g_j^{(1-\xi_j)x_j} s_j^{\xi_j(1-x_j)} (1 - g_j)^{(1-\xi_j)(1-x_j)} \\
=& (1 - s_j)^{x_j} s_j^{1-x_j} = g_j^{x_j} (1 - g_j)^{1-x_j}.
\end{aligned}$$

Thus, for every $\boldsymbol{x} \in \{0, 1\}^J$,

$$\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{\xi}^1, s, g) =& \prod_{j=1}^{m} p(x_j|\xi_j^1, s_j, g_j) \\
=& \prod_{\{i: 1 - s_j = g_j\}} p(x_j|\xi_j^1, s_j, g_j) \prod_{\{i: 1 - s_j \neq g_j\}} p(x_j|\xi_j^1, s_j, g_j) \\
=& \prod_{\{i: 1 - s_j = g_j\}} p(x_j|\xi_j^2, s_j, g_j) \prod_{\{i: 1 - s_j \neq g_j\}} p(x_j|\xi_j^2, s_j, g_j) = p(\boldsymbol{x}|\boldsymbol{\xi}^2, s, g)
\end{aligned}$$

and $\boldsymbol{\alpha}^1$ cannot be separated from $\boldsymbol{\alpha}^2$.

Now suppose that $\xi_j^1 \neq \xi_j^2$ for some j such that $1 - s_j \neq g_j$. Then

$$\Pr(x_j = 1|\boldsymbol{\xi}^1, s, g) = (1-s_j)^{\xi_j^1} g_j^{1-\xi_j^1} \neq g_j^{\xi_j^1}(1-s_j)^{1-\xi_j^1} = g_j^{1-\boldsymbol{\xi}^2}(1-s_j)^{\xi_j^2} = \Pr(x_j = 1|\boldsymbol{\xi}^2, s, g),$$

so the response distributions differ. ■

Proof. [Proposition 9] Suppose that Q is complete. WLOG, for $j = 1, \ldots, K$ let $\boldsymbol{q}^j = e_j$. Then, for $j = 1, \ldots, K$, $\xi_j(Q, \boldsymbol{\alpha}) = \alpha_j$ and given any two attribute profiles $\boldsymbol{\alpha}^1 \neq \boldsymbol{\alpha}^2$,

$$\xi_{1:K}(Q, \boldsymbol{\alpha}^1) = \boldsymbol{\alpha}^1 \neq \boldsymbol{\alpha}^2 = \xi_{1:K}(Q, \boldsymbol{\alpha}^2)$$

By Proposition 7, Q separates any $\boldsymbol{\alpha}^1 \neq \boldsymbol{\alpha}^2$.

Now suppose that $\exists k_* \in \{1, \ldots, K\}$ such that $e_{k_*} \notin R_Q$. WLOG, suppose $k_* = 1$. Consider profiles $\boldsymbol{\alpha} = e_1$ and $\boldsymbol{\alpha}' = \mathbf{0}$, the zero column-vector. For each item $j = 1, \ldots, J$, if $q_{j,1} = 0$ then

$$\xi_j(Q, e_1) = \left(1^0\right) \prod_{k \neq 1} 0^{q_{j,k}} = \left(0^0\right) \prod_{k \neq 1} 0^{q_{j,k}} = \xi_j(Q, \mathbf{0}).$$

Else, $q_{j,1} = 1$ and there exists some $k_{**} \neq 1$ such that $q_{j,k_{**}} = 1$ and

$$\xi_j(Q, e_1) = \left(0^1\right) \prod_{k \neq k_{**}} [\mathbf{1}(k = 1)]^{q_{j,k}} = 0 = \left(0^1\right) \prod_{k \neq k_{**}} 0^{q_{j,k}} = \xi_j(Q, \mathbf{0}).$$

Thus, $\boldsymbol{\xi}(Q, e_1) = \boldsymbol{\xi}(Q, \mathbf{0})$ and by Proposition 7, attribute profiles e_1 and $\mathbf{0}$ cannot be separated. ■

Proof. [Theorem 11] We can write the likelihood

$$L(\boldsymbol{\pi}) = p(X|\boldsymbol{\pi}) = \prod_{i=1}^{N} p(\boldsymbol{x}^i|\boldsymbol{\pi})$$

and the log-likelihood as

$$\ell(\boldsymbol{\pi}) = \sum_{i=1}^{N} \log p(\boldsymbol{x}^i|\boldsymbol{\pi}) = \sum_{\boldsymbol{x}\in\{0,1\}^J} N_{\boldsymbol{x}} \log\left(\sum_{[\boldsymbol{\alpha}]} p\left(\boldsymbol{x}\,|[\boldsymbol{\alpha}]\,\right)\pi_{[\boldsymbol{\alpha}]}\right),$$

where $N_{\boldsymbol{x}} = \#\{i : \boldsymbol{x}^i = \boldsymbol{x}\}$.

Suppose there are a total of L distinct equivalence classes partitioning the attribute profile space. Rather than indexing the column vectors of the T-matrix by the attribute profiles $\boldsymbol{\alpha}$, we index by the attribute profile equivalence classes $[\boldsymbol{\alpha}]$. Then the T-matrix is the $2^J \times L$ matrix $T = (t_{\boldsymbol{x},[\boldsymbol{\alpha}]})$, indexed over all response vectors $\boldsymbol{x}$ and equivalence classes $[\boldsymbol{\alpha}]$, such that $t_{\boldsymbol{x},[\boldsymbol{\alpha}]} = \Pr(\boldsymbol{X} \geq \boldsymbol{x}|[\boldsymbol{\alpha}])$. This is well defined, since $t_{\boldsymbol{x},\boldsymbol{\alpha}}$ is constant over $\boldsymbol{\alpha}$ in the same equivalence class. The survival probabilities for each vector $\boldsymbol{x}$ can be calculated via matrix multiplication as $T(Q,\boldsymbol{c},\boldsymbol{g})\boldsymbol{\pi}$, and we have identifiability iff $T(Q,\boldsymbol{c},\boldsymbol{g})\hat{\boldsymbol{\pi}} = T(Q,\boldsymbol{c},\boldsymbol{g})\bar{\boldsymbol{\pi}} \Rightarrow \hat{\boldsymbol{\pi}} = \bar{\boldsymbol{\pi}}$. Then identifiability is equivalent to T being a rank L matrix.

First, suppose that $g \equiv 0$. WLOG, it can be assumed that the L equivalence classes $[\boldsymbol{\alpha}^1], \ldots, [\boldsymbol{\alpha}^L]$ are ordered lexicographically by their minimal representatives, $\boldsymbol{\alpha}^{1*}, \ldots, \boldsymbol{\alpha}^{L*}$. Thus, if $\boldsymbol{\alpha}^{k*} \geq \boldsymbol{\alpha}^{\ell*}$, then $k \geq \ell$. Also, let $\boldsymbol{x}^\ell = \boldsymbol{\xi}(Q,[\boldsymbol{\alpha}^\ell])$ for $\ell \in \{1,\ldots,L\}$. Define $T^* = (t^*_{k,\ell}$, where $t^*_{k,\ell} = t_{\boldsymbol{x}^k,[\boldsymbol{\alpha}^\ell]}$. Then T^* is an $L \times L$ sub-matrix of T, containing the specified rows $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^L$. Moreover, T^* is an upper triangular matrix. This is a consequence of the fact that for any $k > \ell$, $\boldsymbol{\alpha}^{\ell*} \not\geq \boldsymbol{\alpha}^{k*}$. Thus, there must be some item $j \in \{1,\ldots,J\}$ for which individuals with profiles $\boldsymbol{\alpha} \in [\boldsymbol{\alpha}^k]$ possess the necessary attributes, but individuals with profiles $\boldsymbol{\alpha} \in [\boldsymbol{\alpha}^\ell]$ do not. Then $\Pr(X_j = 1|[\boldsymbol{\alpha}^\ell]) = g_j = 0 \Rightarrow t_{\boldsymbol{x}^k,[\boldsymbol{\alpha}^\ell]} = t^*_{k,\ell} = 0$. In addition, on the diagonal, $t^*_{\ell,\ell} = \prod_{\{i:x^\ell_j=1\}}(1-s_j) \neq 0$. Thus, T^* is a rank L matrix, as is T.

Next suppose that $g \not\equiv 0$. Consider the T matrix as a function of the item probabilities of obtaining a correct answer given sufficient knowledge, $\boldsymbol{c} = 1 - \boldsymbol{s}$

and the probabilities given insufficient knowledge, also known as the guessing vector $\boldsymbol{g}$. Then the T-matrix $T(\boldsymbol{c}, \boldsymbol{g})$ can be written as a linear transformation of another T-matrix $T(\boldsymbol{c} - \boldsymbol{g}, \boldsymbol{0})$. For any subset of the items S and any constants b_j,

$$\prod_{j \in S}(b_j - g_j) = \prod_{j \in S} b_j - \sum_{j \in S} g_j \prod_{k \neq j} b_k + \sum_{k \neq j \in S} g_j g_k \prod_{\ell \neq j,k} b_\ell - \cdots + (-1)^{\#S} \prod_{j \in S} g_j$$

In the case of the entries of the T-matrix, the b_j will correspond to either c_j or g_j, depending on the value of $\xi_j(Q, [\boldsymbol{\alpha}])$. Using this relationship, it is possible to write $T(\boldsymbol{c} - \boldsymbol{g}, 0) = D(\boldsymbol{g}) \cdot T(\boldsymbol{c}, \boldsymbol{g})$, where the transformation matrix $D(g)$ is a $2^J \times 2^J$ matrix depending solely on $\boldsymbol{g}$. Since the rows of T are ordered lexicographically by $\boldsymbol{x}$, $D(g)$ is a lower triangular matrix with diagonal $diag(D) \equiv 1$. Thus, D is a full-rank matrix and $\text{rank}(T(\boldsymbol{c}, \boldsymbol{g})) = \text{rank}(T(\boldsymbol{c} - \boldsymbol{g}, \boldsymbol{0})) = L$. The identifiability condition has been fulfilled, and all other conditions for the consistency of the maximum likelihood estimator are clearly evident. ∎

Proof. [Proposition 12] Since

$$\sum_{\{\boldsymbol{\alpha}:\delta_{[\boldsymbol{\alpha}],k}=1\}} \pi_{\boldsymbol{\alpha}} = \sum_{\{[\boldsymbol{\alpha}]:\delta_{[\boldsymbol{\alpha}],k}=1\}} \sum_{\boldsymbol{\alpha}' \in [\boldsymbol{\alpha}]} \pi_{\boldsymbol{\alpha}} = \sum_{\{[\boldsymbol{\alpha}]:\delta_{[\boldsymbol{\alpha}],k}=1\}} \pi_{[\boldsymbol{\alpha}]},$$

ζ_k can be written in terms of $\pi_{[\boldsymbol{\alpha}]}$ as

$$\zeta_k = \sum_{\{[\boldsymbol{\alpha}]:\delta_{[\boldsymbol{\alpha}],k}=1\}} \pi_{[\boldsymbol{\alpha}]}.$$

By Theorem 11, the MLE $\hat{\pi}_{[\boldsymbol{\alpha}]}$ is consistent as $N \to \infty$ under the conditions of the proposition. Thus, $\hat{\zeta}_k$ is consistent as $N \to \infty$. ∎