

NEW PARADIGM OF IDENTIFIABLE GENERAL-RESPONSE COGNITIVE DIAGNOSTIC MODELS: BEYOND CATEGORICAL DATA

SEUNGHYUN LEE[✉]

COLUMBIA UNIVERSITY

YUQI GU[✉]

COLUMBIA UNIVERSITY

Cognitive diagnostic models (CDMs) are a popular family of discrete latent variable models that model students' mastery or deficiency of multiple fine-grained skills. CDMs have been most widely used to model categorical item response data such as binary or polytomous responses. With advances in technology and the emergence of varying test formats in modern educational assessments, new response types, including continuous responses such as response times, and count-valued responses from tests with repetitive tasks or eye-tracking sensors, have also become available. Variants of CDMs have been proposed recently for modeling such responses. However, whether these extended CDMs are identifiable and estimable is entirely unknown. We propose a very general cognitive diagnostic modeling framework for arbitrary types of multivariate responses with minimal assumptions, and establish identifiability in this general setting. Surprisingly, we prove that our general-response CDMs are identifiable under Q-matrix-based conditions similar to those for traditional categorical-response CDMs. Our conclusions set up a new paradigm of identifiable general-response CDMs. We propose an EM algorithm to efficiently estimate a broad class of exponential family-based general-response CDMs. We conduct simulation studies under various response types. The simulation results not only corroborate our identifiability theory, but also demonstrate the superior empirical performance of our estimation algorithms. We illustrate our methodology by applying it to a TIMSS 2019 response time dataset.

Key words: cognitive diagnostic model, diagnostic classification model, EM algorithm, exponential family, general responses, generalized linear model, identifiability, Q-matrix.

1. Introduction

Cognitive diagnostic models (CDMs, also called diagnostic classification models; see Rupp et al. 2010) are a popular family of discrete latent variable models in educational and psychological measurement. CDMs employ multiple discrete latent variables to model and diagnose students' mastery or deficiency of a set of fine-grained skills. Popular examples in the literature include the Deterministic Input Noisy output "And" gate model (DINA model; Junker and Sijtsma, 2001), the log-linear CDM (LCDM; Henson et al., 2009), the additive CDM (ACDM; de la Torre, 2011), and general diagnostic models (GDM; von Davier, 2008).

Originally, CDMs were developed mostly to model binary responses and later generalized to polytomous responses, which are both categorical. Recently, with advances in technology and the emergence of varying test formats in modern educational assessments, new response types have become available. In particular, *multivariate continuous* and *count* responses are especially common. Continuous responses arise in the following scenarios: (a) responses that place a mark

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11336-024-09983-4>.

Correspondence should be made to Yuqi Gu, Department of Statistics, Columbia University, Room 928 SSW, 1255 Amsterdam Avenue, New York, NY 10027, USA. Email: yuqi.gu@columbia.edu

on a line segment (such as the visual analog scale), (b) assessments that record the probability of each option being correct (probability testing), (c) computer-based tests that record the response time (Minchen et al., 2017). In particular, modeling response times has received great interest for a long time, and many different models have been proposed to this end; see De Boeck and Jeon (2019) for an overview. Another common response type is count responses. They arise in the following scenarios: (a) assessments with repetitive tasks where the number of correct responses is recorded, (b) assessments where examinees read aloud a passage and the number of errors is recorded, (c) modern exams with eye tracking sensors that record students' visual fixation counts, (d) computer-based tests that record the visit count per item (Man & Harring, 2019; Liu et al., 2022). Rasch (1993) first proposed a Poisson-based item response theory (IRT) model for count data, and many other models have been proposed (Magnus & Thissen, 2017; Man & Harring, 2019, 2022).

Many existing latent variable models for general responses are based on IRT that use continuous latent traits to model the unobserved constructs (Thissen, 1983; van der Linden, 2007; Wang & Xu, 2015). On the other hand, using discrete latent variables as in CDMs can provide students with valuable personalized diagnoses of the mastery/deficiency profiles of the latent skills. Currently, there are a few CDMs proposed for modeling non-categorical responses. For instance, Minchen et al. (2017) proposed a DINA model with a lognormal link to model response time data, whereas Liu et al. (2022, 2023) proposed GDMs (which include the DINA model as a submodel) with a Poisson link and a negative binomial link, respectively, to model visual fixation count data. It is desirable to propose a general modeling framework of CDMs for flexible types of responses and their associated estimation methods.

When proposing new statistical models, identifiability is a crucial consideration, because it is a fundamental prerequisite for valid statistical estimation and inference. A model is identifiable if the parameters can be uniquely recovered from the observed data distribution. In recent years, many identifiability results have been established for the categorical-response CDMs, including both binary and polytomous cases (e.g., Chen et al., 2015; Xu and Zhang, 2016; Fang et al., 2019; Gu and Xu, 2019, 2020; Culpepper, 2019, 2023). These studies typically show that CDMs for multivariate categorical data are identifiable under structural constraints on the **Q**-matrix. The **Q**-matrix here is a key component in a CDM that specifies how the observed responses depend on the latent attributes (see its formal definition in Sect. 2). However, it is entirely unknown whether or not those extended CDMs for continuous or count data (such as those in Minchen et al. (2017) and Liu et al. (2022)) are identifiable, let alone identifiability of models for more general response types.

This manuscript makes the following contributions. *First*, we propose a very general new framework of **Q**-matrix-based CDMs for modeling rich types of responses. In particular, this framework includes a sub-family of *exponential family-based CDMs* (ExpCDMs), which is a wide class of parametric CDMs using exponential families to model general responses. Our general modeling framework covers existing CDMs for continuous and count responses as special cases (Minchen et al., 2017; Minchen & de la Torre, 2018; Liu et al., 2022, 2023). *Second*, we provide the crucial identifiability guarantees for the proposed new models. Somewhat surprisingly, we prove that our general model is identifiable under similar structural conditions on the **Q**-matrix, just as those for traditional categorical-response CDMs. This is the first identifiability result for CDMs with non-categorical responses. Our conclusions set up a new paradigm of identifiable general-response CDMs and significantly advances the psychometric theory of diagnostic modeling. *Third*, we propose an EM algorithm to efficiently estimate model parameters. As concrete demonstrations, we consider the DINA and main-effect-based CDMs for general responses and derive explicit updates in the EM algorithms. Our simulation results corroborate the identifiability theory, and also demonstrate the superior empirical performance of the proposed estimation algorithms.

The remainder of this manuscript is organized as follows. Section 2 formally introduces our general-response CDM framework and gives many examples of its parametric submodels. Section 3 provides conditions for model identifiability. Section 4 proposes a general-purpose EM algorithm to estimate the model parameters. Section 5 presents simulation studies under various response types. Section 6 illustrates our methodology via application to a real-world response time dataset from the Trends in International Mathematics and Science Study (TIMSS) in 2019. Section 7 concludes and discusses future research directions. The Supplementary Material contains the proofs of the theorems and more details of the simulation studies and real data analysis.

2. General-Response Cognitive Diagnostic Models

2.1. General Model Setup

Consider an educational assessment with J items, and denote an observed response vector by $\mathbf{Y} = (Y_1, \dots, Y_J) \in \times_{j=1}^J \mathcal{Y}_j$, where each \mathcal{Y}_j is the sample space of the random variable Y_j . In particular, our main examples are (a) continuous responses with $\mathcal{Y}_j = \mathbb{R}$, and (b) count responses with $\mathcal{Y}_j = \{0, 1, 2, \dots\}$ being the set of all nonnegative integers. Beyond continuous and count responses, our general modeling framework covers a much wider class of responses with a general sample space \mathcal{Y}_j . This includes vector-valued responses that arise in the context of modern assessment data, such as a joint vector of response accuracy, response time, and visual fixation counts (De Boeck & Jeon, 2019; Man & Harring, 2022). Another example of responses is continuous features learned from process data via dimension reduction methods (He & Von Davier, 2016; Tang et al., 2020). For a more rigorous presentation with minimal assumptions on \mathcal{Y}_j , measure-theoretic definitions are required, and we provide those details in Supplementary Material S.1. Note that we also allow the response type \mathcal{Y}_j to differ across $j = 1, \dots, J$, that is, mixed-type responses (e.g., see Moustaki and Knott, 2000).

To define the general-response CDMs, we start by specifying the latent part. We model each student's latent skill profile \mathbf{A} as a binary K -dimensional vector $\mathbf{A} = (A_1, \dots, A_K) \in \{0, 1\}^K$ for diagnostic modeling purposes. Here $A_k = 1$ or 0 represents the presence or absence of the k th latent skill. Following most existing studies on CDMs (Chen et al., 2015, 2018), we adopt the saturated model for the latent attributes; that is, for each skill profile $\boldsymbol{\alpha} \in \{0, 1\}^K$, define its population proportion parameter as $p_{\boldsymbol{\alpha}} = \mathbb{P}(\mathbf{A} = \boldsymbol{\alpha})$, which satisfy

$$\sum_{\boldsymbol{\alpha} \in \{0, 1\}^K} p_{\boldsymbol{\alpha}} = 1, \quad p_{\boldsymbol{\alpha}} > 0. \quad (1)$$

We collect all proportion parameters in the vector $\mathbf{p} = (p_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \{0, 1\}^K)$.

Next, we define the measurement model, which specifies the conditional distribution of the observed responses \mathbf{Y} given the latent skills \mathbf{A} . The main ingredients in this definition are (a) the local independence assumption, and (b) constraints induced by the \mathbf{Q} -matrix (Tatsuoka, 1983). Let $\mathbb{P}_{j, \boldsymbol{\alpha}}(Y_j | \mathbf{A} = \boldsymbol{\alpha})$ (or $\mathbb{P}_{j, \boldsymbol{\alpha}}$ for short) denote the conditional distribution of Y_j given $\mathbf{A} = \boldsymbol{\alpha}$. For any positive integer M , let $[M] := \{1, 2, \dots, M\}$ denote the set of all positive integers no greater than M . First, under local independence, the observed Y_1, \dots, Y_J are conditionally independent given the latent A_1, \dots, A_K , which implies:

$$\mathbb{P}(\mathbf{Y} \in \times_{j=1}^J S_j | \mathbf{A} = \boldsymbol{\alpha}) = \prod_{j=1}^J \mathbb{P}_{j, \boldsymbol{\alpha}}(Y_j \in S_j | \mathbf{A} = \boldsymbol{\alpha}), \quad \forall S_j \in \mathcal{F}_j, \quad j \in [J]. \quad (2)$$

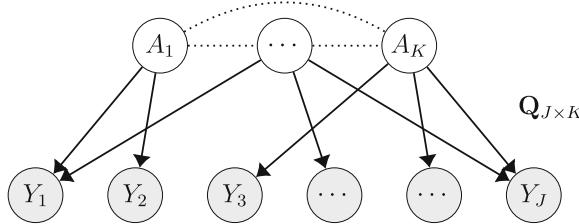


FIGURE 1.

Graphical model of the general-response CDM with $Y_j \in \mathcal{Y}_j$. White nodes are latent attributes, and gray nodes are observed responses. The directed arrows from the latent to the observed capture the conditional dependence of \mathbf{Y} given \mathbf{A} , which is exactly encoded in the \mathbf{Q} -matrix $\mathbf{Q}_{J \times K}$. There is an directed arrow from A_k to Y_j if and only if $q_{j,k} = 1$.

Here, the notation $\mathbf{Y} \in \times_{j=1}^J S_j$ means that $Y_j \in S_j$ for all j , and \mathcal{F}_j is a collection of measurable subsets of the sample space \mathcal{Y}_j that determines a probability distribution on \mathcal{Y}_j . For continuous responses, \mathcal{F}_j can be the collection of all open intervals in \mathbb{R} ; for count responses, \mathcal{F}_j can be the collection of all subsets of \mathcal{Y}_j .

Next, we describe the constraints on the conditional distributions $\mathbb{P}_{j,\alpha}$ imposed by the \mathbf{Q} -matrix. The \mathbf{Q} -matrix was initially proposed by Tatsuoka (1983) for modeling binary responses using the cognitive diagnostic assumption. The \mathbf{Q} -matrix $\mathbf{Q} = (q_{j,k})$ is a $J \times K$ matrix with binary entries, where the (j, k) -th entry $q_{j,k}$ equals 1 if item j requires or measures the latent attribute k , and 0 otherwise. Statistically, the entries in the \mathbf{Q} -matrix describe how the J observed item responses depend on the latent attributes. For convenience, we also use the notation $\text{pa}(j) = \{k \in [K] : q_{j,k} = 1\}$ to represent the set of attributes required/measured by item j . Here “pa” is short for “parent”, motivated by the graphical model perspective of the proposed model; see Remark 1 below for details. We write $\alpha_{\text{pa}(j)}$ as the sub-vector of $\alpha = (\alpha_1, \dots, \alpha_K)$ that contains those entries indexed by $k \in \text{pa}(j)$. Then, by the definition of the \mathbf{Q} -matrix, the distribution $\mathbb{P}_{j,\alpha}(Y_j \in S_j \mid \mathbf{A} = \alpha, \mathbf{Q})$ must satisfy

$$\mathbb{P}_{j,\alpha}(Y_j \in S_j \mid \mathbf{A} = \alpha, \mathbf{Q}) = \mathbb{P}_{j,\alpha}(Y_j \in S_j \mid \mathbf{A}_{\text{pa}(j)} = \alpha_{\text{pa}(j)}), \quad \forall S_j \in \mathcal{F}_j; \quad (3)$$

that is, the conditional distribution of Y_j given all the latent attributes is the same as the conditional distribution of Y_j given the required attributes of item j as indicated by the \mathbf{Q} -matrix. To make the \mathbf{Q} -matrix play a meaningful role in describing the conditional distribution of the observed variables given the latent variables, we make a mild assumption that for each $j \in [J]$, there exist latent patterns α and α' with $\alpha_{\text{pa}(j)} \neq \alpha'_{\text{pa}(j)}$ such that $\mathbb{P}_{j,\alpha} \neq \mathbb{P}_{j,\alpha'}$. In the special case of binary response CDMs, this assumption is readily satisfied by many popular existing models. In this work, we consider the confirmatory modeling setting where the \mathbf{Q} -matrix is known.

Remark 1. Viewed from a probabilistic graphical modeling perspective, the \mathbf{Q} -matrix can also be represented by a bipartite graph with directed arrows pointing from the latent variables to the observed ones; see Fig. 1 for a graphical model illustration. Then, $\text{pa}(j)$ exactly denotes the *parent* attributes that have a directed arrow toward Y_j , and Assumption (3) can be interpreted as the graphical model being faithful (e.g., see Definition 3.8 in Koller and Friedman, 2009). For example, in the setting of Fig. 1, A_1 is the only parent attribute of Y_2 , thus $\text{pa}(2) = \{1\}$.

Now, we formally define the general-response CDM based on the above assumptions.

Definition 1. (General-response CDM) A general-response CDM with K binary latent attributes $\mathbf{A} = (A_1, \dots, A_K)$, J observed responses $\mathbf{Y} = (Y_1, \dots, Y_J)$, and model components $(\mathbf{p}, \{\mathbb{P}_{j,\alpha}\})$ is a statistical model for \mathbf{Y} that satisfies (1), (2), and (3).

Under Definition 1, the marginal distribution of \mathbf{Y} in a general-response CDM is:

$$\mathbb{P}(\mathbf{Y} \in \times_{j=1}^J S_j \mid \mathbf{Q}) = \sum_{\alpha \in \{0, 1\}^K} p_\alpha \prod_{j=1}^J \mathbb{P}_{j, \alpha}(Y_j \in S_j \mid \mathbf{A}_{\text{pa}(j)} = \alpha_{\text{pa}(j)}), \quad \forall S_j \in \mathcal{F}_j. \quad (4)$$

The proposed general-response CDMs form a very general and flexible framework of cognitive diagnostic models. *First*, this framework includes existing popular CDMs such as the DINA model, Reduced-RUM, LCDM, and GDM for categorical responses as submodels. This is because one can let the conditional distribution $\mathbb{P}_{j, \alpha}(Y_j \mid \mathbf{A} = \alpha)$ be a specific parametric distribution such as the Bernoulli with a certain link function for modeling binary responses (elaborated in the next paragraph). *Second*, our framework also includes those recently proposed specific CDMs for continuous or count responses (e.g., Minchen and de la Torre, 2018; Liu et al., 2022) as special cases, and is further able to model other response types.

As a concrete illustration, we next present the special case of our model (4) when the responses are all binary, which is the most widely considered scenario for CDMs. In this case, $\mathcal{Y}_j = \{0, 1\}$ for all $j \in [J]$ and $\mathbb{P}_{j, \alpha}$ is just the Bernoulli distribution. We write the Bernoulli parameter of $\mathbb{P}_{j, \alpha}$ as $\theta_{j, \alpha}$. Note that in the classical binary-response CDM literature, $\theta_{j, \alpha}$ is often called the positive response probability. The conditional independence condition (3) boils down to the following equality constraints on the parameters $\theta_{j, \alpha}$:

$$\theta_{j, \alpha} = \theta_{j, \alpha'}, \quad \forall \alpha, \alpha' \in \{0, 1\}^K \text{ such that } \alpha_{\text{pa}(j)} = \alpha'_{\text{pa}(j)}. \quad (5)$$

The above constraints are satisfied in all existing CDMs for binary responses, and they also appeared as assumptions in the study of identifiability for the related restricted latent class models (Xu, 2017; Gu & Xu, 2020). As a more concrete example, $\theta_{j, \alpha}$ under the DINA model with the slipping parameter s_j and guessing parameter g_j can be written as

$$\theta_{j, \alpha} = (1 - s_j) \prod_{k=1}^K \alpha^{q_{j,k}} + g_j (1 - \prod_{k=1}^K \alpha^{q_{j,k}}) = (1 - s_j) \prod_{k \in \text{pa}(j)} \alpha_k + g_j (1 - \prod_{k \in \text{pa}(j)} \alpha_k),$$

where $1 - s_j > g_j$ is typically assumed. This parametrization clearly satisfies (5).

2.2. Parametric and Exponential Family-Based CDMs (ExpCDMs)

We next define parametric general-response CDMs, and illustrate how our general framework can be specified according to various modeling assumptions and response types. In this section, for notational simplicity, we additionally assume that the response types are the same across the items (i.e., $(\mathcal{Y}_j, \mathcal{F}_j)$ is the same for all $j \in [J]$) and follow the same parametric family $\mathcal{P} = \{g(\cdot; \eta)\}$. Under this assumption, we can further write $\mathbb{P}_{j, \alpha}$ as

$$\mathbb{P}_{j, \alpha}(Y_j \in S_j \mid \mathbf{A} = \alpha) = \begin{cases} \int_{S_j} g(y; \eta_{j, \alpha}) dy, & \text{if } \mathcal{Y}_j = (a, b) \text{ or } \mathbb{R}, \\ \sum_{y \in S_j} g(y; \eta_{j, \alpha}), & \text{if } \mathcal{Y}_j = \{0, 1, 2, \dots\}. \end{cases} \quad (6)$$

Here, $g(\cdot; \eta)$ is the probability density/mass function of a specific identifiable parametric family \mathcal{P} (that is, $\eta \rightarrow g(\cdot; \eta)$ is a one-to-one mapping), and $\eta_{j, \alpha}$'s are parameters that can be a scalar or a vector. We next formally define a parametric CDM.

TABLE 1.
Examples of exponential families, and their natural parameters and sufficient statistics.

| Distribution family | Natural parameter η | Sufficient statistics $\mathbf{T}(Y)$ |
|------------------------------|---|---------------------------------------|
| Bernoulli(θ) | $\log\left(\frac{\theta}{1-\theta}\right)$ | Y |
| Normal(μ, σ^2) | $\left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$ | (Y, Y^2) |
| lognormal(μ, σ^2) | $\left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$ | $(\log Y, (\log Y)^2)$ |
| Poisson(λ) | $\log \lambda$ | Y |

Definition 2. (Parametric general-response CDM) Given an identifiable parametric family $\mathcal{P} = \{g(\cdot; \eta)\}$, the parametric general-response CDM with parameters $(\{\eta_{j,\alpha}\}, \mathbf{p})$ is a general-response CDM whose $\mathbb{P}_{j,\alpha}$ satisfies (6).

For parametric general-response CDMs, the \mathbf{Q} -matrix constraints (3) reduce to the following constraints on the parameters η , which are analogous to (5) in the binary-response case:

$$\eta_{j,\alpha} = \eta_{j,\alpha'}, \quad \forall \alpha, \alpha' \in \{0, 1\}^K \text{ such that } \alpha_{\text{pa}(j)} = \alpha'_{\text{pa}(j)}. \quad (7)$$

In the context of modeling response times in psychometrics, the Gamma (Maris, 1993), Weibull (Loeys et al., 2011), and inverse Gaussian (Lo & Andrews, 2015) distributions have been commonly used in addition to the aforementioned lognormal distribution. In the context of modeling visual fixation counts or correct answer counts, the negative binomial distribution is a popular choice (Man & Harring, 2019; Liu et al., 2023) in addition to the aforementioned Poisson distribution (Liu et al., 2022). In the later sections on estimation methodology and simulation studies, we will consider a variety of distributions for \mathcal{P} : Normal, transformed-Normal (lognormal, logistic-Normal), Poisson, and negative binomial.

Except for the negative binomial distribution, all parametric distributions mentioned in the previous paragraph are *exponential family distributions* (e.g., see Section 3.4 in Casella and Berger, 2021, for which the probability density/mass function can be written as:

$$g(y_j; \eta) = h(y_j) \exp\{\eta^\top \mathbf{T}(y_j) - A(\eta)\}. \quad (8)$$

Following the convention for the exponential family distributions, η collects the natural parameters, $\mathbf{T}(Y_j)$ collects the sufficient statistics, and $A(\eta)$ is the log-partition function. Both η and $\mathbf{T}(Y_j)$ are typically multidimensional in our general-response CDM context. Exponential families are a natural and flexible choice to model rich types of responses and are widely used in statistics and psychometrics. For example, the Bernoulli distribution with success parameter θ , commonly used to model binary responses, is an exponential family with natural parameter $\eta = \log(\theta/(1-\theta))$ and sufficient statistic $T(Y) = Y$. In Table 1, we also present the natural parameter and sufficient statistics for the exponential families that we will later use to model general responses.

We next define parametric general-response CDMs based on the exponential families; we call these models *ExpCDMs*.

Definition 3. (ExpCDM) We define an ExpCDM as a parametric CDM whose parametric family $\mathcal{P} = \{g(\cdot; \boldsymbol{\eta})\}_{\boldsymbol{\eta}}$ is an exponential family distribution.

The general framework of ExpCDMs encompasses models for various response types including categorical, count, bounded and unbounded continuous data. In the end of this section, we will also briefly mention how to deal with even more general distributions outside of the exponential family, such as the negative binomial.

Remark 2. Using exponential family distributions to model general-response multivariate data with latent variables has a long history. Moustaki and Knott (2000) extended the IRT model to general responses, by modeling the natural parameter of the exponential family distribution as the linear combination of the continuous latent traits. This approach extends the generalized linear model (GLM) in Nelder and Wedderburn (1972) to the latent variable setting. In a similar spirit, Dunson (2000) used continuous latent variables and exponential family distributions to model multiple clustered outcomes.

In the context of factor analysis, Skrondal and Rabe-Hesketh (2004) extended the linear factor analysis to model general responses by using exponential families and proposed the generalized linear factor model. We emphasize that all these previous models use continuous latent variables, and there do not exist works that adopt a framework as flexible as the exponential family with multidimensional discrete latent variables that serve the diagnostic purposes.

To define a general-response CDM, there are two parts that need to be specified: (a) the type of parametric family that models the response distribution, i.e., the forms of \mathcal{P} and $g(\cdot; \boldsymbol{\eta})$; and (b) the type of interactions between the latent attributes and the observed responses, i.e., in what way $\boldsymbol{\eta}_{j,\alpha}$ are subject to the \mathbf{Q} -matrix constraints, or equivalently, whether the parameters for Y_j are impacted by the main effects, or interaction effects, or both, of the required attributes. We have already shown that the ExpCDM framework provides a flexible way of specifying part (a). As for part (b), all the existing binary-response CDMs including the conjunctive DINA model, disjunctive DINO model, additive models (ACDM, reduced RUM), and more flexible all-effect models such as GDINA and GDM can be easily extended and incorporated into our ExpCDM framework.

We next show how all the above different types of attribute-item interactions have their ExpCDM counterparts for modeling rich types of data. In the remainder of this section, we describe the exponential family-based DINA (ExpDINA) model and the exponential family-based additive CDM (ExpACDM). We define these models by further specifying how the natural parameters $\boldsymbol{\eta}_{j,\alpha}$ satisfy the \mathbf{Q} -matrix constraints in (7). Specifically, for each model, we present the lognormal distribution and the Poisson distribution as example exponential families to illustrate our general framework. These two distributions are suitable for modeling positive continuous data and nonnegative count data, respectively.

2.2.1. ExpDINA for General Responses The DINA model was proposed by Junker and Sijtsma (2001) for modeling binary responses using a conjunctive assumption. Define the *ideal response* to item j given attribute pattern $\mathbf{A} = \boldsymbol{\alpha}$ to be

$$\Gamma_{j,\alpha} = \prod_{k=1}^K \alpha_k^{q_{j,k}} = \prod_{k \in \text{pa}(j)} \alpha_k. \quad (9)$$

In other words, $\Gamma_{j,\alpha}$ is the binary indicator of whether skill pattern $\boldsymbol{\alpha}$ masters all of the required skills of item j and hence is “capable” of item j .

Under the binary-response DINA model, the positive response probability $\theta_{j,\alpha}$ can be written as

$$\theta_{j,\alpha} = \begin{cases} 1 - s_j, & \text{if } \Gamma_{j,\alpha} = 1, \\ g_j, & \text{if } \Gamma_{j,\alpha} = 0. \end{cases} \quad (10)$$

Following the conventional notation in the literature, s_j is the slipping parameter that gives the probability that a capable subject of item j provides an incorrect answer, and g_j is the guessing parameter that gives the probability that an incapable subject provides a correct answer. Continuing from (10), we can also write the Bernoulli natural parameters in terms of the item parameters s_j and g_j by using the following transformation:

$$\eta_{j,(h)} := \begin{cases} \log\left(\frac{1 - s_j}{s_j}\right), & \text{if } h = 1, \\ \log\left(\frac{g_j}{1 - g_j}\right), & \text{if } h = 0. \end{cases}$$

Here, h equals the value of the DINA model ideal response $\Gamma_{j,\alpha}$, and $\eta_{j,(h)}$ denotes the value of the natural parameter $\eta_{j,\alpha}$ given $\Gamma_{j,\alpha} = h$. We are now ready to define the general-response ExpDINA model.

Definition 4. (ExpDINA) The ExpDINA model is a submodel of the ExpCDM for which the natural parameters can be written as $\eta_{j,\alpha} = \eta_{j,(\Gamma_{j,\alpha})}$ for $j \in [J]$ and $\alpha \in \{0, 1\}^K$. The parameters in an ExpDINA include the item parameters $\{\eta_{j,(0)}, \eta_{j,(1)} : j \in [J]\}$ and proportion parameters \mathbf{p} .

Remark 3. We remark that for certain distributions there may exist a more conventional parametrization than directly parametrizing the natural parameters to depend on the latent attributes. For instance, in the ExpDINA with a normal distribution for $Y_j | \mathbf{A} = \alpha$ (see Example 1), it is more intuitive to parameterize using the mean μ and the variance σ^2 as depending on α instead of assuming so for the natural parameters $\eta = (\mu/\sigma^2, -1/2\sigma^2)^\top$. Such equivalent re-parametrizations still fit well into our definition of ExpCDMs.

Under an ExpDINA model, each natural parameter for each item j takes exactly two possible values, for capable ($\Gamma_{j,\alpha} = 1$) and incapable ($\Gamma_{j,\alpha} = 0$) subjects, respectively. Generally, ExpDINA still adopts the conjunctive assumption of required attributes and can be used to model rich types of response data. Under Definition 4, the conditional distribution $Y_j | \mathbf{A} = \alpha$ in (6) is:

$$Y_j | \mathbf{A} = \alpha \sim \begin{cases} g(Y_j; \eta_{j,(1)}), & \text{if } \Gamma_{j,\alpha} = 1; \\ g(Y_j; \eta_{j,(0)}), & \text{if } \Gamma_{j,\alpha} = 0. \end{cases} \quad (11)$$

Therefore, ExpDINA essentially models each response Y_j as a local mixture of two distributions, similar in spirit to the binary-response DINA model. Indeed, ExpDINA covers the binary-response DINA and polytomous-response DINA as special cases. From the definition of $\Gamma_{j,\alpha}$, it is clear that the \mathbf{Q} -matrix constraints (7) are satisfied.

Example 1. (Lognormal-DINA, Logistic-Normal-DINA, and transformed-Normal-DINA) As an illustrative example, consider using the lognormal distribution to model positive continuous response. The lognormal is very commonly used to model response times (van der Linden,

2006, 2007; Minchen et al., 2017). The lognormal density with mean parameter μ and variance parameter σ^2 takes the form of

$$g^{\text{lognormal}}(y; \mu, \sigma^2) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log y - \mu)^2}{2\sigma^2}\right], \quad y \in (0, \infty).$$

As mentioned in Remark 3, we directly parameterize the mean $\mu_{j,h}$ and variance $\sigma_{j,h}^2$ for the capable or incapable students based on the DINA model ideal response values $h = 1$ or 0 , respectively, instead of using the less interpretable natural parameters $\eta = (\mu/\sigma^2, -1/2\sigma^2)^\top$. Then, our lognormal-DINA model becomes the so-called C-DINA model proposed by Minchen et al. (2017):

$$Y_j \mid \mathbf{A} = \boldsymbol{\alpha} \sim \begin{cases} \text{lognormal}(\mu_{j,1}, \sigma_{j,1}^2), & \text{if } \Gamma_{j,\boldsymbol{\alpha}} = 1, \\ \text{lognormal}(\mu_{j,0}, \sigma_{j,0}^2), & \text{if } \Gamma_{j,\boldsymbol{\alpha}} = 0. \end{cases} \quad (12)$$

Note that under the same parameter values (μ, σ^2) , the lognormal distribution is just an exponential transformation of the Normal distribution. Hence, the lognormal-DINA and normal-DINA models are equivalent up to an invertible transform of Y . Importantly, ExpCDMs based on *any monotone-increasing transformation* of the Normal are equivalent to the *Normal-CDMs*, that is, when $Y_j \mid \mathbf{A} = \boldsymbol{\alpha}$ follows a Normal distribution. This observation can be used to build ExpCDMs for continuous responses that take values in a restricted range. For example, in addition to lognormal-CDM, we can define logistic-Normal-CDM by applying a logit transform of a Normal random variable to model bounded responses ranging in the interval $(0, 1)$. Here, the density function of the logistic-Normal(μ, σ^2) is

$$g^{\text{logistic-Normal}}(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}y(1-y)} \exp\left[-\frac{(\log(y/(1-y)) - \mu)^2}{2\sigma^2}\right], \quad y \in (0, 1).$$

The *logistic-Normal-DINA* model can be defined as:

$$Y_j \mid \mathbf{A} = \boldsymbol{\alpha} \sim \text{logistic-Normal}(\mu_{j,h}, \sigma_{j,h}^2), \quad h = \Gamma_{j,\boldsymbol{\alpha}}.$$

This newly proposed logistic-Normal-DINA model may be useful for analyzing continuous bounded responses in psychological or educational assessments.

Example 2. (Poisson-DINA) As another example, we consider using the Poisson distribution to model count responses. The Poisson distribution is a canonical choice for modeling count responses, and was recently used in Liu et al. (2022) to model the number of correct answers in repetitive tasks under a diagnostic model. Poisson distribution with mean parameter λ has the probability mass function

$$g(y; \lambda) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y \in \mathbb{N},$$

where \mathbb{N} denotes the collection of all nonnegative integers. By defining the mean for $h = 0, 1$ as $\lambda_{j,h}$ (or equivalently setting the natural parameters as $\eta_{j,(h)} = \log \lambda_{j,h}$), the distribution $Y_j | \mathbf{A}$ can be written as

$$Y_j | \mathbf{A} = \alpha \sim \begin{cases} \text{Poisson}(\lambda_{j,1}), & \text{if } \Gamma_{j,\alpha} = 1, \\ \text{Poisson}(\lambda_{j,0}), & \text{if } \Gamma_{j,\alpha} = 0. \end{cases} \quad (13)$$

This can be viewed as a reparameterization of the PDCM-DINA model in Liu et al. (2022).

2.2.2. ExpACDM for General Responses We next define the *exponential family-based Additive CDMs* (ExpACDM) for general responses. The word *additive* refers to the assumption that the required skills of item j enter the conditional distribution of Y_j through an additive linear combination of the latent attributes A_1, \dots, A_K . Variants of such additive diagnostic models are very popular in the binary-response CDM literature. Examples include the reduced reparameterized unified model (R-RUM; DiBello et al., 1995), the additive cognitive diagnostic model (ACDM; de la Torre, 2011), and the logistic linear model (LLM; Maris, 1999).

We continue to work in the general framework of general-response CDMs in (6). Under the additive assumption, we write the parameters $\eta_{j,\alpha}$ as

$$\eta_{j,\alpha} = \mathbf{h} \left(\beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k, \gamma_j \right) = \mathbf{h} \left(\beta_{j,0} + \sum_{k \in \text{pa}(j)} \beta_{j,k} \alpha_k, \gamma_j \right). \quad (14)$$

We explain the notations in the above expression in turn. First, similar to binary-response additive CDMs, the corresponding $\beta_{j,k}$ are not needed to define the model when $q_{j,k} = 0$. So without the loss of generality, we assume $\beta_{j,k} = 0$ if $q_{j,k} = 0$. If $q_{j,k} = 1$, then $\beta_{j,k}$ is the main-effect coefficient for the k th latent attribute. One can see that the linear combination $\beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k$ depends only on those required attributes of item j , so $\eta_{j,\alpha}$ defined by (14) satisfies the \mathbf{Q} -matrix constraints (7).

The γ_j in (14) denotes potential additional model parameters that do not depend on the latent attributes (e.g., γ_j may be the dispersion parameter if an exponential family distribution is used), and \mathbf{h} in (14) is a link function that plays a very similar role as the link function in the generalized linear models (e.g., Nelder and Wedderburn, 1972). Here, \mathbf{h} is introduced to map the linear combination of the required attributes to the natural parameters η , as the natural parameter space may be different from the space of the linear combinations. Note that η may be multidimensional, and hence \mathbf{h} can be a mapping between multidimensional spaces.

With these notations at hand, we next formally define ExpACDMs.

Definition 5. (ExpACDM) An ExpACDM is a submodel of the ExpCDM for which the natural parameters $\eta_{j,\alpha}$ satisfy (7).

The parameters in an ExpACDM with a link function \mathbf{h} include $\{\beta_{j,0} : j \in [J]\}$, $\{\beta_{j,k} : q_{j,k} = 1\}$, $\{\gamma_j : j \in [J]\}$ and proportion parameters \mathbf{p} .

Note that for the ExpDINA model defined earlier in Sect. 2.2.1, we have parameterized the item parameters as $(\eta_{j,(0)}, \eta_{j,(1)})$, because there are only two possible values of $\{\eta_{j,\alpha} : \alpha \in \{0, 1\}^K\}$ due to the conjunctive modeling assumption. In contrast, the ExpACDM has more possible values of $\{\eta_{j,\alpha}\}$ when the attribute profile α ranges in $\{0, 1\}^K$. So in Definition 5, we have parameterized ExpACDMs using the β - and γ -coefficients.

Example 3. (lognormal-ACDM) Continuing from Example 1, we consider the lognormal distribution for modeling the responses but now with an additive structure of the attributes.

For the lognormal parameters $(\mu_{j,\alpha}, \sigma_{j,\alpha}^2)$, one can choose to model $\mu_{j,\alpha}$ as additive in α_k 's and $\sigma_{j,\alpha}^2$ to not depend on α :

$$\mu_{j,\alpha} = \beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k, \quad \sigma_{j,\alpha}^2 = \gamma_j.$$

Here, we can view the link function \mathbf{h} as $\mathbf{h}(\mu, \sigma^2) = (\mu/\sigma^2, -1/2\sigma^2)$ that maps the mean and variance to the natural parameters under the lognormal distribution.

By plugging this parametrization in (6), $Y_j | \mathbf{A}$ can be written as

$$Y_j | \mathbf{A} = \alpha \sim \text{lognormal} \left(\beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k, \gamma_j \right). \quad (15)$$

Note that there are multiple modeling choices regarding how to specify the dependence of $\mu_{j,\alpha}$ and $\sigma_{j,\alpha}^2$ on the linear (i.e., additive) combinations of α_k 's.

One could model an additive mean and a constant variance after the log transformation as done in the above (15). Alternatively, one could even model the variance also as additive in α_k 's.

Example 4. (Poisson-ACDM) As an counterpart of the Poisson-DINA model in Example 2, we define an ExpACDM that uses the Poisson distribution to model count data.

Since Poisson only has one rate parameter, there is no need to introduce γ_j and we only need to express the rate parameter as the linear combination of those required attributes:

$$\lambda_{j,\alpha} = \beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k.$$

Recall that the natural parameter under Poisson is $\eta = \log \lambda$, so we define the link function \mathbf{h} as $\mathbf{h}(\lambda) = \log \lambda$. We additionally assume $\beta_{j,k} \geq 0$ to ensure that the rate parameters $\lambda_{j,\alpha}$ are nonnegative. By plugging the above expression in (6), $Y_j | \mathbf{A}$ can be written as

$$Y_j | \mathbf{A} = \alpha \sim \text{Poisson} \left(\beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k \right).$$

Alternatively, one could let \mathbf{h} to be the identity link $\mathbf{h}(\lambda) = \lambda$ without assuming $\beta_{j,k} \geq 0$.

In Supplementary Material S.1, we define the exponential family-based general diagnostic model (ExpGDM) for general responses, and also discuss how to define the general-response DINA and ACDM for distributions outside of the exponential family, such as the negative binomial distribution. In this section, we have focused on cases where the parametric families \mathcal{P} do not change across different items $j \in [J]$. But it is actually straightforward to use our framework to model different distribution families for different items.

3. Identifiability of General-response CDMs

3.1. Strict Identifiability of General-Response CDMs

Model identifiability is a fundamental prerequisite for valid statistical estimation and inference. We will prove that our general-response CDMs are identifiable under transparent conditions on the \mathbf{Q} -matrix. In this subsection, we first define strict identifiability for general-response CDMs and then propose conditions for strict identifiability.

Recall that $(\mathbf{p}, \{\mathbb{P}_{j,\alpha} : j \in [J], \alpha \in \{0, 1\}^K\})$ can be viewed as the model parameters of a general-response CDM. We will say that $(\mathbf{p}, \{\mathbb{P}_{j,\alpha}\})$ is equal to $(\bar{\mathbf{p}}, \{\bar{\mathbb{P}}_{j,\alpha}\})$ up to a sign flip for each coordinate of α , if there exist permutation maps $\tau_k : \{0, 1\} \rightarrow \{0, 1\}$ such that $p_\alpha = \bar{p}_{(\tau_1(\alpha_1), \dots, \tau_K(\alpha_K))}$ and $\mathbb{P}_{j,\alpha} = \bar{\mathbb{P}}_{j,(\tau_1(\alpha_1), \dots, \tau_K(\alpha_K))}$ for any $j \in [p]$ and $\alpha \in \{0, 1\}^K$. The sign flipping issue of the binary latent variables in Definition 6 is inevitable for general-response CDMs under the minimal assumptions that we make on the response types. Nevertheless, sign flipping is a trivial ambiguity and can be easily fixed in parametric submodels including most ExpCDMs. For example, in traditional CDMs for binary responses, the sign of latent variables is fixed by making the monotonicity assumption that students who possess all required skills of an item are more likely to correctly answer it compared to those who lack some required skills. Similarly, to resolve the sign flipping issue, in the lognormal-DINA model for continuous responses, one can assume that the students who possess all skills required by an item have a larger mean parameter in the lognormal distribution than those who do not: $\mu_{j,1} > \mu_{j,0}$. More generally, for ExpACDMs, assuming $\beta_{j,k} > 0$ when $q_{j,k} = 1$ resolves the sign flipping issue. This claim is formally presented in the later Proposition 1. However, it is less convincing to assume monotonicity for every response type, as not all responses increase with more skill mastery, and moreover, an ordering may not be defined on the sample space of the response \mathcal{Y}_j if it is not a subset of the real line \mathbb{R} . We next formally define strict identifiability up to the sign flipping.

Definition 6. (Strict identifiability) Consider a general-response CDM with a known \mathbf{Q} -matrix. The model is strictly identifiable if for any parameters $(\mathbf{p}, \{\mathbb{P}_{j,\alpha}\})$ and $(\bar{\mathbf{p}}, \{\bar{\mathbb{P}}_{j,\alpha}\})$ satisfying (3), the following equations hold if and only if $(\mathbf{p}, \{\mathbb{P}_{j,\alpha}\})$ is equal to $(\bar{\mathbf{p}}, \{\bar{\mathbb{P}}_{j,\alpha}\})$ up to a sign flip for each coordinate of α :

$$\mathbb{P}(\mathbf{Y} \in \times_{j=1}^J S_j \mid \bar{\mathbf{p}}, \bar{\mathbb{P}}, \mathbf{Q}) = \mathbb{P}(\mathbf{Y} \in \times_{j=1}^J S_j \mid \mathbf{p}, \mathbb{P}, \mathbf{Q}), \quad \forall S_j \in \mathcal{F}_j. \quad (16)$$

Now we are ready to state the main theorem on strict identifiability. We emphasize that no further parametric assumptions are imposed beyond the three assumptions for the general-response CDM, given in Definition 1. Hence, the result applies for general-response CDMs with an arbitrary response type \mathcal{Y}_j .¹

Theorem 1. *Under the general-response CDM, the model components $(\mathbf{p}, \{\mathbb{P}_{j,\alpha}\})$ are strictly identifiable if the following conditions hold.*

A. *The true \mathbf{Q} -matrix contains two identity submatrices \mathbf{I}_K after row swapping, i.e., \mathbf{Q} can be written as*

$$\mathbf{Q} = [\mathbf{I}_K, \mathbf{I}_K, \mathbf{Q}^{*\top}]^\top.$$

B. *Suppose that the \mathbf{Q} -matrix is written as in A. For any $\alpha \neq \alpha'$, there exists $j > 2K$ such that $\mathbb{P}_{j,\alpha} \neq \mathbb{P}_{j,\alpha'}$.*

¹We prove Theorem 1 under the minimal assumption (in Supplementary Material S.1) that \mathcal{Y}_j is a separable metric space.

The conditions in Theorem 1 are similar in spirit to existing identifiability conditions for CDMs with binary or polytomous responses, such as Xu (2017) and Culpepper (2019). When considering unidimensional responses Y_j 's for each item, we can additionally make a monotonicity assumption to avoid the sign flipping issue. The following proposition establishes the strongest possible notion of identifiability under a monotonicity assumption (17). Here, $\mathbf{0}_K$ and $\mathbf{1}_K$ denote the all-zero and all-one vectors of length K , respectively.

Proposition 1. *Suppose that a general-response CDM satisfies conditions A and B. Without loss of generality, suppose the first K rows in \mathbf{Q} form an identity matrix \mathbf{I}_K .*

Additionally, suppose that $\mathcal{Y}_j \subseteq \mathbb{R}$, and the conditional distributions $\{\mathbb{P}_{j,\alpha}\}$ satisfy the following monotonicity assumption:

$$\mathbb{E}(Y_j \mid \mathbf{A} = \mathbf{0}_K) < \mathbb{E}(Y_j \mid \mathbf{A} = \mathbf{1}_K), \quad (17)$$

for $1 \leq j \leq K$.

Then, the model components $(\mathbf{p}, \{\mathbb{P}_{j,\alpha}\})$ are strictly identifiable with no sign flipping issues. In other words, (16) implies $(\mathbf{p}, \{\mathbb{P}_{j,\alpha}\}) = (\bar{\mathbf{p}}, \{\bar{\mathbb{P}}_{j,\alpha}\})$.

The proofs of all theoretical results are deferred to the Supplementary Material. We next provide the high-level intuition of our proof argument. The main idea is to discretize all responses by first constructing a partition of each sample space \mathcal{Y}_j into a finite number of categories, and then defining surrogate categorical variables based on the partitions. In this case, our model in (4) implies a tensor decomposition for the probability mass function of the surrogate categorical variables. We then leverage Kruskal's theorem (Kruskal, 1977) to establish uniqueness of this tensor decomposition under the \mathbf{Q} -matrix conditions. Finally, we link such uniqueness results back to identifiability of parameters based on the original distribution of the general responses $Y_j \in \mathcal{Y}_j$. Our proof strategy is motivated by Theorem 8 in Allman et al. (2009) which established identifiability of mixtures of products of nonparametric probability densities. However, Allman et al. (2009)'s result cannot be directly applied to our setup as we consider more general response types and our conditional distributions are subject to \mathbf{Q} -matrix constraints.

As pointed out by a reviewer, our results extend identifiability conditions for finite mixtures of product distributions for continuous/count responses to the more intricate setting of general-response CDMs with the \mathbf{Q} -matrix constraints. While those traditional mixture models also satisfy our assumptions (1) and (2), there are no \mathbf{Q} -matrix constraints (3) that restrict the space of conditional distributions $\mathbb{P}_{j,\alpha}$'s. Hence, existing identifiability conditions for the general-response finite mixture of products (Teicher, 1967; Yakowitz & Spragins, 1968; Allman et al., 2009) are not directly applicable to general-response CDMs. In particular, these results commonly assume linear independence of the conditional distributions $\{\mathbb{P}_{j,\alpha} : \alpha \in \{0, 1\}^K\}$, for each $j \in [J]$. However, for CDMs, $\{\mathbb{P}_{j,\alpha}\}_{\alpha \in \{0, 1\}^K}$'s are linearly dependent unless we consider the special case of the GDINA model with $\mathbf{q}_j = \mathbf{1}_K$ for all j , corresponding to the case where the \mathbf{Q} -matrix imposes no restrictions at all. We remark that for certain parametric distributions, such as the exponential distribution and Normal distribution, it is known that mixture models with independent conditional distributions are identifiable (see Proposition 1 and 2 in Yakowitz and Spragins, 1968). However, such results heavily depend on the specific parametric family under consideration whereas our results generally apply under the minimal nonparametric assumptions.

Theorem 1 and Proposition 1 hold true without assuming a specific measurement model such as DINA or ACDM. For binary responses, there is an extensive literature on model identifiability under various assumptions (e.g., Chen et al., 2015; Xu, 2017; Xu and Shang, 2018; Chen et al., 2020; Gu and Xu, 2021). It is worth noting that when not assuming a specific measurement model, our results in Theorem 1 and Proposition 1 nearly match existing weakest identifiability

conditions for categorical-response CDMs. For example, Theorem 1 in Xu (2017) stated that the parameters (\mathbf{p}, Θ) in a restricted latent class model are strictly identifiable under condition A and the following condition B1:

B1. For any $k \in [K]$, there exists one $j > 2K$ such that $\theta_{j,\mathbf{e}_k} \neq \theta_{j,\mathbf{0}_K}$.

Here, $\Theta = (\theta_{j,\alpha})_{j \in [J], \alpha \in \{0,1\}^K}$ is the positive response probability matrix with $\theta_{j,\alpha} = \mathbb{P}_{j,\alpha}(Y_j = 1 \mid \mathbf{A} = \alpha)$, and this determines the $\mathbb{P}_{j,\alpha}$'s, as mentioned in Sect. 2. Also, \mathbf{e}_k is the standard basis vector whose k th entry is 1 and other entries are zeros. For exploratory CDMs with polytomous responses, Theorem 2 in Culpepper (2019) proved that the model parameters are identifiable under conditions A, B, and an additional condition C:

C. Each item $j \in [J]$ has distinct item response functions for at least two latent classes.

For general-response CDMs, the above additional assumption is equivalent to “*for all $j \in [J]$, there exist skill patterns $\alpha \neq \alpha'$ such that $\mathbb{P}_{j,\alpha} \neq \mathbb{P}_{j,\alpha'}$* ”. Our Theorem 1 does not impose such an assumption for all items, but only implicitly imposes it for $j = 1, \dots, 2K$ (because the first $2K$ rows in \mathbf{Q} form two copies of the identity matrix \mathbf{I}_K). As for $j > 2K$, we do not require there exist $\alpha \neq \alpha'$ such that $\mathbb{P}_{j,\alpha} \neq \mathbb{P}_{j,\alpha'}$ for each j . In fact, Theorem 1 allows some Y_j for $j > 2K$ to not depend on any attributes; or equivalently, the corresponding row vector \mathbf{q}_j in the \mathbf{Q} -matrix is an all-zero row vector. Such all-zero row vectors can be absorbed into the submatrix \mathbf{Q}^* in condition A of Theorem 1.

Finally, in the following propositions, we propose variations of Theorem 1 by modifying condition B to more easily checkable ones. In Proposition 2, we replace condition B in Theorem 1 by a stronger, but more intuitive condition. This condition has been previously proposed to establish identifiability of exploratory diagnostic models for categorical responses (Fang et al., 2019; Gu & Dunson, 2023).

Proposition 2. *Condition B in Theorem 1 holds when \mathbf{Q}^* contains an identity submatrix \mathbf{I}_K . Hence, the general-response CDM is strictly identifiable when it vertically stacks three identity submatrices \mathbf{I}_K after some row swapping.*

In the next proposition, we consider the ExpACDM with parameters $(\mathbf{p}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{Q})$, instead of the general-response CDM, and show how condition B can be simplified.

Proposition 3. *For the ExpACDM, condition B in Theorem 1 reduces to:*

B2. *For any $\alpha \neq \alpha' \in \{0, 1\}^K$, there exists $j > 2K$ such that $\sum_{k=1}^K q_{j,k} \beta_{j,k} (\alpha_k - \alpha'_k) \neq 0$.*

Hence, the ExpACDM is strictly identifiable when the \mathbf{Q} -matrix satisfies condition A, and the main-effect coefficients $\boldsymbol{\beta}$ satisfies condition B2.

3.2. Generic Identifiability of ExpACDMs

Although strict identifiability is the strongest possible identifiability notion, in practice it may impose too stringent conditions on the \mathbf{Q} -matrix. For instance, the \mathbf{Q} -matrix for the TIMSS 2019 response time dataset analyzed in Sect. 6 requires one content skill and one cognitive skill for each item (see Table 2 and related discussion there). This \mathbf{Q} -matrix does not contain any K -dimensional standard basis vectors as row vectors, and hence does not contain an identity submatrix \mathbf{I}_K . In this subsection, we study generic identifiability, which is a slightly weaker notion than strict identifiability proposed by Allman et al. (2009). Generic identifiability requires that the model parameters are identifiable almost everywhere in the parameter space, in the sense that they are identifiable except in a Lebesgue measure-zero subset of the parameter space. Existing studies such as Gu and Xu (2020) and Chen et al. (2020) proposed practical generic identifiability conditions

for binary-response CDMs that contain main effects of latent attributes. Those conditions are much weaker than the strict identifiability conditions.

To study generic identifiability, we next focus on parametric forms of $\mathbb{P}_{j,\alpha}$ and especially consider ExpACDMs that model the main effects of latent attributes. For notational simplicity, we let $\beta_{\text{main}} = \{\beta_{j,k}\}_{k \in [J], k \in [K]}$ denote all main-effect coefficients, and $\beta = \beta_{\text{main}} \cup \{\beta_{j,0}\}_{j \in [J]}$ denote the collection of all intercepts and main-effect coefficients. Before defining the concept of generic identifiability, we define the true parameter space $\Omega(\beta_{\text{main}}; \mathbf{Q})$ for β_{main} as follows:

$$\Omega(\beta_{\text{main}}; \mathbf{Q}) = \{\beta_{\text{main}} : \beta_{j,k} = 0 \text{ if } q_{j,k} = 0\}. \quad (18)$$

Here, we are using the previous assumption that $\beta_{j,k} = 0$ if $q_{j,k} = 0$ (see the text following (14)). We also define the parameter space of $\gamma = (\gamma_1, \dots, \gamma_J)^\top$ as the set $\Omega(\gamma)$. Recall that in an ExpACDM, γ commonly denotes the dispersion parameters of the exponential family, so we define $\Omega(\gamma)$ as $\Omega(\gamma) = \{\gamma : \gamma_j > 0 \text{ for all } j \in [J]\}$. Also, recall that \mathbf{p} denotes the proportion parameters of the latent attribute patterns. We define the joint parameter space of all parameters \mathbf{p} , β , and γ by

$$\Omega(\mathbf{p}, \beta, \gamma; \mathbf{Q}) = \{(\mathbf{p}, \beta, \gamma) : \sum_{\alpha \in \{0,1\}^K} p_\alpha = 1, p_\alpha \geq 0, \beta_{\text{main}} \in \Omega(\beta_{\text{main}}; \mathbf{Q}), \gamma \in \Omega(\gamma)\}. \quad (19)$$

In order to define generic identifiability without the sign flipping issue, we next adopt the monotonicity assumption in (17). In particular, we assume that $\beta_{j,k} > 0$ when $q_{j,k} = 1$, which means that the main effects of the required attributes are all positive. Here, the actual sign of $\beta_{j,k}$ is not important and one may assume $\beta_{j,k} < 0$ for all j and k instead to achieve identifiability. Now we are ready to define generic identifiability for ExpACDMs.

Definition 7. An ExpACDM with parameters $(\mathbf{p}, \beta, \gamma)$ is generically identifiable if the following set has measure zero with respect to the Lebesgue measure on $\Omega(\mathbf{p}, \beta, \gamma; \mathbf{Q})$:

$$\{(\mathbf{p}, \beta, \gamma) \in \Omega(\mathbf{p}, \beta, \gamma; \mathbf{Q}) : \text{there exist alternative parameters } (\bar{\mathbf{p}}, \bar{\beta}, \bar{\gamma}) \neq (\mathbf{p}, \beta, \gamma) \text{ in } \Omega(\mathbf{p}, \beta, \gamma; \mathbf{Q}) \text{ such that } \mathbb{P}(\mathbf{Y} | \mathbf{p}, \beta, \gamma) = \mathbb{P}(\mathbf{Y} | \bar{\mathbf{p}}, \bar{\beta}, \bar{\gamma})\}.$$

We next propose generic identifiability conditions for ExpACDMs that are substantially weaker than the strict identifiability conditions in Sect. 3.1.

Theorem 2. Consider an ExpACDM. Assume that \mathbf{h} is an analytic function, and that the true parameters $(\mathbf{p}, \beta, \gamma)$ lie in $\Omega(\mathbf{p}, \beta, \gamma; \mathbf{Q})$. Then, the model parameters $(\mathbf{p}, \beta, \gamma)$ are generically identifiable if the following conditions on the \mathbf{Q} -matrix hold.

A*. The true \mathbf{Q} -matrix can be written as $\mathbf{Q} = [\mathbf{Q}_1^\top, \mathbf{Q}_2^\top, \mathbf{Q}^{*\top}]^\top$ after some row permutation, where $\mathbf{Q}_1, \mathbf{Q}_2$'s are $K \times K$ matrices such that all diagonal entries are equal to one. In other words, we can write

$$\mathbf{Q}_i = \begin{pmatrix} 1 & * & \dots & * \\ * & 1 & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & 1 \end{pmatrix}$$

for $i = 1, 2$. Here, $*$ indicates an arbitrary value in $\{0, 1\}$.

B^* . Suppose that the \mathbf{Q} -matrix is written as in A^* , and \mathbf{Q}^* does not have any all-zero columns. In other words, for any k , there exists $j > 2K$ such that $q_{j,k} = 1$.

Conditions in Theorem 2 only depend on the \mathbf{Q} -matrix but not on other model components. Theorem 2 is motivated by Gu and Xu (2020) and Chen et al. (2020), which are the first studies that considered generic identifiability for CDMs. Theorem 4.3 in Gu and Xu (2020) and Theorem 1 in Chen et al. (2020) showed that binary-response CDMs are generically identifiable under conditions A^* and B^* .

Note that Theorem 2 holds without assuming (1) that $p_\alpha > 0$ for all $\alpha \in \{0, 1\}^K$. This is because $\{\mathbf{p} : \exists \alpha \text{ such that } p_\alpha = 0\}$ is a measure-zero subset of $\{\mathbf{p} : \sum_\alpha p_\alpha = 1, p_\alpha \geq 0\}$, and hence relaxing the condition from $p_\alpha > 0$ to $p_\alpha \geq 0$ does not violate generic identifiability. In addition, we point out that the exponential family assumption for \mathcal{P} is not crucial for generic identifiability to hold. In fact, \mathcal{P} can be any parametric family as long as the model parameters $\eta_{j,\alpha}$ satisfy the additive assumption in (14). Hence, the conditions in Theorem 2 also guarantee the generic identifiability of ACDMs with non-exponential family distributions; see more discussions in Remark S.1 in the Supplementary Material S.2.

Theorem 2 may be extended to the setting of ExpGDM and other all-effect CDMs for general responses. This is because these models include the main effects of the latent attributes, which are the key components underlying generic identifiability. We leave the rigorous statement of generic identifiability of ExpGDMs as future work. On the other hand, Theorem 2 does not apply to the DINA model, because the DINA model is conjunctive and does not include the main effects of the latent attributes. Under the DINA model for general responses, we conjecture that another identifiability notion similar to the \mathbf{p} -partial identifiability in Gu and Xu (2020) for conventional categorical-response CDMs can be studied.

4. Universal EM Algorithms for Estimating ExpCDMs

Our identifiability result in Theorem 1 has a nice consequence of statistical consistency of the maximum likelihood estimator (MLE). We next formally establish this result. Consider a sample with N independent and identically distributed response vectors $\mathbf{Y}_{1:N} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_N\}$ from an ExpCDM in (6) with true parameters (η_0, \mathbf{p}_0) . Given the \mathbf{Q} -matrix, the marginal log-likelihood can be written as

$$\ell(\eta, \mathbf{p} \mid \mathbf{Y}_{1:N}, \mathbf{Q}) = \sum_{i=1}^N \log \mathbb{P}(\mathbf{Y} = \mathbf{Y}_i \mid \eta, \mathbf{p}) = \sum_{i=1}^N \log \left(\sum_{\alpha \in \{0, 1\}^K} p_\alpha \prod_{j=1}^J g(Y_{i,j}; \eta_{j,\alpha}) \right) \quad (20)$$

where η is the collection of all $\eta_{j,\alpha}$'s that are subject to the \mathbf{Q} -matrix constraints under the considered ExpCDM. Define the MLE of model parameters as

$$(\hat{\eta}, \hat{\mathbf{p}}) = \underset{\eta, \mathbf{p}}{\operatorname{argmax}} \ell(\eta, \mathbf{p} \mid \mathbf{Y}_{1:N}).$$

Following from the identifiability conclusions, we have the following proposition on parameter estimation consistency.

Proposition 4. (Parameter Estimation Consistency) Consider an ExpCDM in the form (6) where the true parameters (η_0, \mathbf{p}_0) lie in the interior of the parameter space. Suppose the strict identifiability conditions A and B in Theorem 1 hold. Then $(\hat{\eta}, \hat{\mathbf{p}})$ converge to (η_0, \mathbf{p}_0) almost surely as $N \rightarrow \infty$.

Next, consider an ExpACDM with true parameters $(\mathbf{p}_0, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ where the generic identifiability conditions A^* and B^* in Theorem 2 hold. Then, the MLE $(\widehat{\mathbf{p}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}})$ converges to $(\mathbf{p}_0, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ almost surely, for $(\mathbf{p}_0, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0) \in \Omega(\mathbf{p}, \boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{Q}) \setminus \mathcal{N}$. Here, \mathcal{N} is a negligible measure-zero set in $\Omega(\mathbf{p}, \boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{Q})$.

Next we propose an expectation-maximization (EM, Dempster et al., 1977) algorithm for parameter estimation. For the purpose of developing an EM algorithm, consider the log-likelihood for the complete data $(\mathbf{Y}_{1:N}, \mathbf{A}_{1:N}) = (\mathbf{Y}_i, \mathbf{A}_i)_{i=1,\dots,N}$:

$$\begin{aligned}\ell_c(\boldsymbol{\eta}, \mathbf{p} \mid \mathbf{Y}_{1:N}, \mathbf{A}_{1:N}, \mathbf{Q}) &= \log \left[\prod_{i=1}^N \mathbb{P}(\mathbf{A}_i) \mathbb{P}(\mathbf{Y}_i \mid \mathbf{A}_i, \mathbf{Q}) \right] \\ &= \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \sum_{i=1}^N \mathbb{1}(\mathbf{A}_i = \boldsymbol{\alpha}) \left(\log p_{\boldsymbol{\alpha}} + \sum_{j=1}^J \log g(Y_{i,j}; \boldsymbol{\eta}_{j,\boldsymbol{\alpha}}) \right),\end{aligned}\quad (21)$$

where $\mathbf{A}_i \in \{0, 1\}^K$ denotes the latent attribute profile of the i th subject in the sample. In Sects. 4.1 and Sect. 4.2, we present details of the EM algorithms for the general ExpDINA and ExpACDM models.

4.1. EM Algorithm for the ExpDINA

Consider the ExpDINA model defined in Sect. 2.2.1, which is parametrized by item parameters $\{\boldsymbol{\eta}_{j,(h)}\}_{j \in [J], h=0,1}$ and \mathbf{p} . In this subsection, for notational simplicity, we write $\boldsymbol{\eta} = \{\boldsymbol{\eta}_{j,(h)}\}_{j \in [J], h=0,1}$ to denote all item parameters. For the ExpDINA model, the complete data log-likelihood (21) can be rewritten as:

$$\begin{aligned}\ell_c^{\text{ExpDINA}}(\boldsymbol{\eta}, \mathbf{p} \mid \mathbf{Y}_{1:N}, \mathbf{A}_{1:N}) &= \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \sum_{i=1}^N \mathbb{1}(\mathbf{A}_i = \boldsymbol{\alpha}) \log p_{\boldsymbol{\alpha}} \\ &\quad + \sum_{\boldsymbol{\alpha} \in \{0,1\}^K} \sum_{i=1}^N \mathbb{1}(\mathbf{A}_i = \boldsymbol{\alpha}) \sum_{j=1}^J \\ &\quad \left[\left(\boldsymbol{\eta}_{j,(0)}^\top \mathbf{T}(Y_{i,j}) - A(\boldsymbol{\eta}_{j,(0)}) + \log h(Y_{i,j}) \right) (1 - \Gamma_{j,\boldsymbol{\alpha}}) \right. \\ &\quad \left. + \left(\boldsymbol{\eta}_{j,(1)}^\top \mathbf{T}(Y_{i,j}) - A(\boldsymbol{\eta}_{j,(1)}) + \log h(Y_{i,j}) \right) \Gamma_{j,\boldsymbol{\alpha}} \right].\end{aligned}\quad (22)$$

Recall that $\Gamma_{j,\boldsymbol{\alpha}}$ is the binary ideal response, which equals one if and only if the latent skill pattern $\boldsymbol{\alpha}$ possesses all skills required by item j . We next present details of the E step and M step in the EM algorithm.

E step for ExpDINA In the E step of the t -th EM iteration, we calculate the conditional expectation of ℓ_c^{ExpDINA} in (22) given the current parameter values $\boldsymbol{\eta}^{(t)}, \mathbf{p}^{(t)}$. As \mathbf{A}_i is discrete and ranges in $\{0, 1\}^K$, it suffices to evaluate $\varphi_{i,\boldsymbol{\alpha}}^{(t+1)} := \mathbb{P}(\mathbf{A}_i = \boldsymbol{\alpha} \mid \boldsymbol{\eta}^{(t)}, \mathbf{p}^{(t)})$ for all $\boldsymbol{\alpha} \in \{0, 1\}^K$. This can be calculated by noting that

$$\varphi_{i,\boldsymbol{\alpha}}^{(t+1)} = p_{\boldsymbol{\alpha}} \prod_{j=1}^J \mathbb{P}(Y_{i,j} \mid \boldsymbol{\eta}^{(t)}, \mathbf{p}^{(t)}, \mathbf{A}_i = \boldsymbol{\alpha}) \Big/ \left(\sum_{\boldsymbol{\alpha}' \in \{0,1\}^K} p_{\boldsymbol{\alpha}'} \prod_{j=1}^J \mathbb{P}(Y_{i,j} \mid \boldsymbol{\eta}^{(t)}, \mathbf{p}^{(t)}, \mathbf{A}_i = \boldsymbol{\alpha}') \right).$$

The exact form of $\varphi_{i,\alpha}^{(t+1)}$ is given in Algorithm 1. Now using $\varphi_{i,\alpha}^{(t+1)}$, we can calculate the conditional expectation of ℓ_c^{ExpDINA} as follows and denote it as $Q(\eta, p \mid \eta^{(t)}, p^{(t)})$:

$$\begin{aligned} Q(\eta, p \mid \eta^{(t)}, p^{(t)}) &= \mathbb{E} \left[\ell_c(\eta, p \mid \mathbf{Y}, \mathbf{A}) \mid \eta^{(t)}, p^{(t)} \right] \\ &= \sum_{\alpha \in \{0, 1\}^K} \sum_{i=1}^N \varphi_{i,\alpha}^{(t+1)} \log p_{\alpha} \\ &\quad + \sum_{\alpha \in \{0, 1\}^K} \sum_{i=1}^N \varphi_{i,\alpha}^{(t+1)} \sum_{j=1}^J \left[\left(\eta_{j,(0)}^{\top} \mathbf{T}(Y_{i,j}) - A(\eta_{j,(0)}) + \log h(Y_{i,j}) \right) (1 - \Gamma_{j,\alpha}) \right. \\ &\quad \left. + \left(\eta_{j,(1)}^{\top} \mathbf{T}(Y_{i,j}) - A(\eta_{j,(1)}) + \log h(Y_{i,j}) \right) \Gamma_{j,\alpha} \right]. \end{aligned}$$

M step for ExpDINA Next, in the M step in the t th EM iteration, we maximize $Q(\eta, p \mid \eta^{(t)}, p^{(t)})$ with respect to (η, p) and update the model parameters as follows:

$$(\eta^{(t+1)}, p^{(t+1)}) = \underset{\eta, p}{\text{argmax}} Q(\eta, p \mid \eta^{(t)}, p^{(t)}).$$

Since $\eta_{j,(0)}, \eta_{j,(1)}, p$ are continuous parameters, we can set their partial gradients to zero and update parameters as the solutions to the gradient equations. Specifically, we will solve

$$\underset{p}{\text{argmax}} \sum_{\alpha \in \{0, 1\}^K} \sum_{i=1}^N \varphi_{i,\alpha}^{(t+1)} \log p_{\alpha}, \quad \text{subject to} \sum_{\alpha \in \{0, 1\}^K} p_{\alpha} = 1, \quad (23)$$

$$\underset{\eta_{j,(0)}}{\text{argmax}} \sum_{\alpha \in \{0, 1\}^K} \sum_{i=1}^N \varphi_{i,\alpha}^{(t+1)} \sum_{j=1}^J \left(\eta_{j,(0)}^{\top} \mathbf{T}(Y_{i,j}) - A(\eta_{j,(0)}) \right) (1 - \Gamma_{j,\alpha}), \quad \forall j \in [J], \quad (24)$$

$$\underset{\eta_{j,(1)}}{\text{argmax}} \sum_{\alpha \in \{0, 1\}^K} \sum_{i=1}^N \varphi_{i,\alpha}^{(t+1)} \sum_{j=1}^J \left(\eta_{j,(1)}^{\top} \mathbf{T}(Y_{i,j}) - A(\eta_{j,(1)}) \right) \Gamma_{j,\alpha}, \quad \forall j \in [J]. \quad (25)$$

The first optimization problem in (23) has a closed-form solution for p :

$$p_{\alpha} = \frac{\sum_{i=1}^N \varphi_{i,\alpha}^{(t+1)}}{\sum_{\alpha \in \{0, 1\}^K} \sum_{i=1}^N \varphi_{i,\alpha}^{(t+1)}}, \quad \forall \alpha \in \{0, 1\}^K.$$

For optimization problems in (24) and (25) for the item parameters, there also often exist closed-form updates for many distributions. Consider $\eta_{j,(1)}$ in (25) for some item $j \in [J]$.

Setting the partial gradient of (25) with respect to $\eta_{j,(1)}$ to zero, we obtain:

$$\sum_{\alpha} \sum_i \varphi_{i,\alpha}^{(t+1)} \Gamma_{j,\alpha} (\mathbf{T}(Y_{i,j}) - \nabla_{\eta} A(\eta)) = 0, \implies \nabla_{\eta_{j,(1)}} A(\eta) = \frac{\sum_{i,\alpha} \mathbf{T}(Y_{i,j}) \varphi_{i,\alpha}^{(t+1)} \Gamma_{j,\alpha}}{\sum_{i,\alpha} \varphi_{i,\alpha}^{(t+1)} \Gamma_{j,\alpha}}. \quad (26)$$

Now note that an exponential family distribution has a nice property that $A(\eta)$ is always a convex function of η (Casella & Berger, 2021). This is because the Hessian matrix of $A(\eta)$ equals the covariance matrix of the sufficient statistics \mathbf{T} and hence is always positive definite. This property implies that $\eta \rightarrow \nabla_\eta A(\eta)$ is always an invertible map of η and there exists a unique solution to the gradient equation in (26). In fact, for many widely used exponential family distributions, (26) has an explicit solution and hence the corresponding ExpDINA has explicit M step updates.

For example, under the *Lognormal-DINA* model, maximizing (24)–(25) gives the following updates of the mean and variance parameters:

$$\begin{aligned}\mu_{j,1}^{(t+1)} &= \frac{\sum_{i,\alpha} \varphi_{i,\alpha} \Gamma_{j,\alpha} \log Y_{i,j}}{\sum_{i,\alpha} \varphi_{i,\alpha} \Gamma_{j,\alpha}}, & (\sigma_{j,1}^2)^{(t+1)} &= \frac{\sum_{i,\alpha} \varphi_{i,\alpha} \Gamma_{j,\alpha} (\log Y_{i,j} - \mu_{j,1}^{(t+1)})^2}{\sum_{i,\alpha} \varphi_{i,\alpha} \Gamma_{j,\alpha}}; \\ \mu_{j,0}^{(t+1)} &= \frac{\sum_{i,\alpha} \varphi_{i,\alpha} (1 - \Gamma_{j,\alpha}) \log Y_{i,j}}{\sum_{i,\alpha} \varphi_{i,\alpha} (1 - \Gamma_{j,\alpha})}, & (\sigma_{j,0}^2)^{(t+1)} &= \frac{\sum_{i,\alpha} \varphi_{i,\alpha} (1 - \Gamma_{j,\alpha}) (\log Y_{i,j} - \mu_{j,1}^{(t+1)})^2}{\sum_{i,\alpha} \varphi_{i,\alpha} (1 - \Gamma_{j,\alpha})}.\end{aligned}$$

Updates for ExpDINA for other transformed-Normal distributions (e.g., logistic-Normal) can be obtained by simply replacing the $\log Y_{i,j}$ term in the above display with the corresponding transform. As another example, under the *Poisson-DINA* model, maximizing (24)–(25) gives the following updates for the Poisson rate parameters:

$$\lambda_{j,1}^{(t+1)} = \frac{\sum_{i,\alpha} \varphi_{i,\alpha} \Gamma_{j,\alpha} Y_{i,j}}{\sum_{i,\alpha} \varphi_{i,\alpha} \Gamma_{j,\alpha}}, \quad \lambda_{j,0}^{(t+1)} = \frac{\sum_{i,\alpha} \varphi_{i,\alpha} (1 - \Gamma_{j,\alpha}) Y_{i,j}}{\sum_{i,\alpha} \varphi_{i,\alpha} (1 - \Gamma_{j,\alpha})}.$$

4.2. EM Algorithm for the ExpACDM

In this subsection, we propose an EM algorithm to estimate the model parameters under the ExpACDM defined in Sect. 2.2.2. Recall that the ExpACDM is parametrized by $\eta_{j,\alpha} = \mathbf{h}(\beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k, \boldsymbol{\gamma}_j)$ with all parameters collected in $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{p})$. The complete data log-likelihood (21) under an ExpACDM can be written as:

$$\begin{aligned}\ell_c^{\text{ExpACDM}}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{p} \mid \mathbf{Y}, \mathbf{A}) &= \sum_{\alpha \in \{0,1\}^K} \sum_{i=1}^N \mathbb{1}(\mathbf{A}_i = \alpha) \log p_\alpha \\ &+ \sum_{\alpha \in \{0,1\}^K} \sum_{i=1}^N \mathbb{1}(\mathbf{A}_i = \alpha) \sum_{j=1}^J \left\{ \mathbf{h}(\beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k, \boldsymbol{\gamma}_j)^\top \mathbf{T}(Y_{i,j}) \right. \\ &\quad \left. - A(\mathbf{h}(\beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k, \boldsymbol{\gamma}_j)) + \log h(Y_{i,j}) \right\}.\end{aligned}\tag{27}$$

Based on the above expression, we can derive an EM algorithm to estimate parameters in an ExpACDM. The detailed steps are summarized in Algorithm 2.

An interesting fact is that in Algorithm 2, maximizing $(\boldsymbol{\beta}_j, \boldsymbol{\gamma}_j)$ in each M step is similar to obtaining the MLE of the regression coefficients in a *generalized linear model*, but with the observed covariates replaced with the latent attributes evaluated in the E step. In particular, in the special case of the transformed-Normal distributions, this maximization is similar to linear

Algorithm 1: EM Algorithm for the General ExpDINA Model**Data:** Responses $\mathbf{Y} = (Y_{ij})_{N \times J}$, Q-matrix $\mathbf{Q} = (q_{jk})_{J \times K}$.Initialize parameters $(\boldsymbol{\eta}, \mathbf{p})$.**while** *log-likelihood has not converged* **do**In the $(t + 1)$ th iteration,

// E step

for $(i, \boldsymbol{\alpha}) \in [N] \times \{0, 1\}^K$ **do**

$$\varphi_{i,\boldsymbol{\alpha}}^{(t+1)} = \mathbb{P}(\mathbf{A}_i = \boldsymbol{\alpha} \mid \mathbf{Y}, \boldsymbol{\eta}^{(t)}, \mathbf{p}^{(t)})$$

$$= \frac{p_{\boldsymbol{\alpha}}^{(t)} \exp \left(\sum_j [\boldsymbol{\eta}_{j,(0)}^{(t)\top} \mathbf{T}(Y_{i,j}) - A(\boldsymbol{\eta}_{j,(0)}^{(t)})] (1 - \Gamma_{j,\boldsymbol{\alpha}}) + [\boldsymbol{\eta}_{j,(1)}^{(t)\top} \mathbf{T}(Y_{i,j}) - A(\boldsymbol{\eta}_{j,(1)}^{(t)})] \Gamma_{j,\boldsymbol{\alpha}} \right)}{\sum_{\boldsymbol{\alpha}'} p_{\boldsymbol{\alpha}'}^{(t)} \exp \left(\sum_j [\boldsymbol{\eta}_{j,(0)}^{(t)\top} \mathbf{T}(Y_{i,j}) - A(\boldsymbol{\eta}_{j,(0)}^{(t)})] (1 - \Gamma_{j,\boldsymbol{\alpha}'}) + [\boldsymbol{\eta}_{j,(1)}^{(t)\top} \mathbf{T}(Y_{i,j}) - A(\boldsymbol{\eta}_{j,(1)}^{(t)})] \Gamma_{j,\boldsymbol{\alpha}'} \right)};$$

// M step

for $\boldsymbol{\alpha} \in \{0, 1\}^K$ **do**

$$p_{\boldsymbol{\alpha}}^{(t+1)} = \frac{\sum_i \varphi_{i,\boldsymbol{\alpha}}^{(t+1)}}{\sum_{i,\boldsymbol{\alpha}'} \varphi_{i,\boldsymbol{\alpha}'}^{(t+1)}},$$

for $j \in [J], h \in \{0, 1\}$ **do**

$$\boldsymbol{\eta}_{j,(h)}^{(t+1)} = \operatorname{argmax}_{\boldsymbol{\eta}} \sum_{i,\boldsymbol{\alpha}} \left[(1 - \Gamma_{j,\boldsymbol{\alpha}})^{1-h} \Gamma_{j,\boldsymbol{\alpha}}^h (\boldsymbol{\eta}^\top \mathbf{T}(Y_{i,j}) - A(\boldsymbol{\eta})) \varphi_{i,\boldsymbol{\alpha}}^{(t+1)} \right];$$

Or equivalently, solve the following gradient equations for $(\boldsymbol{\eta}_{j,(0)}^{(t+1)}, \boldsymbol{\eta}_{j,(1)}^{(t+1)})$:

$$(\nabla_{\boldsymbol{\eta}} A) (\boldsymbol{\eta}_{j,(1)}^{(t+1)}) = \frac{\sum_{i,\boldsymbol{\alpha}} \mathbf{T}(Y_{i,j}) \varphi_{i,\boldsymbol{\alpha}}^{(t+1)} \Gamma_{j,\boldsymbol{\alpha}}}{\sum_{i,\boldsymbol{\alpha}} \varphi_{i,\boldsymbol{\alpha}}^{(t+1)} \Gamma_{j,\boldsymbol{\alpha}}},$$

$$(\nabla_{\boldsymbol{\eta}} A) (\boldsymbol{\eta}_{j,(0)}^{(t+1)}) = \frac{\sum_{i,\boldsymbol{\alpha}} \mathbf{T}(Y_{i,j}) \varphi_{i,\boldsymbol{\alpha}}^{(t+1)} (1 - \Gamma_{j,\boldsymbol{\alpha}})}{\sum_{i,\boldsymbol{\alpha}} \varphi_{i,\boldsymbol{\alpha}}^{(t+1)} (1 - \Gamma_{j,\boldsymbol{\alpha}})}.$$

Output: Estimated parameters $(\boldsymbol{\eta}, \mathbf{p})$.

regression and has a closed form solution. For example, when \mathcal{P} is the Normal distribution, the maximization for $(\boldsymbol{\beta}_j, \boldsymbol{\gamma}_j)$ in the M step can be written as:

$$(\boldsymbol{\beta}_j, \boldsymbol{\gamma}_j)^{(t+1)} = \operatorname{argmax}_{\boldsymbol{\beta}_j, \boldsymbol{\gamma}_j} \sum_{i,\boldsymbol{\alpha}} \left(-\frac{(Y_{i,j} - \beta_{j,0} - \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k)^2}{2 \boldsymbol{\gamma}_j^2} - \frac{1}{2} \log \boldsymbol{\gamma}_j^2 \right) \varphi_{i,\boldsymbol{\alpha}}^{(t+1)}$$

Algorithm 2: EM Algorithm for the General ExpACDM

Data: Responses $\mathbf{Y} = (Y_{ij})_{N \times J}$, \mathbf{Q} -matrix $\mathbf{Q} = (q_{jk})_{J \times K}$
 Initialize parameters $\beta_j, \gamma_j, p_\alpha$'s.

while *log-likelihood has not converged* **do**

In the $(t + 1)$ th iteration,

// E step

for $j \in [J], \alpha \in \{0, 1\}^K$ **do**

$\eta_{j,\alpha}^{(t)} = \mathbf{h}(\beta_{j,0}^{(t)} + \sum_k \beta_{j,k}^{(t)} \alpha_k, \gamma_j^{(t)})$

for $(i, \alpha) \in [N] \times \{0, 1\}^K$ **do**

$\varphi_{i,\alpha}^{(t+1)} = \mathbb{P}(\mathbf{A}_i = \alpha \mid \mathbf{Y}, \beta_j^{(t)}, \gamma_j^{(t)}, p_\alpha^{(t)}) = \frac{p_\alpha \exp(\sum_j [\eta_{j,\alpha}^{(t)\top} \mathbf{T}(Y_{i,j}) - A(\eta_{j,\alpha}^{(t)})])}{\sum_{\alpha'} p_{\alpha'} \exp(\sum_j [\eta_{j,\alpha'}^{(t)\top} \mathbf{T}(Y_{i,j}) - A(\eta_{j,\alpha'}^{(t)})])};$

// M step

for $\alpha \in \{0, 1\}$ **do**

$p_\alpha^{(t+1)} = \frac{\sum_i \varphi_{i,\alpha}^{(t+1)}}{\sum_{i,\alpha'} \varphi_{i,\alpha'}^{(t+1)}}$

for $j \in [J]$ **do**

$(\beta_j, \gamma_j)^{(t+1)} = \underset{\beta_j, \gamma_j}{\operatorname{argmax}} \sum_{i,\alpha} \left[\mathbf{h}(\beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k, \gamma_j)^\top \mathbf{T}(Y_{i,j}) \varphi_{i,\alpha}^{(t+1)} \right. \\ \left. - A(\mathbf{h}(\beta_{j,0} + \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k, \gamma_j)) \varphi_{i,\alpha}^{(t+1)} \right]$

Output: $\beta_j, \gamma_j, p_\alpha$'s.

for any $j \in [J]$. So for those $q_{j,k} = 1$, we can get the updates

$$\hat{\beta}_{j,k}^{(t+1)} = \frac{\sum_{i,\alpha} Y_{i,j} \alpha_k \varphi_{i,\alpha}^{(t+1)}}{\sum_{i,\alpha} \alpha_k \varphi_{i,\alpha}^{(t+1)}}, \quad \hat{\gamma}_j^{(t+1)} = \sqrt{\frac{\sum_{i,\alpha} (Y_{i,j} - \beta_{j,0} - \sum_{k=1}^K \beta_{j,k} q_{j,k} \alpha_k)^2 \varphi_{i,\alpha}^{(t+1)}}{N}}.$$

The updates under the transformed-Normal distributions can be obtained similarly by applying the corresponding transform of $Y_{i,j}$. For other members in the exponential family beyond the transformed-Normal, maximizing (β_j, γ_j) may not have closed forms and one may use available optimization software to find $(\hat{\beta}_j, \hat{\gamma}_j)$ in the M step.

Although we have focused on describing EM algorithms for ExpDINA and ExpACDM in this section, we remark that an EM algorithm with closed-form M step can be similarly developed for the ExpGDM (ExpGDINA). Furthermore, the exponential family assumption in Algorithms

1 and 2 is not essential to our EM procedures. In Supplementary Material S.4, we demonstrate how our algorithms can be modified to estimate the negative-binomial-based DINA and negative-binomial-based ACDM.

5. Simulation Studies

We conduct simulation studies under various models in the proposed family. We have two goals in the simulations: (a) to empirically verify the theoretical results of identifiability and consistency; and (b) to assess the computational performance of the proposed EM algorithms. Under the ExpDINA and the ExpACDM, we consider the Normal and transformed Normal distributions (i.e., lognormal and logistic normal) for continuous responses in Sect. 5.1, and the Poisson and negative binomial distributions for count responses in Sect. 5.2. We remark that in addition to the distributions considered in this section, it is also easy to use our framework and estimation procedures for other exponential family distributions such as Gamma and Beta to model continuous positive, and continuous bounded data, respectively.

5.1. Simulations Under the Normal- and Transformed Normal-CDMs

We first describe the true parameter settings used in the simulations. In all simulations, we set the \mathbf{Q} -matrix and proportion parameter \mathbf{p} as follows. Consider $K = 5$ latent attributes and $J = 20$ items. The \mathbf{Q} -matrix takes the form

$$\mathbf{Q} = \begin{pmatrix} \mathbf{I}_K \\ \mathbf{I}_K \\ \mathbf{I}_K \\ \mathbf{Q}_1 \end{pmatrix}, \quad \text{where } \mathbf{Q}_1 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & \ddots & \ddots \\ \ddots & \ddots & 1 \\ 0 & 1 & 1 \end{pmatrix}_{K \times K}. \quad (28)$$

The above \mathbf{Q} -matrix satisfies our strict identifiability conditions in Theorem 1 and Proposition 2. The proportion parameters are set to be uniform with $p_{\alpha} = 1/2^K$ for all $\alpha \in \{0, 1\}^K$. We consider varying sample sizes $N = 100, 500, 1000, 1500, 2000$. In each setting, 100 independent simulation replicates are performed. Under the Normal-DINA model, we set the true item parameters as

$$\mu_{j,0} = -1, \quad \mu_{j,1} = 2, \quad \sigma_{j,0} = 1, \quad \sigma_{j,1} = 1.$$

Under the Normal-ACDM, we set the coefficients $\beta_{j,k}$ by

$$\beta_{j,0} = -1; \quad \beta_{j,k} = \frac{3}{\sum_{k'=1}^K q_{j,k'}} \mathbb{1}(q_{j,k} = 1), \quad \forall j \in [J], k \in [K].$$

The variance parameter $\gamma_j = \sigma_j^2$ is fixed to be 1 for all j .

Recall that in Example 1, we have mentioned that general-response CDMs based on any transformed Normal distributions (such as lognormal and logistic Normal) are equivalent to the Normal-based CDM. This fact implies that the estimation procedures for these models are identical up to an invertible transformation in the M step. In preliminary simulations, we have tried

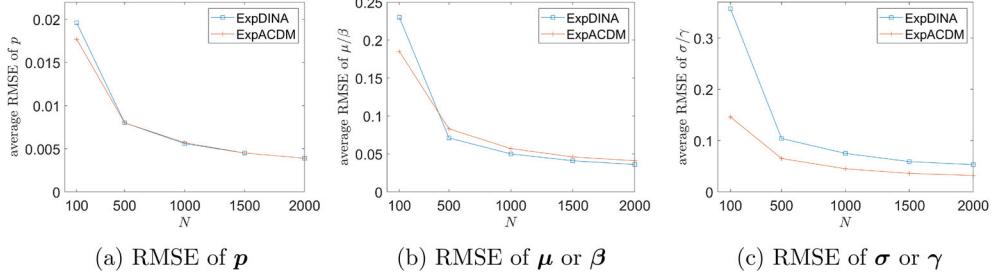


FIGURE 2.
RMSEs of parameters under the Normal-DINA (red) and Normal-ACDM (blue).

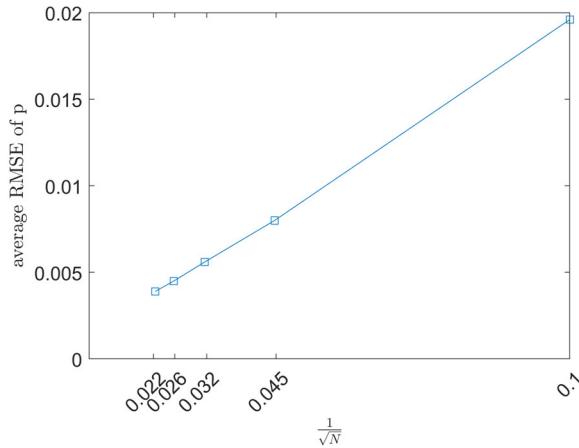


FIGURE 3.
RMSEs of p with respect to $\frac{1}{\sqrt{N}}$ under the Normal-DINA model.

estimating the Normal-based and lognormal-based CDMs independently, and the estimation accuracies for the two models are exactly the same. Hence, we only report the estimation accuracy for the Normal-based CDM in this section.

In each of the $C = 100$ independent replicates, we generate data using the above parameter settings and fit our EM Algorithms 1 or 2 with a random initialization. We calculate the root average mean squared error (RMSE) of the proportion parameters and item parameters based on the simulation replicates. Figure 2 displays the average RMSEs for the Normal-DINA and Normal-ACDM. In each simulation setting, the RMSE is defined as

$$\sqrt{\frac{1}{C} \sum_{c=1}^C \frac{\|\widehat{\boldsymbol{\theta}}^{(c)} - \boldsymbol{\theta}_0\|_2^2}{\dim(\boldsymbol{\theta}_0)}},$$

where $\boldsymbol{\theta}_0$ is the true parameter vector, $\widehat{\boldsymbol{\theta}}^{(c)}$ is the estimator in the c -th simulation replication, and $\dim(\boldsymbol{\theta}_0)$ denotes the dimension of the vector $\boldsymbol{\theta}_0$. The exact RMSE values are included in Tables S.1, S.2, and S.3 in the Supplementary Material.

Figure 2 clearly shows that the RMSE decreases as the sample size N increases. Furthermore, in Fig. 3, we plot the average RMSE of \mathbf{p} under the Normal-DINA model versus $1/\sqrt{N}$, and it is clear that the RMSE is linear with respect to $1/\sqrt{N}$. This observation empirically validates that our estimation procedure is statistically consistent and converges at the usual parametric rate $1/\sqrt{N}$. In addition, the above simulation results also show that our EM algorithms have good computational performance and can efficiently find the MLE.

In the Supplementary Material S.5.2, we conduct additional simulation studies with alternative \mathbf{Q} -matrices that do not satisfy the strict identifiability conditions in Theorem 1, but satisfy the generic identifiability conditions in Theorem 2. The simulation results empirically show that in such settings, the model parameters can still be consistently estimated by our proposed method.

5.2. Simulations Under the Poisson- and Negative Binomial-CDMs

We also conduct simulation studies for general-response CDMs for multivariate count data. Here, we consider Poisson and negative binomial distributions to model the count responses. The \mathbf{Q} -matrix and the proportion parameters \mathbf{p} are set to be the same as described in the previous subsection. The distribution-specific item parameters are set as follows. For the Poisson-DINA model, we set the item parameters as $\lambda_{j,0} = 1$ and $\lambda_{j,1} = 3$. For the Poisson-ACDM, we set the true model parameters as

$$\beta_{j,0} = 1, \quad \beta_{j,k} = \frac{2}{\sum_{k'=1}^K q_{j,k'}} \mathbb{1}(q_{j,k} = 1), \quad \forall j \in [J], \quad k \in [K].$$

For the negative binomial-DINA model (NegBin-DINA, defined in Supplementary Material S.1), we set the item parameters as

$$r_{j,0} = 1, \quad r_{j,1} = 3, \quad \pi_{j,0} = 0.5, \quad \pi_{j,1} = 0.5, \quad \forall j \in [J],$$

where $r_{j,0}, r_{j,1}$ are the number of successes and $\pi_{j,0}, \pi_{j,1}$ are the success probability in a negative binomial distribution. For the negative binomial-ACDM (NegBin-ACDM, defined in Supplementary Material S.1), we set the true model parameters as

$$\beta_{j,0} = 1, \quad \beta_{j,k} = \frac{2}{\sum_{k'=1}^K q_{j,k'}} \mathbb{1}(q_{j,k} = 1), \quad \gamma_j = \pi_j = 0.5, \quad \forall j \in [J], \quad k \in [K].$$

Figures 4 and 5 report the RMSEs of the estimated parameters obtained from the replicated simulations. Similar to the Normal-based CDMs, the RMSEs here also decrease as N increases, at the typical $1/\sqrt{N}$ rate.

Finally, we remark that our estimation methods are computationally quite efficient. In the simulation settings considered in this section, the computation time taken by our EM algorithm is less than one minute on average even for a sample size as large as $N = 2000$. We present the average number of EM iterations and computation time in Supplementary Material S.5.1.

6. Application to the TIMSS 2019 Response Time Data

We demonstrate the proposed general-response CDM by applying it to a response time dataset extracted from the TIMSS 2019 assessment (Fishbein et al., 2021). This dataset is from the mathematics assessment data of eighth-grade students in the United States. It consists of each

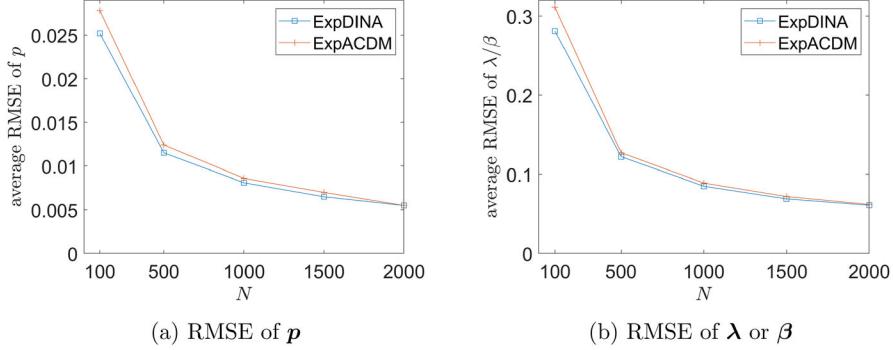


FIGURE 4.
RMSEs of parameters under the Poisson-DINA (red) and Poisson-ACDM (blue).

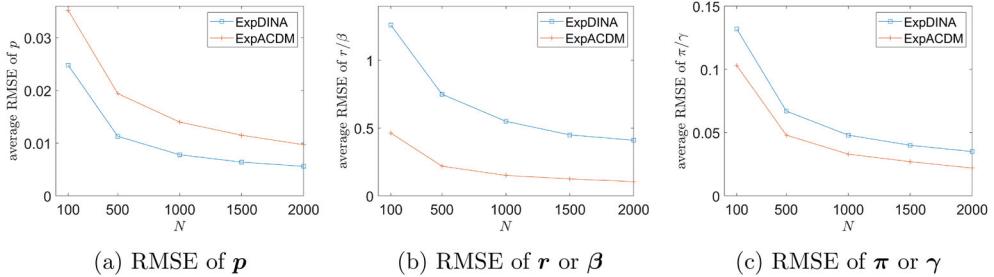


FIGURE 5.
RMSEs of parameters under the NegBin-DINA (red) and NegBin-ACDM (blue).

student's time spent on each item screen (in seconds). If a question has sub-questions that share the same screen, all the sub-questions combined are regarded as a single question, and the overall response time is recorded. We focus on the students who received booklet number 14. After data preprocessing (see details in Supplementary Material S.5), the dataset consists of $N = 620$ students' response times to $J = 29$ items.

The TIMSS 2019 mathematics assessment aims to measure four content skills (Number, Algebra, Geometry, and Data and Probability) and three cognitive skills (Knowing, Applying, and Reasoning). The TIMSS 2019 database specifies how each of the $J = 29$ items measures one content skill and one cognitive skill. Based on this information, we construct a 29×7 \mathbf{Q} -matrix with $K = 7$ skill attributes: A_1 : Number, A_2 : Algebra, A_3 : Geometry, A_4 : Data and Probability, A_5 : Knowing, A_6 : Applying, A_7 : Reasoning. Each row of this \mathbf{Q} -matrix has exactly two nonzero entries in one content skill and one cognitive skill. We provide the details of this \mathbf{Q} -matrix in Table 2. It is not hard to verify that this \mathbf{Q} -matrix satisfies our generic identifiability conditions in Theorem 2 under any general-response ExpACDMs, and we choose to use the lognormal-ACDM to analyze this dataset. The TIMSS database also provides additional item information, including a brief description of the item type (whether it is a multiple choice item or a constructed response item), and the correct response percentage among the U.S. students. We present this information in Table 3.

Given the \mathbf{Q} -matrix in Table 2, we fit the lognormal-ACDM using Algorithm 2. In addition to the identifiability consideration, our rationale for adopting the additive model assumption is that in order to solve each problem, students need to perform operations (i.e., attributes) that are

TABLE 2.
Q-matrix for the TIMSS 2019 math assessment booklet 14.

| Item ID. | Number | Algebra | Geometry | Data and Prob. | Knowing | Applying | Reasoning |
|----------|--------|---------|----------|----------------|---------|----------|-----------|
| 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 2 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 3 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 4 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| 6 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 7 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 8 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| 9 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 10 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| 11 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| 12 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 13 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 14 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 15 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 16 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 17 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 18 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 19 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 20 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| 21 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 22 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| 23 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 24 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 25 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 26 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| 27 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 28 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 29 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |

specified by the **Q**-matrix additively. We assume that each operation is carried out separately, and the total log-time is the sum of the log-times for each operation. For instance, the question “Value of X in $10/15 = X/18$ ” (this is the 15th question in our dataset; see Tables 2 and 3 for more details) measures the content skill “Number” and cognitive skill “Knowing”, and we assume that the log-time is the sum of:

- $\beta_{j,0}$: how long it takes to read the problem and click/type the answer;
- $\beta_{j,K} A_K$: how long it takes to re-formulate the problem as “ $10 \times 18 \div 15 = X$ ”, regarding whether a student “knows” the necessary concepts, i.e., whether the student possesses the cognitive skill “Knowing”;
- $\beta_{j,N} A_N$: how long it takes to compute and find the answer $X = 12$, regarding whether the student possesses the content skill “Number”.

TABLE 3.

Additional item information for items in TIMSS 2019 math assessment booklet 14. Starred (*) items are items that are composed of sub-questions; for these items, we display the smallest correct response percentage among all sub-questions.

| Item | Description | Item type | % Correct response |
|------|--|-----------|--------------------|
| 1 | Octagon with equivalent shading | MC | 55 |
| 2 | Percentage of laps finished | CR | 32* |
| 3 | Multiples of 3 | MC | 42 |
| 4 | Convert decimal to a fraction | CR | 42 |
| 5 | Expression for area of rectangle | MC | 51 |
| 6 | Ratio of boys | MC | 56 |
| 7 | Rule for number of matches | CR | 11* |
| 8 | Graph of $y = 2x$ | MC | 41 |
| 9 | Rotation and reflection | MC | 34 |
| 10 | Surface area of the prism | MC | 41 |
| 11 | Value of angle x outside triangle | MC | 44 |
| 12 | Number of balls in a bag | MC | 69 |
| 13 | Liv's smartphone use | MC | 35* |
| 14 | Number with most factors | MC | 54 |
| 15 | Value of X in $10/15 = X/18$ | CR | 50 |
| 16 | Make two fractions with largest product | CR | 41 |
| 17 | Ratio of boys to girls - eighth grade | CR | 53* |
| 18 | Expression for money in savings account | MC | 73 |
| 19 | Value of k when $(k + 7) / 3 = 6$ | CR | 60 |
| 20 | Formula for students' overall scores in course | CR | 54 |
| 21 | Expression equivalent to $x + 2y + 6$ | MC | 42 |
| 22 | Pair of equations for ticket price (Derived) | MC | 30 |
| 23 | Shape reflected over dotted line (Derived) | MC | 58 |
| 24 | Value of angle x with parallel lines a and b | CR | 11 |
| 25 | Areas of triangles ABC and ABD | CR | 8 |
| 26 | Surface area of rectangular prism | CR | 19 |
| 27 | Bar graph of Internet users in world | MC | 60 |
| 28 | Graph that best shows favorite fruit | MC | 65 |
| 29 | Mean heights of basketball team | CR | 30 |

Summing up the above components and adding a normal error gives the following response time distribution, which corresponds to the lognormal-ACDM defined in Example 3:

$$\log Y_{i,j} \mid \mathbf{A} = \underset{\text{intercept}}{\beta_{j,0}} + \underset{\text{cognitive skill}}{\beta_{j,K} A_K} + \underset{\text{content skill}}{\beta_{j,N} A_N} + \epsilon_{i,j}, \text{ where } \epsilon_{i,j} \stackrel{\text{i.i.d.}}{\sim} N(0, \gamma_j).$$

Decomposing the response time into several components has a long history. Sternberg (1969) and Sternberg (1980) considered a linear regression model based on a sequence of hypothetical processes that students go through to solve a problem. However, these studies do not introduce individual-level latent variables. Maris (1993) considered a latent variable model with Gamma distribution for the response time, with a similar additive combination of the latent variables. These models are also referred to as exploratory response time models in the literature (e.g., De Boeck and Jeon, 2019). The recent papers Minchen et al. (2017) and Minchen and de la Torre (2018) incorporate this line of thinking into a CDM framework with the lognormal distribution for modeling the response times. The empirical results in these papers show that students possessing

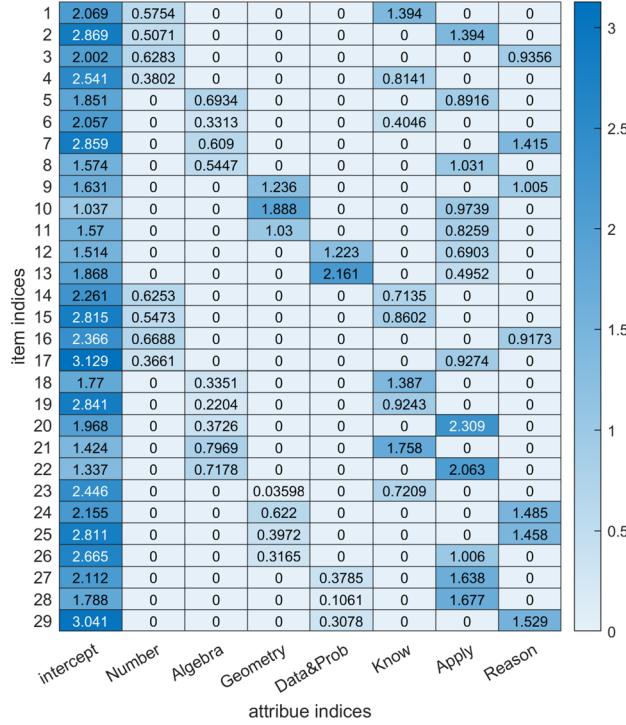


FIGURE 6.

Heatmap of the estimated β -parameters under the lognormal-ACDM for the TIMSS 2019 response time dataset.

more required attributes take a longer time to respond, so we anticipate that our main-effect coefficients $\beta_{j,k}$ for the attributes are nonnegative.

We run our EM Algorithm 2 with 20 random parameter initializations with $\beta_{j,k} > 0$. We do not impose any strict constraints on the sign of $\beta_{j,k}$ and allow the algorithm to update parameters in the unconstrained space. Among the resulting 20 parameter estimates, we select the estimate with the largest log-likelihood. Figure 6 presents the estimated main-effect coefficients. This figure shows that students spend the most time reading the question and clicking/typing the answer (represented by the intercepts in the first column), compared to the time it takes to re-formulate and solve the problem (represented by the main-effect coefficients from the second column to the last column). The intercept values vary a lot across different questions, ranging from 1.0 to 3.1. This indicates that the length and the abstractness of the questions vary substantially across items. One interesting observation is the relation between the magnitude of the estimated intercept $\beta_{j,0}$ and the type of each item. Recall that Table 3 provides information about whether each item is a multiple choice item or a constructed response item. Nonetheless, Fig. 6 reveals that, among the first ten items, the three items that ask students to construct the response (items 2, 4, 7) have the largest intercept values. This is clearly visible by looking at the darker-colored entries (or the numbers shown in white) in the first column of Fig. 6. This result can be interpreted as that students spend more time typing and checking the answer when solving constructed response items. Note that the item type information is *not used* in our estimation procedure, but our method can automatically distinguish the multiple choice items from the constructed response items via the estimated parameters.

As for the main-effect coefficients $\beta_{j,k}$ for $k \geq 1$, the computation time represented by the content skills' coefficients and formulation time represented by the cognitive skills' coefficients

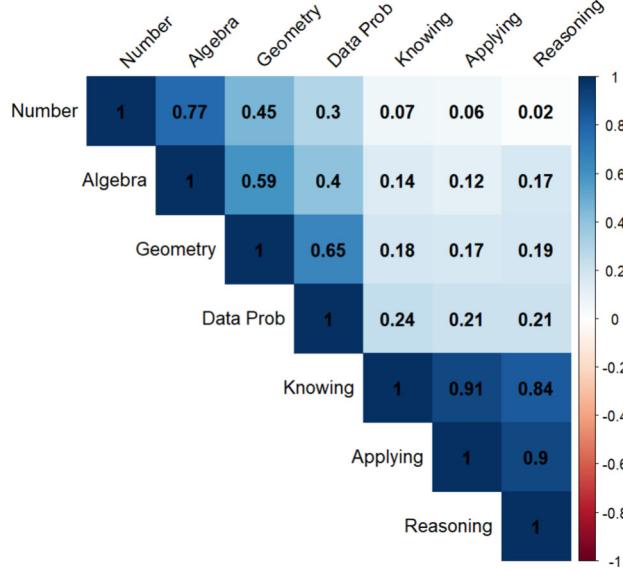


FIGURE 7.

Correlation plot of the latent attributes under the estimated lognormal-ACDM for the TIMSS 2019 response time dataset.

also differ a lot across the items. These coefficients range from 0.03 to 2.3. Even though we do not constrain $\beta_{j,k}$ to be positive, all estimated main-effect coefficients turn out to be positive, which is consistent with empirical findings in previous studies of response time modeling. This observation also indicates that our additive model assumption is indeed plausible and yields interpretable parameter estimates.

We also provide the correlation plot of the estimated latent attributes in Fig. 7. The EM algorithm does not estimate the individual latent attributes directly, but instead computes the conditional expectations $\hat{\varphi}_{i,\alpha} = \mathbb{P}(\mathbf{A}_i = \alpha \mid \mathbf{Y}, \hat{\eta}, \hat{\mathbf{p}})$ for all $i \in [N]$ and $\alpha \in \{0, 1\}^K$. Therefore, we estimate the latent attribute profiles $\hat{\mathbf{A}}_i$ by

$$\hat{\mathbf{A}}_i = \underset{\alpha \in \{0, 1\}^K}{\operatorname{argmax}} \hat{\varphi}_{i,\alpha},$$

and use these estimated $\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_N$ to compute the sample correlation between the attributes. Figure 7 reveals interesting observations about the intrinsic dependence among the attributes. There are higher correlations among the four content skills and also higher correlations among the three cognitive skills, compared to the correlations between one content skill and one cognitive skill. This phenomenon not only supports that our model is a reasonable choice as it outputs interpretable correlation structures among latent attributes, but also implies that it may be plausible to model high-order latent traits behind the seven fine-grained attributes (de la Torre & Douglas, 2004; Zhan et al., 2018a). Future studies are warranted to explore the identifiability and interpretability of such high-order extensions of our general-response CDMs.

To assess the model fit, we use the popular Bayesian information criterion (BIC) and compare the lognormal-ACDM to an alternative lognormal-DINA model in terms of the BIC values. In order to satisfy the identifiability conditions in Theorem 1 under the lognormal-DINA model, and also motivated by the high correlation among the cognitive skills in Fig. 7, we consider a more parsimonious lognormal-DINA model with $K = 4$ content skills only. The corresponding

Q-matrix for lognormal-DINA has four columns, which are the first four columns of the **Q**-matrix for lognormal-ACDM in Table 2. The BIC value for the lognormal-ACDM is 17,005, whereas the BIC for the lognormal-DINA (fitted via Algorithm 1) is 17,067. Based on this result, we conclude that the lognormal-ACDM with seven attributes is more suitable and fits better to this dataset, despite its higher model complexity.

Furthermore, in Supplementary Material S.6.2, we also provide an analysis of the binary response accuracy data. We find an interesting relationship between the item parameters estimated from the response time data and those from the response accuracy data, as well as a similar connection between the intercept (guessing parameter) of an item and the type of the item; see the Supplementary Material for details.

7. Discussion

In this paper, we have proposed a flexible new framework of cognitive diagnostic models for multivariate general responses beyond the traditional binary or polytomous responses. Our modeling framework incorporates the **Q**-matrix constraints in a unified way and covers popular existing CDMs as submodels. An important contribution of this work is to provide the crucial identifiability theory for all these general-response CDMs. Interestingly and somewhat surprisingly, we have shown that the general-response CDMs are identifiable under similar conditions on the **Q**-matrix as the binary-response CDMs. Our identifiability theory has the nice implication of consistent parameter estimation via the maximum likelihood. For computation, we have proposed an efficient EM algorithm for parameter estimation under various response types. Extensive simulation studies not only corroborate the identifiability conclusions, but also demonstrate the favorable computational performance of our algorithms. We have analyzed a response time dataset from the TIMSS 2019 assessment using the proposed lognormal-ACDM and obtained interpretable results.

The proposed new paradigm of identifiable general-response cognitive diagnostic models open up a number of interesting possibilities for future research. *First*, our current identifiability results and estimation procedure assume that the **Q**-matrix is known. In practice, many modern assessment datasets may not come with a readily available **Q**-matrix. In these exploratory settings, it would be interesting to directly identify and estimate the **Q**-matrix along with other model parameters for general-response CDMs. In terms of identifiability, we conjecture that our identifiability proof technique could be generalized to also handle the identifiability of the **Q**-matrix. In terms of estimation, it may be possible to extend existing methods for estimating the **Q**-matrix such as penalized likelihood methods in Chen et al. (2015) and Ma et al. (2023), or Bayesian approaches in Chen et al. (2018) and Liu et al. (2020) from the binary-response CDMs to general-response CDMs. We plan to pursue these directions of exploratory general-response CDMs in the future.

Second, it would be interesting to explore possibilities of strengthening our identifiability results by developing weaker identifiability conditions for special submodels of general-response CDMs. Our main theorems provide transparent sufficient conditions for identifiability in a very general setting. But for binary-response CDMs under some special modeling assumptions, existing studies show that there exist weaker conditions that can guarantee identifiability. For example, for the DINA model with binary responses, Gu and Xu (2019) proposed weaker conditions that are necessary and sufficient for strict identifiability. Establishing necessary identifiability conditions in the most general setting with arbitrary response types is a non-trivial but interesting future direction.

Third, many educational assessments naturally contain multiple types of responses for each item. For instance, it is common to record the response accuracy and visual fixation counts/visit

counts in addition to the response time for each item (Zhan et al., 2022; Fishbein et al., 2021). Jointly modeling multiple types of responses has received great attention in the measurement literature (van der Linden, 2007; Molenaar et al., 2015; Zhan et al., 2018b; Wang et al., 2018; Man & Harring, 2022; Kang et al., 2023). Among these modeling approaches, CDM-based methods such as Zhan et al. (2018b) and Zhan et al. (2022) use binary latent variables to model the response accuracy only (i.e., binary responses of correct or wrong answers to items), and use an IRT-based model with continuous latent variables to model the response time and fixation counts. In fact, as suggested by our analysis of the TIMSS response time data, using binary latent skills to model the response time can also yield interpretable results. Therefore, this work offers useful insights for future research to propose identifiable general-response CDMs to jointly model multiple types of responses, such as one's response accuracy, response time, and visit counts to an item. We believe that our ExpCDM framework is flexible enough to allow for such extensions. Indeed, one naive model would be to assume conditional independence between the response accuracy (Y_{ij}) and the response time (T_{ij}) given one's latent attributes (\mathbf{A}_i), and we conjecture that similar identifiability conditions as proposed in this work still suffice for identifiability in this setting. More investigations about modeling methodology and identifiability theory along this line are left for future research.

Acknowledgments

This research is partially supported by NSF grant DMS-2210796. The authors thank the editor Prof. Sandip Sinharay and three reviewers for their helpful and constructive comments.

Code availability The Matlab codes implementing the proposed methods are available via this <https://github.com/seunghyunstats/General-Response-CDM/tree/main>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

References

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

Casella, G., & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.

Chen, Y., Culpepper, S., & Liang, F. (2020). A sparse latent class model for cognitive diagnosis. *Psychometrika*, 85(1), 121–153.

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

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

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

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

De Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. *Frontiers in Psychology*, 10, 102.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2), 179–199.

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

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 39(1), 1–22.

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. *Cognitively diagnostic assessment*, 361389.

Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 62(2), 355–366.

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

Fishbein, B., Foy, P., & Yin, L. (2021). TIMSS 2019 user guide for the international database. *Hentet fra <https://timssandpirls.bc.edu/timss2019/international-database>*.

Gu, Y., & Dunson, D. B. (2023). Bayesian Pyramids: identifiable multilayer discrete latent structure models for discrete data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(2), 399–426.

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

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

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

He, Q., & Von Davier, M. (2016). Analyzing process data from problem-solving items with n-grams: Insights from a computer-based large-scale assessment. In *Handbook of research on technology tools for real-world skill development*, pages 750–777. IGI Global.

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.

Kang, I., Jeon, M., & Partchev, I. (2023). A latent space diffusion item response theory model to explore conditional dependence between responses and response times. *Psychometrika*, pages 1–35.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

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

Liu, C.-W., Andersson, B., & Skrondal, A. (2020). A constrained Metropolis-Hastings Robbins–Monro algorithm for Q matrix estimation in DINA models. *Psychometrika*, 85(2), 322–357.

Liu, R., Heo, I., Liu, H., Shi, D., & Jiang, Z. (2023). Applying negative binomial distribution in diagnostic classification models for analyzing count data. *Applied Psychological Measurement*, 47(1), 64–75.

Liu, R., Liu, H., Shi, D., & Jiang, Z. (2022). Poisson diagnostic classification models: A framework and an exploratory example. *Educational and Psychological Measurement*, 82(3), 506–516.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 6, 1171.

Loeys, T., Rosseel, Y., & Baten, K. (2011). A joint modeling approach for reaction time and accuracy in psycholinguistic experiments. *Psychometrika*, 76, 487–503.

Ma, C., Ouyang, J., & Xu, G. (2023). Learning latent and hierarchical structures in cognitive diagnosis models. *Psychometrika*, 88(1), 175–207.

Magnus, B. E., & Thissen, D. (2017). Item response modeling of multivariate count data with zero inflation, maximum inflation, and heaping. *Journal of Educational and Behavioral Statistics*, 42(5), 531–558.

Man, K., & Harring, J. R. (2019). Negative binomial models for visual fixation counts on test items. *Educational and Psychological Measurement*, 79(4), 617–635.

Man, K., & Harring, J. R. (2022). Detecting preknowledge cheating via innovative measures: A mixture hierarchical model for jointly modeling item responses, response times, and visual fixation counts. *Educational and Psychological Measurement*, page 00131644221136142.

Maris, E. (1993). Additive and multiplicative models for gamma distributed random variables, and their application as psychometric models for response times. *Psychometrika*, 58, 445–469.

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

Minchen, N., & de la Torre, J. (2018). A general cognitive diagnosis model for continuous-response data. *Measurement: Interdisciplinary Research and Perspectives*, 16(1), 30–44.

Minchen, N., de la Torre, J., & Liu, Y. (2017). A cognitive diagnosis model for continuous response. *Journal of Educational and Behavioral Statistics*, 42(6), 651–677.

Molenaar, D., Tuerlinckx, F., & van der Maas, H. L. (2015). A bivariate generalized linear item response theory modeling framework to the analysis of responses and response times. *Multivariate Behavioral Research*, 50(1), 56–74.

Moustaki, I., & Knott, M. (2000). Generalized latent trait models. *Psychometrika*, 65, 391–411.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3), 370–384.

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests*. ERIC.

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

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Crc Press.

Sternberg, R. J. (1980). Representation and process in linear syllogistic reasoning. *Journal of Experimental Psychology: General*, 109(2), 119.

Sternberg, S. (1969). The discovery of processing stages: Extensions of donders' method. *Acta psychologica*, 30, 276–315.

Tang, X., Wang, Z., He, Q., Liu, J., & Ying, Z. (2020). Latent feature extraction for process data via multidimensional scaling. *Psychometrika*, 85(2), 378–397.

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

Teicher, H. (1967). Identifiability of mixtures of product measures. *The Annals of Mathematical Statistics*, 38(4), 1300–1302.

Thissen, D. (1983). Timed testing: An approach using item response theory. In *New horizons in testing*, pages 179–203. Elsevier.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, 31(2), 181–204.

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, 72(3), 287–308.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61, 287–307.

Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology*, 68(3), 456–477.

Wang, S., Zhang, S., Douglas, J., & Culpepper, S. (2018). Using response times to assess learning progress: A joint model for responses and response times. *Measurement: Interdisciplinary Research and Perspectives*, 16(1), 45–58.

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

Xu, G., & Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association*, 113(523), 1284–1295.

Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, 81(3), 625–649.

Yakowitz, S. J., & Spragins, J. D. (1968). On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 39(1), 209–214.

Zhan, P., Jiao, H., & Liao, D. (2018). Cognitive diagnosis modelling incorporating item response times. *British Journal of Mathematical and Statistical Psychology*, 71(2), 262–286.

Zhan, P., Liao, M., & Bian, Y. (2018). Joint testlet cognitive diagnosis modeling for paired local item dependence in response times and response accuracy. *Frontiers in Psychology*, 9, 607.

Zhan, P., Man, K., Wind, S. A., & Malone, J. (2022). Cognitive diagnosis modeling incorporating response times and fixation counts: Providing comprehensive feedback and accurate diagnosis. *Journal of Educational and Behavioral Statistics*.

Manuscript Received: 20 OCT 2023