PSYCHOMETRIKA
https://doi.org/10.1007/s11336-024-09962-9

DIAGNOSTIC CLASSIFICATION MODELS FOR TESTLETS: METHODS AND THEORY

XIN XU

BEIJING NORMAL UNIVERSITY

GUANHUA FANG

FUDAN UNIVERSITY

JINXIN GUO

MINZU UNIVERSITY OF CHINA

ZHILIANG YING

COLUMBIA UNIVERSITY

SUSU ZHANG

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Diagnostic classification models (DCMs) have seen wide applications in educational and psychological measurement, especially in formative assessment. DCMs in the presence of testlets have been studied in recent literature. A key ingredient in the statistical modeling and analysis of testlet-based DCMs is the superposition of two latent structures, the attribute profile and the testlet effect. This paper extends the standard testlet DINA (T-DINA) model to accommodate the potential correlation between the two latent structures. Model identifiability is studied and a set of sufficient conditions are proposed. As a byproduct, the identifiability of the standard T-DINA is also established. The proposed model is applied to a dataset from the 2015 Programme for International Student Assessment. Comparisons are made with DINA and T-DINA, showing that there is substantial improvement in terms of the goodness of fit. Simulations are conducted to assess the performance of the new method under various settings.

Key words: diagnostic classification model, testlet DINA, identifiability, PISA, Q-matrix, interaction, hypothesis testing, model selection.

In educational and psychological assessments, there is a surge of interest in adapting psychometric models to informed learning. Diagnostic classification models (DCMs; see Rupp et al. 2010), with the capability of measuring examinees' mastery of fine-grained skills, have gained increasing attention and popularity. Over the past few decades, a wide range of DCMs has been developed, for example, the deterministic inputs, noisy "and" gate (DINA; Macready & Dayton, 1977; Junker & Sijtsma, 2001) model, the deterministic inputs, noisy "or" gate (DINO; Templin & Henson, 2006) model, the generalized DINA (G-DINA; de la Torre, 2011) model, and the log-linear CDM (LCDM; Henson et al., 2009).

One fundamental assumption in latent variable modeling is the local independence of item responses given the measured trait or profile. Practical assessments, however, may comprise testlets (e.g., Bradlow et al., 1999), that is, sets of items based on a common stimulus, for instance, a passage in a reading assessment. Correlation among item responses within a testlet often cannot

Correspondence should be made to Susu Zhang, University of Illinois Urbana-Champaign, Champaign, USA. Email: szhan105@illinois.edu

Published online: 26 March 2024

be fully explained by the measured trait or profile, and such a testlet effect leads to the violation of the commonly assumed local independence. Under item response theory (IRT) models for continuous traits, studies have shown that ignoring the testlet effect leads to biased estimation of parameters and standard errors (e.g., Bradlow et al., 1999; Wainer et al., 2007; Sireci et al., 1991). Under DCMs, Hansen (2013) and Chen et al. (2018) discussed the consequences of ignoring testlet effects, including inaccurate estimation of parameters and misclassification of examinee attribute profiles, especially when testlet effects are large.

To account for the possible local dependence due to the testlet effect, a commonly adopted approach in compensatory IRT is to superimpose a testlet-specific random effect term on each testlet in addition to the general trait(s) to be measured by the whole test. The response distribution of an item within a testlet is typically assumed to depend on a linear combination of the general trait(s) and the corresponding testlet-specific random effect term. Examples of this include the bi-factor model (e.g., Gibbons & Hedeker, 1992; DeMars, 2006), the two-tier model (Cai, 2010), and the testlet-effect model Bradlow et al. (1999). The general and testlet-specific traits are usually assumed to be normally distributed, and three additional assumptions are required: (1) the primary and testlet-specific traits jointly work to capture the local dependency within a testlet; (2) the testlets are assumed to be mutually exclusive, meaning each item loads on at most one testlet-specific dimension; (3) the general and specific effects are independent or, equivalently under the normal distribution, uncorrelated. Assumption (1) is the fundamental and essential idea behind incorporating testlet effects. Assumption (2) reduces computational complexity in parameter estimation (e.g., Cai, 2010), and the independence between testlet effects can be relaxed to model the additional correlation between specific dimensions unexplained by the general trait (e.g., Jennrich & Bentler, 2012). Assumption (3), uncorrelated general and testlet effects, is essential, as Fang et al. (2021) proved that bi-factor-type models treating general and specific effects as dependent cannot be statistically identified.

For diagnostic assessments with testlets, similar testlet-effect DCMs have been proposed, where a testlet-specific random effect is incorporated into the item response function (Hansen, 2013; Hansen et al., 2016; Zhan et al., 2015, 2018; Ma et al., 2023) : Assumption (1) above straightforwardly carries over to testlet-effect DCMs. Specifically, the proposed models leverage the fact that both general and specific DCMs can be parameterized as a logistic model (e.g., Henson et al., 2009; de la Torre, 2011), where the linear component contains attribute main effects and higher-order interactions. Subsequently, a testlet random effect is included as an additive term to the linear component to account for additional within-testlet dependency. Assumption (2), each item loading on at most one testlet effect, is also imposed in most models, with the exceptions of Zhan et al. (2015, 2018), where in general, an item is allowed to crossload on multiple testlet effects. To impose Assumption (3) in testlet-effect DCMs, existing models still imposed the structural independence assumption, by assuming that the testlet-specific traits each stem from an independent normal distribution, either uncorrelated with a higher-order continuous trait (de la Torre & Douglas, 2004) underlying the discrete attributes (Hansen, 2013; Hansen et al., 2016; Zhan et al., 2015, 2018) , or independent from the attribute pattern stemming from a multinomial distribution over permissible latent classes (Ma et al., 2023) .

In the current paper, however, we show that imposing Assumption (3) under testlet-effect DCMs can be restrictive, resulting in a model that does not adequately account for the within-testlet local dependency. The reasoning is as follows: In confirmatory IRT and DCMs, focal traits are typically expert-defined and cannot be rotated without altering substantive interpretations. A testlet effect, which enters the measurement model as an additive term, defines a person-specific effect that explains the dependency in responses within the same testlet that has not been accounted for by the existing terms in the measurement model. For compensatory IRT models, it is natural to restrict the testlet effect to be independent of the focal traits. This is because, when the compensatory item response model is linear and the relationship between testlet-specific and

focal traits is linear, any correlation between the testlet-specific and focal traits can be removed by projecting the testlet-specific factor to the vector space spanned by the focal factor(s) and, subsequently, treating the residual of the projection as the testlet-specific effect. However, for DCMs, where the focal traits are discrete and the measurement model involves higher-order interactions of latent attributes, such projection cannot be performed. Under testlet-effect DCMs, one thus faces a dilemma where the orthogonality between testlet-specific and focal effects is key to identifiability but cannot be naturally granted. To this end, we propose an extension to the testlet-effect DINA model, to account for the additional local dependency by adding an interaction term involving testlet-specific and focal effects.

The second issue addressed by the current work is testlet-effect DCM identifiability. Model identifiability is fundamental to consistent parameter estimation and valid statistical inference, and thus it is not only crucial to latent variable modeling but also has direct implications in test designs. It is well known that, without suitable assumptions, neither DCMs nor bi-factor models are identifiable. Identifiability conditions for various DCMs and restricted latent class models without testlet effects have been studied extensively (see e.g. Allman et al., 2009; Chen et al., 2015; Xu & Zhang, 2016; Xu, 2017; Fang et al., 2019; Gu & Xu, 2019). For bi-factor models, which are commonly used to account for testlet effects in continuous latent trait models, Fang et al. (2021) provided relatively complete characterizations of their identifiability conditions. On the other hand, establishing the identifiability conditions for testlet-effect DCMs is much more challenging as it involves both continuous testlet effects and discrete attribute profiles. To this end, the current work provides theoretical results on the identifiability conditions for testlet-effect DINA and DINO models under the probit link.

It is worth noting that, aside from testlet-effect DCMs, there exist other approaches for handling the local dependency induced by testlets in diagnostic assessments. This includes treating the items in a testlet as one single polytomous item (Ma & de la Torre, 2016) , graphical DCM that additionally models item pairwise dependence (Kang et al., 2017) , and nonparametric diagnostic classification that employs a testlet-effect-weighted distance measure (Sha, 2016) . In addition, under testlet-effect DCMs, more general DCMs than the DINA and the DINO have been adopted (e.g., Hansen, 2013; Ma et al., 2023). However, the current paper will focus on accounting for the additional dependency and identifiability theory of the testlet-effect DINA and DINO models. While the DINA and DINO models are restrictive on the permissible attribute interactions, we aim to provide a thorough analysis for this simple case, in the hope that these initial results contribute to further discussions on more general models.

The rest of the paper is organized as follows. Section 1 introduces the extended testlet DINA model and discusses its connection to the independent testlet DINA model. Section 2 presents the main theoretical results for model identifiability. Applications to the 2015 Program for International Student Assessment (PISA) collaborative problem solving (CPS) assessment data are presented in Sect. 3. Section 4 is devoted to simulation studies. A discussion of the main results, limitations, and future work is provided in Sect. 5.

1. Testlet DINA with Dependency

In this section, we introduce our extension to the testlet DINA model, the interaction testlet DINA (IT-DINA) model. The IT-DINA model incorporates the potential dependency between testlet effects and latent attribute profiles via an interaction term. To explain the rationale behind incorporating the additional latent interaction, we begin with introducing the testlet-effect DINA measurement model as seen in Hansen (2013), Hansen et al. (2016), Zhan et al. (2015), Zhan et al. (2018), and Ma et al. (2023).

Consider a diagnostic test with J items designed to measure K attributes. The examinee's attribute profile is represented by a K-dimensional binary vector $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_K)$, where $\alpha_k = 1$ or 0, indicating the mastery or nonmastery of the kth attribute. Under the DINA model (Tatsuoka, 1983; Junker & Sijtsma, 2001) , the relationship between item responses and attributes is specified by an incidence matrix $Q = (q_{jk})_{J\times K}$, where $q_{jk} = 1$ if the kth attribute is required for answering the jth item correctly, and $q_{jk} = 0$ otherwise. The ideal response to item j, $\eta_j = \mathcal{I}(\boldsymbol{\alpha} \succeq \mathbf{q}_j)$, is 1 when the examinee masters all attributes required by item j and is 0 otherwise. Under the DINA model, the probability of correct response can be parameterized as a generalized linear model,

$$P(Y_j = 1 \mid \boldsymbol{\alpha}) = h(\delta_{0j} + \delta_{1j}\eta_j), \tag{1}$$

where $h(\cdot)$ is a link function, such as the probit or the logit link.

In the presence of testlets, suppose that the J-item test comprises of S testlets. Let J_s denote the number of items in the sth testlet, with $\sum_{s=1}^{S} J_s = J$, and $\mathcal{I}_s = \{j \mid \text{ item } j \text{ belongs to testlet } s\}$ denote the set of items in testlet s. Within a testlet, dependency among items often cannot be fully explained by the measured attributes ($\boldsymbol{\alpha}$). As in testlet-effect IRT models (e.g. Bradlow et al., 1999), testlet-effect DCMs assume that items within a testlet additionally measure the same specific factor known as the testlet effect. For an item j belonging to testlet s,

$$P(Y_j = 1 \mid \boldsymbol{\alpha}, \zeta_s) = h(\delta_{0j} + \delta_{1j}\eta_j + \delta_{sj}\zeta_s). \tag{2}$$

Here, ζ_s is the examinee's testlet-specific trait for testlet s. Inside the linear component, the item parameters, δ_{0j}, δ_{1j}, and δ_{sj}, represent the intercept, the slope for mastering all requisite attributes, and the slope for the testlet-specific trait, respectively.

At the population level, similar to Ma et al. (2023), we assume that a randomly chosen examinee's attribute profile follows a multinomial distribution with proportion parameter vector $\boldsymbol{\pi} = (\pi_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in\{0,1\}^K}$, where $\sum_{\boldsymbol{\alpha}\in\{0,1\}^K} \pi_{\boldsymbol{\alpha}} = 1$. In general, ζ_s, which accounts for the additional dependence in correct response probability for all $j \in \mathcal{I}_s$ yet to be explained by the *ideal responses* (η_js), can still depend on $\boldsymbol{\alpha}$. To capture this dependency, we decompose the last term in Eq. (2), $\delta_{sj}\zeta_s$, which controls the expected change in item j's linear component due to the testlet effect, as follows:

$$\delta_{sj}\zeta_s = f_{s1j}(\boldsymbol{\alpha}) + f_{s2j}(\boldsymbol{\alpha})\xi_{s2} + \beta_{js}\xi_{s1}. \tag{3}$$

Here, $\xi_{s1}, \xi_{s2} \overset{i.i.d.}{\sim} N(0, 1)$ are two examinee-specific traits on testlet s, and $f_{s1j}(\cdot)$, $f_{s2j}(\cdot)$ are functions of $\boldsymbol{\alpha}$. Specifically, $f_{s1j}(\boldsymbol{\alpha})$ must be orthogonal to η_j and describes how $\boldsymbol{\alpha}$ further influences the mean of the testlet effect term ($\delta_{sj}\zeta_s$), and $f_{s2j}(\boldsymbol{\alpha})$ describes how $\boldsymbol{\alpha}$ influences the variance of the testlet effect term. The remaining component, $\beta_{js}\xi_{s1}$, is the unique testlet effect unrelated to $\boldsymbol{\alpha}$. With this parameterization for the distribution of the testlet effect, (2) becomes

$$P(Y_j = 1 \mid \boldsymbol{\alpha}, \zeta_s) = h(\delta_{0j} + \delta_{1j}\eta_j + f_{s1j}(\boldsymbol{\alpha}) + f_{s2j}(\boldsymbol{\alpha})\xi_{s2} + \beta_{js}\xi_{s1}). \tag{4}$$

This general formulation is not directly suitable for practical use. We discuss a few special cases.

- **T-DINA**: Conditioning on η_j, if the distribution of the testlet-specific term ($\delta_{sj}\zeta_s$) is independent of $\boldsymbol{\alpha}$, $f_{s1j}(\boldsymbol{\alpha})$ and $f_{s2j}(\boldsymbol{\alpha})$ reduce to 0, and Eq. (4) becomes

$$P(Y_j = 1 \mid \boldsymbol{\alpha}, \zeta_s) = h(\delta_{0j} + \delta_{1j}\eta_j + \beta_{js}\xi_{s1}), \tag{5}$$

  which is the T-DINA model (e.g., Zhan et al., 2015).

- **General testlet-effect DCM**: Conditioning on η_j, if the mean of the testlet effect term depends on the requisite skills of j but the variance is the same across attribute classes, then $f_{s2j}(\boldsymbol{\alpha}) = 0$. Equation (4) reduces to

$$P(Y_j = 1 \mid \boldsymbol{\alpha}, \zeta_s) = h(\delta_{0j} + \delta_{1j}\eta_j + f_{s1j}(\boldsymbol{\alpha}) + \beta_{js}\xi_{s1}), \tag{6}$$

  where $f_{s1j}(\boldsymbol{\alpha})$ can be written as the main effect and higher-order interactions of select requisite skills. This model is equivalent to the testlet-effect general DCM (Hansen, 2013; Ma et al., 2023) . In this case, the additional dependency in responses within a testlet is due to the DINA ideal response, η_j, not being able to fully describe the attribute-response relationship, and adopting a more general DCM resolves this issue.
- **IT-DINA**: Conditioning on η_j, if the mean of the testlet effect term does not further depend on $\boldsymbol{\alpha}$, but the variance differs depending on the ideal response η_j, then $f_{s1j}(\boldsymbol{\alpha}) = 0$, and $f_{s2j}(\boldsymbol{\alpha})$ can be parameterized as $\gamma_{js}(2\eta_j - 1)$. In this case, Eq. (4) reduces to

$$P(Y_j = 1 \mid \boldsymbol{\alpha}, \zeta_s) = h(\delta_{0j} + \delta_{1j}\eta_j + \beta_{js}\xi_{s1} + \gamma_{js}(2\eta_j^{\boldsymbol{\alpha}} - 1)\xi_{s2}). \tag{7}$$

  Adopting the probit link for $h(\cdot)$, we arrive at the proposed IT-DINA model,

$$P(Y_j = 1 \mid \boldsymbol{\alpha}, \boldsymbol{\xi}_s) = \Phi(\delta_{0j} + \delta_{1j}\eta_j + \beta_{js}\xi_{s1} + \gamma_{js}(2\eta_j - 1)\xi_{s2}), \tag{8}$$

  where $\Phi(\cdot)$ is the standard normal cumulative distribution function, and $\boldsymbol{\xi}_s = (\xi_{s1}, \xi_{s2})'$.

Equation 8 gives our proposed IT-DINA model, which will be the focal scenario considered in the current paper. Let $\boldsymbol{\delta}_j = (\delta_{0j}, \delta_{1j})$, $\boldsymbol{\delta} = (\boldsymbol{\delta}_j)_{1\le j\le J}$, $\boldsymbol{\beta}_s = (\beta_{js})_{j\in\mathcal{I}_s}$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_s)_{1\le s\le S}$, $\boldsymbol{\gamma}_s = (\gamma_{js})_{j\in\mathcal{I}_s}$, and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_s)_{1\le s\le S}$. The joint marginal distribution of $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_J)$ can be expressed as

$$P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi}) = \sum_{\boldsymbol{\alpha}} \pi_{\boldsymbol{\alpha}} \prod_{s=1}^{S} \int \prod_{j\in\mathcal{I}_s} P(Y_j = y_j \mid \boldsymbol{\alpha}, \boldsymbol{\xi}_s, \boldsymbol{\delta}_j, \beta_{js}, \gamma_{js}) p_{\boldsymbol{\xi}}(\boldsymbol{\xi}_s) d\boldsymbol{\xi}_s, \tag{9}$$

where $p_{\boldsymbol{\xi}}(\boldsymbol{\xi}_s)$ is the standard bivariate normal density.

Below, we present a few remarks on the IT-DINA model's interpretation, its relationship to existing models, and an alternative parameterization based on model-implied first and second moments given ideal responses.

Remark 1. (Differential testlet effect for masterers and non-masterers). The IT-DINA model simplifies the attribute-response relationship to involve only the ideal response, η_j, but affords nice interpretations. Without explicitly modeling the structural relationship between item j's ideal response and the testlet effect ζ_s, the interaction term $(2\eta_j - 1)\xi_{s2}$ controls how the testlet-specific effect that enters the measurement model will differ, depending on whether the examinee has mastered all of item j's requisite skills. We coin this the "differential testlet effect": When an examinee has all requisite skills, $\eta_j = 1$, $(2\eta_j - 1)\xi_{s2} = \xi_{s2}$, and the combined testlet effect becomes $\beta_{js}\xi_{s1} + \gamma_{js}\xi_{s2}$; when an examinee misses any requisite skill for item j, $\eta_j = 0$, $(2\eta_j - 1)\xi_{s2} = -\xi_{s_2}$, and the combined testlet effect becomes $\beta_{js}\xi_{s1} - \gamma_{js}\xi_{s2}$. In other words, the testlet-specific trait that enters the measurement model will differ by $2\gamma_{js}\xi_{s2}$, depending on the mastery status of the required skills. With the variance of both ξ_{s1}, ξ_{s2} fixed to 1, the coefficient γ_{js} will capture the extent to which the combined testlet effect on item j is dependent on

the possession of requisite skills. As a hypothetical example, consider a testlet on an elementary probability assessment, where the items involve the permutations and combinations based on a deck of cards. It is natural to imagine cases where one unit of increase in card game familiarity (the testlet effect) will influence the odds of correct response on an item differently, depending on whether the examinee has the relevant probability knowledge for solving the item.

Remark 2. (Relationship with T-DINA and DINA) When item parameters $\gamma_{js} = 0$ for all j, the measurement model in Eq. (8) reduces to the T-DINA model (e.g., Hansen, 2013; Hansen et al., 2016; Zhan et al., 2015). It further reduces to the DINA model (reparameterized in probit form), when $\beta_{js} = 0$ for all j. Hypothesis testing on the coefficients for the testlet main effect (βs) and interaction (γs) offers a natural way for model comparison, via empirically testing the presence of testlet effect and testlet-attribute dependence on an item.

Remark 3. (Differences with the multidimensional T-DINA model) Zhan et al. (2015, 2018) proposed a multidimensional T-DINA model, where each item can load on more than one testlet-specific dimension. While the proposed IT-DINA model also involves each item loading on two testlet-specific dimensions, the two models are inherently different. The multidimensional T-DINA model essentially allows testlet-specific dimensions to go beyond a simple loading structure. The IT-DINA model, other the other hand, still assumes that each item in testlet s only loads on the two-dimensional testlet effects specific to testlet s. The two dimensions here are introduced for the purpose of capturing the differential testlet effects associated with different ideal responses.

Remark 4. (η-**implied first moment: DINA slipping and guessing**) Under the DINA model, the uncertainty at the item level is captured by the slipping parameter $\boldsymbol{s}$ and guessing parameter $\boldsymbol{g}$: for item j, $g_j := P(Y_j = 1 \mid \eta_j^{\boldsymbol{\alpha}} = 0)$ and $c_j := 1 - s_j = P(Y_j = 1 \mid \eta_j^{\boldsymbol{\alpha}} = 1)$. The new IT-DINA model can be reparameterized to obtain $\boldsymbol{c}$ and $\boldsymbol{g}$. For item j in testlet s, given the testlet effect $\boldsymbol{\xi}_s$, we have

$$
\begin{aligned}
P(Y_j = 1 \mid \eta_j^{\boldsymbol{\alpha}} = 1, \boldsymbol{\xi}_s, \boldsymbol{\delta}_j, \beta_{js}, \gamma_{js}) &= \Phi(\delta_{0j} + \delta_{1j} + \beta_{js}\xi_{s1} + \gamma_{js}\xi_{s2}), \\
P(Y_j = 1 \mid \eta_j^{\boldsymbol{\alpha}} = 0, \boldsymbol{\xi}_s, \boldsymbol{\delta}_j, \beta_{js}, \gamma_{js}) &= \Phi(\delta_{0j} + \beta_{js}\xi_{s1} - \gamma_{js}\xi_{s2}).
\end{aligned}
$$

With the probit link, by integrating out $\boldsymbol{\xi}_s$, we get

$$
c_{js} = \mathbb{E}_{\boldsymbol{\xi}_s}\left[P(Y_j = 1 \mid \eta_j^{\boldsymbol{\alpha}} = 1, \boldsymbol{\xi}_s, \boldsymbol{\delta}_j, \beta_{js}, \gamma_{js})\right] = \Phi\left(\frac{\delta_{0j} + \delta_{1j}}{\sqrt{1 + \beta_{js}^2 + \gamma_{js}^2}}\right), \tag{10}
$$

$$
g_{js} = \mathbb{E}_{\boldsymbol{\xi}_s}\left[P(Y_j = 1 \mid \eta_j^{\boldsymbol{\alpha}} = 0, \boldsymbol{\xi}_s, \boldsymbol{\delta}_j, \beta_{js}, \gamma_{js})\right] = \Phi\left(\frac{\delta_{0j}}{\sqrt{1 + \beta_{js}^2 + \gamma_{js}^2}}\right). \tag{11}
$$

For the case of $\gamma_{js} = 0, c_{js} = \Phi\left((\delta_{0j} + \delta_{1j})/\sqrt{1 + \beta_{js}^2}\right)$ and $g_{js} = \Phi\left(\delta_{0j}/\sqrt{1 + \beta_{js}^2}\right)$, which, when $\beta_{js} = 0$, further reduce to $c_j = \Phi(\delta_{0j} + \delta_{1j})$ and $g_j = \Phi(\delta_{0j})$.

Remark 5. ($\eta-$implied second moments) Following Fang et al. (2021), we can get the joint marginal probability of correct responses for a pair of items j_1 and j_2 in testlet s,

$$
P(Y_{j_1} = 1, Y_{j_2} = 1 \mid \boldsymbol{\delta}_s, \boldsymbol{\beta}_s, \boldsymbol{\gamma}_s, \boldsymbol{\pi})
$$

$$= \sum_{\boldsymbol{\alpha}} \pi_{\boldsymbol{\alpha}} \Phi_2 \left(\Phi^{-1} \left(c_{j_1}^{\eta_{j_1}^{\boldsymbol{\alpha}}} g_{j_1}^{\left(1-\eta_{j_1}^{\boldsymbol{\alpha}}\right)} \right), \Phi^{-1} \left(c_{j_2}^{\eta_{j_2}^{\boldsymbol{\alpha}}} g_{j_2}^{\left(1-\eta_{j_2}^{\boldsymbol{\alpha}}\right)} \right); \rho_{j_1 j_2 s}^{\boldsymbol{\alpha}} \right), \tag{12}$$

where

$$\rho_{j_1 j_2 s}^{\boldsymbol{\alpha}} = \frac{\beta_{j_1 s}\beta_{j_2 s} + \gamma_{j_1 s}\gamma_{j_2 s}(2\eta_{j_1}^{\boldsymbol{\alpha}} - 1)(2\eta_{j_2}^{\boldsymbol{\alpha}} - 1)}{\sqrt{(1 + \beta_{j_1 s}^2 + \gamma_{j_1 s}^2)(1 + \beta_{j_2 s}^2 + \gamma_{j_2 s}^2)}},$$

$\boldsymbol{\delta}_s = \boldsymbol{\delta}[\mathcal{I}_s, :]$ denotes the submatrix of matrix $\boldsymbol{\delta}$ consisting only of rows in the set $\mathcal{I}_s$, and $\Phi_2(a, b; \rho) := P(X_1 \leq a, X_2 \leq b)$ where $X_1, X_2 \sim N(0, 1)$, and $\text{corr}(X_1, X_2) = \rho$. Typically, a, b, and ρ are referred to as the thresholds and the tetrachoric correlation.

Because η_j is either 0 or 1, $\rho_{j_1 j_2 s}^{\boldsymbol{\alpha}}$ can only take two values,

$$\rho_{j_1 j_2}^1 = \frac{\beta_{j_1 s}\beta_{j_2 s} + \gamma_{j_1 s}\gamma_{j_2 s}}{\sqrt{(1 + \beta_{j_1 s}^2 + \gamma_{j_1 s}^2)(1 + \beta_{j_2 s}^2 + \gamma_{j_2 s}^2)}}, \text{ and } \rho_{j_1 j_2}^2 = \frac{\beta_{j_1 s}\beta_{j_2 s} - \gamma_{j_1 s}\gamma_{j_2 s}}{\sqrt{(1 + \beta_{j_1 s}^2 + \gamma_{j_1 s}^2)(1 + \beta_{j_2 s}^2 + \gamma_{j_2 s}^2)}}.$$

For a more concise presentation, we let $\boldsymbol{\rho}_s^1 = (\rho_{j_1 j_2 s}^1)_{j_1 < j_2, j_1, j_2 \in \mathcal{I}_s}$ and $\boldsymbol{\rho}_s^2 = (\rho_{j_1 j_2 s}^2)_{j_1 < j_2, j_1, j_2 \in \mathcal{I}_s}$, and further let $\boldsymbol{\rho} = (\boldsymbol{\rho}_s^1, \boldsymbol{\rho}_s^2)_{s=1,\ldots,S}$ denote the collection of all the tetrachoric correlation parameter matrices.

2. Identifiability

In this section, we develop the theoretical results for the identifiability of the proposed IT-DINA model. Let $\mathcal{P} := \{(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi}) \mid \delta_{1j} > 0, \pi_{\boldsymbol{\alpha}} > 0, \sum_{\boldsymbol{\alpha}} \pi_{\boldsymbol{\alpha}} = 1\}$ denote the parameter space. We also introduce its reparameterized version $\mathcal{P}^{(r)} = \{(\boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\rho}, \boldsymbol{\pi}) \mid c_{js} > g_{js}, \pi_{\boldsymbol{\alpha}} > 0, \sum_{\boldsymbol{\alpha}} \pi_{\boldsymbol{\alpha}} = 1\}$ ($\delta_{1j} > 0$ implies $c_{js} > g_{js}$).

2.1. Definition of Identifiability

We introduce two definitions of identifiability, *strict identifiability* (Definition 1) and *$(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma})$-partial identifiability* (Definition 2).

Definition 1. *(strict identifiability)* We say the IT-DINA is identifiable at $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi})$ if the following holds: If there exists another set of parameters $(\bar{\boldsymbol{\delta}}, \bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\gamma}}, \bar{\boldsymbol{\pi}}) \in \mathcal{P}$ that defines the same probability distribution in Eq. (9), it must hold that $(|\boldsymbol{\delta}|, |\boldsymbol{\beta}|, |\boldsymbol{\gamma}|, |\boldsymbol{\pi}|) = (|\bar{\boldsymbol{\delta}}|, |\bar{\boldsymbol{\beta}}|, |\bar{\boldsymbol{\gamma}}|, |\bar{\boldsymbol{\pi}}|)$.

Definition 2. *($(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma})$-partial identifiability)* For the IT-DINA, the model parameters $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi})$ are said to be $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma})$-partially identifiable if $(\boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\rho}, \boldsymbol{\pi})$ are identifiable.

Definition 2 is weaker than Definition 1, i.e., the strict identifiability of $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi})$ implies $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma})$-partial identifiability. Nevertheless, $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma})$-partial identifiability can uniquely determine parameters $(\boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\rho})$ but can fail to guarantee the strict identifiability. Examples in the appendix provide an illustration of models whose $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ is non-identifiable, while $(\boldsymbol{c}, \boldsymbol{g}, \boldsymbol{\rho})$ can be identified.

2.2. Main Results

For reader convenience, we now first summarize our main theoretical results.

- Theorem 1 provides checkable sufficient conditions for the $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma})$-partial identifiability.
- Theorem 2 gives the conditions which guarantee strict identifiability.
- Proposition 1, essentially a byproduct of Theorem 2, provides sufficient conditions for the identifiability of the independent testlet DINA model.

We require the following two conditions.

C1 After row permutations, there exist S_0 testlets whose Q_s-matrix can be partitioned into a $K \times K$ identity matrix I_K and a matrix Q', where Q' is a $(J_s - K) \times K$ matrix, for $s = 1, 2, \ldots, S_0$. In other words, the Q_s submatrix is complete (Chiu et al., 2009) in S_0 testlets.

C2 The number S_0 in C1 is no smaller than $\max\{K, 3\}$.

Theorem 1. *Let Q_s denote the $J_s \times K$ submatrix of Q that corresponds to items in testlet s. Under the model specified in Eq. (8), Conditions C1 and C2 are sufficient for $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma})$-partial identifiability of $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi})$.*

In addition to the first two conditions, we further impose the following requirement.

C3 For each testlet $s = 1, 2, \cdots, S$, there exist at least three items with nonzero β parameters, and at least three items with nonzero γ parameters. The two sets of three items are allowed to differ.

By meeting all three conditions above, the following Theorem 2 provides sufficient conditions for strict identifiability.

Theorem 2. *Under the model specified in Eq. (8), if conditions C1, C2 and C3 are satisfied, then the parameters $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi})$ are identifiable.*

The T-DINA model is a special case of the proposed model, where $\gamma_{js} \equiv 0$ for all s and all j. By modification of Condition C3, we can achieve the strict identifiability of the T-DINA model. This result provides the first-ever sufficient conditions for the identifiability of T-DINA in the literature.

C4 For testlet s, $s = 1, 2, \cdots, S$, there exist at least three items such that their β parameters are nonzero.

Proposition 1. *For the standard T-DINA model specified in Eq. (8) with $\gamma_{js} \equiv 0$, Conditions C1, C2, and C4 are sufficient for the identifiability of $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi})$.*

Remark 6. (Extension to DINO model) The result presented in Theorems 1 can be directly applied to the DINO-based model by the mathematical duality between the DINA and DINO models (Proposition 1 in Chen et al., 2015; Köhn & Chiu, 2016). Additionally, Theorem 2 and Proposition 1 can also be extended directly to DINO-based models as the $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma})$-partial identifiability of $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi})$.

2.3. Technical Discussions

There are some technical issues that need further clarification. In the remarks below, we provide discussions on some of the issues.

Remark 7. (Identifiability up to sign changes) By the definition of parameters $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi})$, the identifiability holds up to sign changes to the slope coefficients within a testlet, particularly, $\boldsymbol{\beta}_s, \boldsymbol{\gamma}_s$. For example, we consider β_{js} identifiable if $\beta_{js}/\bar{\beta}_{js} = \pm 1$, for item j.

Remark 8. (DINA and T-DINA identifiability) Theorem 1 implies the strict identifiability of the DINA model. The Q-matrix in Theorem 1 satisfies the conditions in Gu and Xu (2020); Gu (2020); Gu and Xu (2019, 2022) for identifying the slipping, guessing, and latent membership parameters. However, the proposed conditions are stricter because Theorem 1 aims at identifying a more general IT-DINA model. Further, for a T-DINA model, a complete characterization of identifiability is not yet available to our knowledge. Theorem 1 can provide the partial identifiability of a T-DINA model and Proposition 1 implies its strict identifiability. However, under a T-DINA model, because of the additional restriction $\rho^1_{j_1 j_2 s} = \rho^2_{j_1 j_2 s} = \beta_{j_1 s}\beta_{j_2 s}/\sqrt{(1+\beta^2_{j_1 s})(1+\beta^2_{j_2 s})}$, there might exist more relaxed identifiability conditions.

Remark 9. Condition C1 states that the Q-matrix, after swapping of testlets, can be partitioned into matrices $Q_1, \ldots, Q_{S_0}$ and Q^* where Q^* is a $(J - \sum_{s=1}^{S_0} J_s) \times K$ matrix. And for $s = 1, 2, \cdots, S_0$, Q_s contains an identity matrix I_K.

Remark 10. For a given testlet s, Condition C3 states, after swapping indices of items, β and γ parameters can be expressed as follows,

$$(\beta_{j_1 s}, \beta_{j_2 s}, \beta_{j_3 s}, \beta_{j_4 s}, \ldots), (\gamma_{l_1 s}, \gamma_{l_2 s}, \gamma_{l_3 s}, \gamma_{l_4 s}, \ldots)$$

where $\beta_{j_1 s}, \beta_{j_2 s}, \beta_{j_3 s}$ and $\gamma_{l_1 s}, \gamma_{l_2 s}, \gamma_{l_3 s}$ are all nonzero. Note that $\{j_1, j_2, j_3\}$ and $\{l_1, l_2, l_3\}$ are subsets of $\mathcal{I}_s$, which can either intersect or be disjoint.

Remark 11. Since J_s can be as small as one, the result here covers the situation in which some of the testlets may contain a single item.

Remark 12. The probit model assumption here is indeed crucial in developing our theory. It allows us to compute the tetrachoric correlation so that the explicit formula of the joint probability of item responses, i.e. (12), can be obtained. Whether other types of link functions can provide such mathematical convenience is worth future study.

Remark 13. Theorems 1–2 guarantee the identifiability of model parameters when the Q-matrix remains fixed. Since Q-matrix is usually unknown in real applications, a common practice is to estimate the Q-matrix in advance from the data. Therefore, it is important and challenging to identify the conditions for the identifiability of Q-matrix. One possible solution is to follow the techniques in Gu and Xu (2021) where they developed necessary and sufficient conditions of identifiability of Q-matrix for DINA model. We leave this to future work.

2.4. Examples

To help readers to gain intuitions of the identifiability conditions, we end this section with two illustrative examples.

Example 1. Consider the IT-DINA model with $K = 2$ attributes and $S = 3$ testlets, with the Q-matrix given by Table 1, and $\mathcal{I}_1 = \{1, 2, 3\}$, $\mathcal{I}_2 = \{4, 5, 6\}$, and $\mathcal{I}_3 = \{7, 8, 9\}$. Assume β_{js} and γ_{js} are nonzero, for $j = 1, \ldots, 9$, $s = 1, 2, 3$.

For each testlet, there exists an identity matrix in the corresponding Q_s-matrix. Additionally, there are three items in each testlet. Thus Conditions C1, C2 and C3 are satisfied. By Theorem 2, we know that the model is identifiable.

TABLE 1.
Q-matrix in Example 1.

Item	α_1	α_2	Item	α_1	α_2	Item	α_1	α_2
1	1	0	4	1	0	7	1	0
2	0	1	5	0	1	8	0	1
3	0	1	6	1	1	9	1	0

TABLE 2.
Q-matrix in Example 2.

Item	α_1	α_2	α_3	α_4	Item	α_1	α_2	α_3	α_4
1	1	0	0	0	13	0	1	0	0
2	1	1	0	0	14	0	0	1	0
3	0	0	1	0	15	0	1	1	1
4	0	1	0	0	16	0	0	0	1
5	0	0	0	1	17	1	0	0	0
6	1	0	1	1	18	0	1	0	0
7	1	0	0	0	19	1	1	1	1
8	0	1	0	0	20	0	0	1	0
9	1	0	1	1	21	0	0	0	1
10	0	0	1	0	22	1	0	1	0
11	0	0	0	1	23	0	1	0	1
12	1	0	0	0	24	1	1	1	1

Example 2. Next we consider a model with $K = 4$ and $S = 4$, with the Q-matrix given by Table 2, and $\mathcal{I}_1 = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{I}_2 = \{7, 8, 9, 10, 11\}$, $\mathcal{I}_3 = \{12, 13, 14, 15, 16\}$, and $\mathcal{I}_4 = \{17, 18, 19, 20, 21\}$. Note that items 22–24 are single items. Assume β_{js} and γ_{js} are nonzero, for $j = 1, \ldots, 21$, $s = 1, \ldots, 4$.

It can be seen that for each testlet, there exists an identity matrix in Q_s. There also exist at least three items in each testlet, and $S_0 \geq \min\{K, 3\}$. By Theorem 2, we know that this model is also identifiable.

3. Applications

In this section, we apply the proposed IT-DINA model (model (8)), the T-DINA model (model (5)), and the DINA model (model (1)) to a dataset from the 2015 PISA. We conduct hypothesis testing for the existence of possible dependency between testlet effects. We also provide parameter estimation for the attribute profiles.

3.1. Data

We fitted the models to the data from the 2015 computer-based PISA collaborative problem-solving (CPS) test. Four of five testlets in main survey cluster 1 were selected. Within the framework of the 2015 PISA survey, CPS skills were represented by a cross matrix of three collaboration competencies and four core problem-solving competencies (OECD, 2016) . Yavuz and Atar

TABLE 3.
The Q-matrix in PISA data and the testlet structure. $*$ indicates which testlet an item belongs to.

Item	α_1	α_2	α_3	s_1	s_2	s_3	s_4	Item	α_1	α_2	α_3	s_1	s_2	s_3	s_4
CC104101	0	1	0	*				CC104202	1	0	0			*	
CC104102	1	0	0	*				CC104203	1	0	0			*	
CC104103	0	1	0	*				CC104204	0	0	1			*	
CC104105	1	0	0	*				CC104205	0	1	0			*	
CC104106	1	0	0	*				CC104206	1	0	0			*	
CC104107	1	0	0	*				CC106201	1	0	0				*
CC106101	1	0	0		*			CC106202	0	0	1				*
CC106102	1	0	0		*			CC106203	0	0	1				*
CC106103	1	0	0		*			CC106204	0	1	0				*
CC106104	0	0	1		*			CC106205	0	1	0				*
CC106105	0	1	0		*			CC106206	0	0	1				*
CC106106	1	0	0		*			CC106207	0	0	1				*
CC106107C	0	0	1		*			CC106208	0	0	1				*
CC104201	0	0	1			*		CC106209	0	1	0				*

(2020) presented three core competencies that could provide the best model fit. Therefore, we take the assignment of the three competencies as attributes in this paper. $K = 3$ attributes (i.e., CPS competencies) were measured, specifically, α_1: establishing and maintaining shared understanding; α_2: taking appropriate action to solve the problem; α_3: establishing and maintaining team organization. This resulted in a total of $J = 28$ items grouped in $S = 4$ testlets. The corresponding Q-matrix and structure of testlets are given in Table 3. It could be seen that testlets 2, 3, and 4 contained complete Q_s matrices. This empirical example hence met the identifiability conditions for Theorem 1. To evaluate the appropriateness of retrofitting a DCM to this data set, following the approach of de La Torre and Karelitz (2009), we have calculated item diagnosticity and found that almost all the items exhibited typical/high diagnosticity. Therefore, we fitted DCMs on PISA CPS data in an attempt to extract diagnostic information.

For the current study, test takers were chosen from all OECD countries and regions where the English version of the exam was administered. For simplicity, all polytomous responses were recoded to dichotomous scores based on whether the highest score was received. Examinees with missingness were removed, resulting in a total of $N = 8880$ examinees.

3.2. Methods and Results

To assess the goodness of fit, the likelihood ratio test (LRT), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC) were used. For the LRT, we use the usual χ^2 approximation for the log-likelihood ratio statistic. The AIC and BIC are given by

$$\begin{aligned} \text{AIC} &= -2\log(\hat{L}) + 2p, \\ \text{BIC} &= -2\log(\hat{L}) + p\log(N), \end{aligned}$$

where p denotes the number of model parameters, and N the sample size.

Table 4 presents the results of the LRT. It shows that, for DINA versus T-DINA, the DINA should be rejected in favor of T-DINA as the corresponding p-value was much less than 0.05. Furthermore, for T-DINA versus the proposed IT-DINA, T-DINA should be rejected. Using AIC and BIC, we also conclude that IT-DINA provided the best fit as shown by Table 5.

TABLE 4.
Results on the LRT on the PISA data. H_0 is the null hypothesis, H_1 is the alternative hypothesis, χ^2 denotes the values of the LRT statistic and df denotes the degrees of freedom of the LRT.

H_0	H_1	χ^2	df	p-value	Result
DINA	T-DINA	2117.491	28	$< 1 \times 10^{-16}$	Reject H_0
T-DINA	IT-DINA	644.933	28	$< 1 \times 10^{-16}$	Reject H_0

TABLE 5.
Results on goodness of fit for PISA data.

Model	p	$-2\log(\hat{L})$	AIC	BIC
IT-DINA	120	**271574.276**	**271814.276**	**272665.263**
T-DINA	92	272219.209	272403.209	273055.632
DINA	64	274336.7	274464.699	274918.559

Tables 6 and 7 provide the parameter estimates under the IT-DINA and T-DINA, respectively. The interaction on some items was much higher, such as $\gamma_{25,4} = 0.768$, while the corresponding $\beta_{25,4} = -0.078$. However, under the T-DINA model, $\beta_{25,4} = 0.534$ which was higher than that under the proposed model. For classes $\eta^{\boldsymbol{\alpha}}_{25} = 1$ and $\eta^{\boldsymbol{\alpha}}_{25} = 0$, the impact of testlet for item 25 is mainly reflected through $\xi_{4,2}$ not $\xi_{4,1}$.

Tables 8 and 9 display the frequencies of latent classes across various models estimated using expected a posteriori (EAP). Both the proposed IT-DINA and T-DINA models modified the classification of examinees when compared to DINA model. For instance, 65 and 84 students with class (0, 0, 0) in the DINA model were instead classified into class (1, 1, 1). One explanation is that the DINA model could potentially classify examinees who possess all the required attributes but are low on the context-based stimuli as non-masterers. However, the IT-DINA and the T-DINA model incorporated additional testlet effects, resulting in adjustments for these students. A similar trend was observed in other classes.

Furthermore, upon comparing Tables 8 and 9, differences were found in the attribute classification under the IT-DINA and the T-DINA. This discrepancy arose because the T-DINA model and the IT-DINA model imply different levels of the testlet effects for the participants, which could in turn shift the attribute classifications.

4. Simulation Study

This section provides a simulation study that evaluates parameter recovery under the proposed testlet DINA model and the consequences of ignoring the testlet and the interaction effects.

4.1. Design

The Q-matrix and testlet structures are specified by Tables 3 and 10 for $K = 3$ and 5, respectively. In both cases, the identifiability conditions in Theorem 1 are met. Sample sizes of $N = 500$, 1000, 2000 and 4000 were considered. The data were generated according to the following specifications. Five parameter settings were considered. Cases 1–3 considered different

TABLE 6.
Parameter estimates under the proposed IT-DINA model for PISA data. For πs, the order is $\boldsymbol{\pi}_{(0,0,0)}$, $\boldsymbol{\pi}_{(1,0,0)}$, $\boldsymbol{\pi}_{(0,1,0)}$, $\boldsymbol{\pi}_{(0,0,1)}$, $\boldsymbol{\pi}_{(1,1,0)}$, $\boldsymbol{\pi}_{(1,0,1)}$, $\boldsymbol{\pi}_{(0,1,1)}$, $\boldsymbol{\pi}_{(1,1,1)}$.

Item	δ_0	δ_1	β_1	β_2	β_2	β_4	γ_1	γ_2	γ_3	γ_4	π
1	0.382	0.890	0.121				0.317				0.339
2	−0.304	1.714	0.629				0.252				0.009
3	−0.501	1.602	0.538				0.431				0.000
4	−1.330	1.413	0.604				0.001				0.001
5	0.760	0.655	−0.150				0.439				0.008
6	0.432	0.633	−0.080				0.463				0.002
7	−0.767	1.810		−0.604				0.730			0.005
8	−0.659	0.403		0.292				0.007			0.636
9	−0.593	1.560		−0.220				0.459			
10	−0.058	0.946		−0.078				0.101			
11	−0.240	1.152		0.091				0.143			
12	−0.672	0.377		0.378				0.296			
13	−1.927	1.284		0.376				0.566			
14	−0.570	1.463			0.930				0.323		
15	−0.465	0.733			0.022				0.162		
16	−0.104	1.944			0.266				0.855		
17	−0.231	1.318			0.224				0.449		
18	0.386	0.315			−0.100				0.195		
19	0.751	0.887			0.035				0.446		
20	−0.312	0.804				0.174				0.189	
21	−0.982	0.334				0.165				0.043	
22	−0.393	0.471				0.155				0.298	
23	−1.044	1.331				0.583				0.292	
24	−0.594	1.598				0.327				0.236	
25	0.210	1.319				−0.078				0.768	
26	−0.785	1.467				0.464				0.092	
27	0.891	1.594				−0.139				0.548	
28	−0.676	1.421				0.099				0.524	

scenarios where the true data generating process was the IT-DINA, and cases 4 and 5 consider scenarios where the DINA and the T-DINA were the true data-generating models, respectively.

- Case 1: $\gamma_{js} = 0.5$, $\beta_{js} = 0.5$, $K = 3$, and attribute profiles were generated from a discrete uniform distribution, i.e., $\boldsymbol{\pi}$ was set to $\frac{1}{8}$ for all $\boldsymbol{\alpha} \in \{0, 1\}^3$.
- Case 2: $\gamma_{js} = 1$, $\beta_{js} = 0.5$, $K = 3$, and $\boldsymbol{\pi}$ was set to $\frac{1}{8}$ for all $\boldsymbol{\alpha} \in \{0, 1\}^3$.
- Case 3: Each testlet contains two γ_{js} and two β_{js} with negative values, and $|\gamma_{js}| = 1$, $|\beta_{js}| = 0.5$, $K = 5$, and $\boldsymbol{\pi}$ was set as an non-uniform distribution following Culpepper (2015) in which the probability of belonging to a class with three or more skills is twice as likely as belonging to a class with zero, one, or two skills. That is, $\pi_c = 0.021$ for all $\boldsymbol{\alpha}_c$ with two or fewer skills and $\pi_c = 0.042$ for $\boldsymbol{\alpha}_c$ with three or more skills.
- Case 4: $\gamma_{js} = 0$, $\beta_{js} = 0$, $K = 3$, and $\boldsymbol{\pi}$ was set to $\frac{1}{8}$ for all $\boldsymbol{\alpha} \in \{0, 1\}^3$.
- Case 5: $\gamma_{js} = 0$, $\beta_{js} = 1$, $K = 3$, and $\boldsymbol{\pi}$ was set to $\frac{1}{8}$ for all $\boldsymbol{\alpha} \in \{0, 1\}^3$.
- For more details on other parameters, please refer to Tables 1–3 in Appendix.

Each condition was replicated for $T = 500$ times. To evaluate performance of the proposed model, we first consider recovery of the original parameters $(\boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\gamma})$. To compare the perfor-

TABLE 7.
Parameter estimates under the T-DINA model for PISA data. For πs, the order is $\boldsymbol{\pi}_{(0,0,0)}$, $\boldsymbol{\pi}_{(1,0,0)}$, $\boldsymbol{\pi}_{(0,1,0)}$, $\boldsymbol{\pi}_{(0,0,1)}$, $\boldsymbol{\pi}_{(1,1,0)}$, $\boldsymbol{\pi}_{(1,0,1)}$, $\boldsymbol{\pi}_{(0,1,1)}$, $\boldsymbol{\pi}_{(1,1,1)}$.

Item	δ_0	δ_1	β_1	β_2	β_2	β_4	π
1	0.376	0.866	0.257				0.337
2	−0.308	1.709	0.685				0.014
3	−0.498	1.602	0.704				0.000
4	−1.238	1.307	0.448				0.000
5	0.689	0.599	0.086				0.006
6	0.396	0.573	0.141				0.000
7	−0.625	1.421		0.175			0.005
8	−0.602	0.337		0.121			0.638
9	−0.599	1.512		0.246			
10	−0.058	0.944		0.053			
11	−0.233	1.139		0.169			
12	−0.630	0.339		0.361			
13	−2.016	1.330		0.803			
14	−0.444	1.144			0.455		
15	−0.471	0.737			0.154		
16	−0.112	1.928			0.883		
17	−0.229	1.325			0.526		
18	0.380	0.309			0.111		
19	0.736	0.871			0.402		
20	−0.314	0.800				0.254	
21	−0.972	0.328				0.122	
22	−0.387	0.462				0.349	
23	−0.971	1.238				0.461	
24	−0.585	1.575				0.369	
25	0.209	1.146				0.534	
26	−0.729	1.370				0.275	
27	0.855	1.468				0.429	
28	−0.653	1.380				0.511	

TABLE 8.
Tabulation of attribute classification based on DINA and IT-DINA models on PISA data.

	DINA								
		(0,0,0)	(1,0,0)	(0,1,0)	(0,0,1)	(1,1,0)	(1,0,1)	(0,1,1)	(1,1,1)	Sum
IT-DINA	(0,0,0)	2866	55	6	5	1	6	8	14	2961
	(1,0,0)	12	53	0	0	2	1	0	1	69
	(0,1,0)	1	0	0	0	0	0	0	1	2
	(0,0,1)	2	0	0	3	0	0	0	1	6
	(1,1,0)	2	8	1	0	18	0	0	2	31
	(1,0,1)	2	2	0	0	0	4	0	0	8
	(0,1,1)	8	0	2	1	0	0	19	3	33
	(1,1,1)	65	62	5	5	48	61	13	5511	5770
	Sum	2958	180	14	14	69	72	40	5533	8880

TABLE 9.
Tabulation of attribute classification based on DINA and T-DINA models on PISA data.

	DINA								
		(0,0,0)	(1,0,0)	(0,1,0)	(0,0,1)	(1,1,0)	(1,0,1)	(0,1,1)	(1,1,1)	Sum
T-DINA	(0,0,0)	2847	30	7	8	2	6	13	27	2940
	(1,0,0)	15	76	0	0	1	3	0	3	98
	(0,1,0)	1	0	0	0	0	0	2	1	4
	(0,0,1)	0	0	0	0	0	0	0	0	0
	(1,1,0)	4	11	2	0	17	0	0	2	36
	(1,0,1)	0	0	0	0	0	0	0	0	0
	(0,1,1)	7	0	1	0	0	0	15	3	26
	(1,1,1)	84	63	4	6	49	63	10	5497	5776
	sum	2958	180	14	14	69	72	40	5533	8880

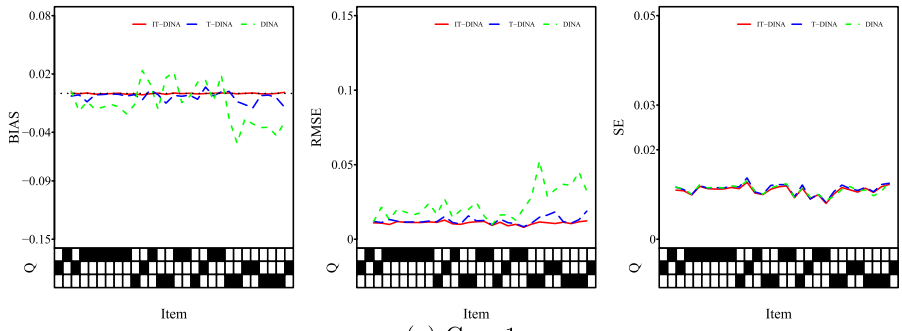
TABLE 10.
The Q-matrix for $K = 5$ in the simulation study and the testlet structure. $*$ indicates which testlet an item belongs to.

Item	α_1	α_2	α_3	α_4	α_5	s_1	s_2	s_3	s_4	Item	α_1	α_2	α_3	α_4	α_5	s_1	s_2	s_3	s_4
1	1	0	0	0	0	*				15	0	1	0	0	0			*	
2	0	1	0	0	0	*				16	0	0	1	0	0			*	
3	0	0	1	0	0	*				17	0	0	0	1	0			*	
4	0	0	0	1	0	*				18	0	0	0	0	1			*	
5	0	0	0	0	1	*				19	0	0	0	1	1			*	
6	1	1	0	0	0	*				20	1	0	0	0	0				*
7	1	0	0	0	0		*			21	0	1	0	0	0				*
8	0	1	0	0	0		*			22	0	0	1	0	0				*
9	0	0	1	0	0		*			23	0	0	0	1	0				*
10	0	0	0	1	0		*			24	0	0	0	0	1				*
11	0	0	0	0	1		*			25	1	0	1	1	0				*
12	0	0	1	1	0		*			26	0	1	1	1	0				*
13	1	0	0	1	0		*			27	1	0	0	0	1				*
14	1	0	0	0	0			*		28	0	1	0	0	1				*

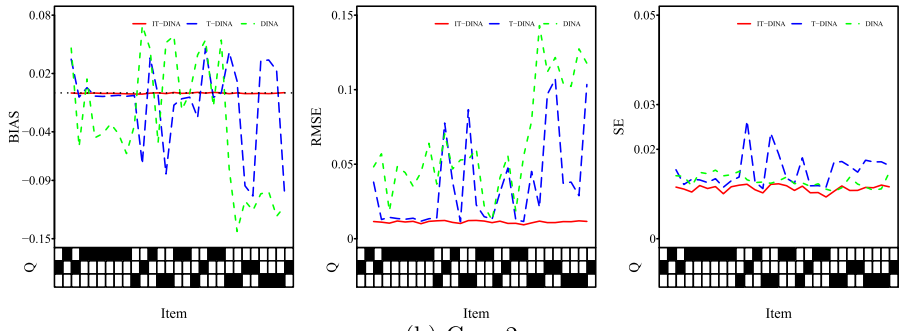
mances of DINA, T-DINA and IT-DINA, recovery of transformed parameters $(\boldsymbol{s}, \boldsymbol{g})$ was evaluated. Recovery was evaluated in terms of average bias (BIAS), root-mean-squared error (RMSE), standard error (SE) and BIAS ratio (BR) (Kish et al., 1962; Meng, 1993; Coelho & Pereira, 2011) for each parameter, across all the replications. For instance, for the slipping parameter s_j, denote its estimate from the tth replication by $\hat{s}_j^{(t)}$, then

$$\text{BIAS}(\hat{s}_j) = \frac{1}{T}\sum_{t=1}^{T}(\hat{s}_j^{(t)} - s_j),$$

$$\text{RMSE}(\hat{s}_j) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{s}_j^{(t)} - s_j)^2},$$



(a) Case 1



(b) Case 2

FIGURE 1.
BIAS, RMSE, and SE of slipping parameters under Cases 1 and 2 when $N = 4000$. X-axis is the Q-matrix, where a black square denotes the requirement of the attribute for an item.

$$\mathrm{SE}(\hat{s}_j) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{s}_j^{(t)} - \frac{1}{T}\sum_{t=1}^{T}\hat{s}_j^{(t)}\right)^2},$$

$$\mathrm{BR}(\hat{s}_j) = \frac{\mathrm{BIAS}(\hat{s}_j)}{\mathrm{SE}(\hat{s}_j)}.$$

4.2. Results

First, we analyzed the BIC-based model selection outcomes, calculating the percentage of instances the true data-generating model was selected using BIC. Across all cases, Case 1 (IT-DINA with weaker interaction effect) was the only condition where the data-generating model was not favored by BIC-based selection across all sample sizes: In Case 1, with sample sizes of $N = 500$ and 1000, the IT-DINA (true model) was selected in 11% and 98.6% of the replications, and the remaining 89% and 1.4% of instances favored the simpler T-DINA model. Under larger sample sizes of 2000 or 4000, BIC selected the IT-DINA in 100% of replications.

For Cases 2 and 3, IT-DINA (the true model) was consistently chosen in 100% of replications, irrespective of the sample sizes considered. In Case 4, the DINA model (true model) emerged as the preferred choice across 100% of replications across different sample sizes. In Case 5, the BIC consistently favored the T-DINA model (true model), with a 100% selection rate across all

TABLE 11.
The average BIAS and RMSE of item parameters under the proposed testlet DINA model.

	N	Case 1				Case 2			
		500	1000	2000	4000	500	1000	2000	4000
BIAS	δ_0	−0.0600	−0.0261	−0.0130	−0.0063	−0.0540	−0.0247	−0.0115	−0.0059
	δ_1	0.1001	0.0452	0.0209	0.0110	0.0914	0.0431	0.0197	0.0112
	β_1	0.0326	0.0212	0.0106	0.0064	0.0265	0.0263	0.0218	0.0190
	β_2	0.0250	0.0146	0.0060	0.0035	0.0133	0.0131	0.0040	0.0021
	β_3	0.0339	0.0137	0.0058	0.0029	0.0149	0.0106	0.0064	0.0041
	β_4	0.0248	0.0115	0.0044	0.0032	0.0090	0.0079	0.0052	0.0034
	γ_1	0.0297	0.0096	0.0045	0.0005	0.0623	0.0153	0.0041	−0.0009
	γ_2	0.0223	0.0077	0.0071	0.0016	0.0586	0.0184	0.0104	0.0052
	γ_3	0.0304	0.0152	0.0050	0.0041	0.0622	0.0258	0.0101	0.0052
	γ_4	0.0127	0.0123	0.0041	0.0009	0.0464	0.0243	0.0086	0.0030
RMSE	δ_0	0.2057	0.1198	0.0781	0.0527	0.2140	0.1322	0.0893	0.0601
	δ_1	0.3211	0.1847	0.1177	0.0798	0.3421	0.2120	0.1402	0.0947
	β_1	0.2774	0.1805	0.1281	0.0912	0.3461	0.2168	0.1547	0.1177
	β_2	0.2142	0.1385	0.0906	0.0623	0.2862	0.1584	0.1040	0.0721
	β_3	0.2285	0.1392	0.0939	0.0655	0.3065	0.1727	0.1087	0.0748
	β_4	0.2265	0.1396	0.0946	0.0639	0.2677	0.1637	0.1022	0.0704
	γ_1	0.3103	0.1929	0.1384	0.0986	0.2559	0.1584	0.1148	0.0827
	γ_2	0.2608	0.1488	0.0976	0.0680	0.2248	0.1366	0.0936	0.0644
	γ_3	0.2682	0.1592	0.1016	0.0710	0.2470	0.1482	0.0989	0.0676
	γ_4	0.2647	0.1579	0.1031	0.0709	0.2174	0.1404	0.0941	0.0646
SE	δ_0	0.1853	0.1118	0.0745	0.0509	0.1997	0.1263	0.0861	0.0586
	δ_1	0.2891	0.1708	0.1120	0.0768	0.3187	0.2009	0.1344	0.0920
	β_1	0.2701	0.1786	0.1273	0.0910	0.3447	0.2119	0.1469	0.1098
	β_2	0.2121	0.1373	0.0903	0.0622	0.2835	0.1577	0.1037	0.0721
	β_3	0.2257	0.1382	0.0938	0.0653	0.3046	0.1720	0.1082	0.0747
	β_4	0.2245	0.1387	0.0940	0.0637	0.2674	0.1632	0.1016	0.0702
	γ_1	0.3079	0.1910	0.1365	0.0974	0.2435	0.1523	0.1093	0.0785
	γ_2	0.2572	0.1481	0.0970	0.0679	0.2146	0.1348	0.0923	0.0639
	γ_3	0.2654	0.1580	0.1014	0.0708	0.2371	0.1449	0.0977	0.0672
	γ_4	0.2637	0.1568	0.1027	0.0707	0.2107	0.1373	0.0926	0.0644
BR	δ_0	−0.3067	−0.2213	−0.1633	−0.1134	−0.2572	−0.1879	−0.1222	−0.0938
	δ_1	0.3363	0.2512	0.1788	0.1368	0.2778	0.2071	0.1367	0.1187
	β_1	0.1085	0.1151	0.0827	0.0710	0.0759	0.1248	0.1443	0.1614
	β_2	0.1170	0.1051	0.0672	0.0565	0.0458	0.0830	0.0387	0.0295
	β_3	0.1509	0.0980	0.0625	0.0433	0.0454	0.0597	0.0596	0.0546
	β_4	0.1122	0.0825	0.0443	0.0502	0.0297	0.0544	0.0494	0.0497
	γ_1	0.0989	0.0441	0.0222	−0.0031	0.2433	0.0741	0.0032	−0.0387
	γ_2	0.0842	0.0454	0.0757	0.0217	0.2699	0.1347	0.1081	0.0778
	γ_3	0.1108	0.0940	0.0485	0.0570	0.2623	0.1736	0.1000	0.0750
	γ_4	0.0359	0.0760	0.0410	0.0120	0.2204	0.1740	0.0865	0.0453

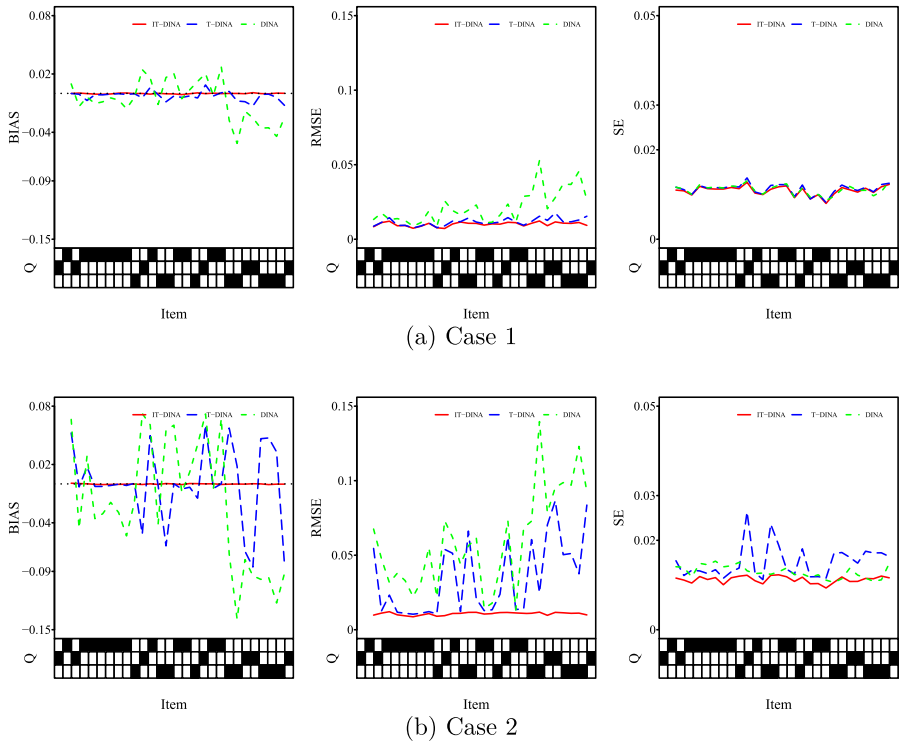


FIGURE 2.
BIAS, RMSE, and SE of guessing parameters under Cases 1 and 2 when $N = 4000$. X-axis is the Q-matrix, where a black square denotes the requirement of the attribute for an item

conditions. In sum, BIC-based model selection could usually select the correct data-generating model but tended to favor simpler models when both the true dependency (γ) and the sample size were smaller.

Under IT-DINA, the BIAS, RMSE, SE and BR of item parameters $\boldsymbol{\delta}$ were averaged across all the items, and those of item parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ were averaged across items in corresponding testlets. The results are presented in Table 11 for Cases 1 and 2. When the sample size increased, BIAS decreased to 0. Furthermore, RMSEs, SEs and BRs under the proposed model decreased as sample size increased, indicating consistency of the estimators. Similar trends for parameter recovery were also found in Cases 3–5.

Figure 1 presents the recovery of recalculated slipping parameters when $N = 4000$. Parameter estimates under both the T-DINA and DINA models showed a larger bias compared with the proposed model. Further, under DINA and T-DINA models, the BIAS of slipping parameters under Case 2 was larger than that under Case 1. Therefore, the values of γ would influence the consequences of ignoring interactions.

Figure 2 presents the recovery of recalculated guessing parameters when $N = 4000$. BIAS under the three models showed the same trend as in Fig. 1. Fitting T-DINA and DINA models to data generated under the proposed model produced biased estimates, and the BIAS was more obvious under a larger γ. Further, when comparing Figs. 1 and 2, BIAS of slipping parameters under T-DINA and DINA models was more obvious than that of guessing parameters, while the proposed model yielded unbiased estimates for both slipping and guessing.

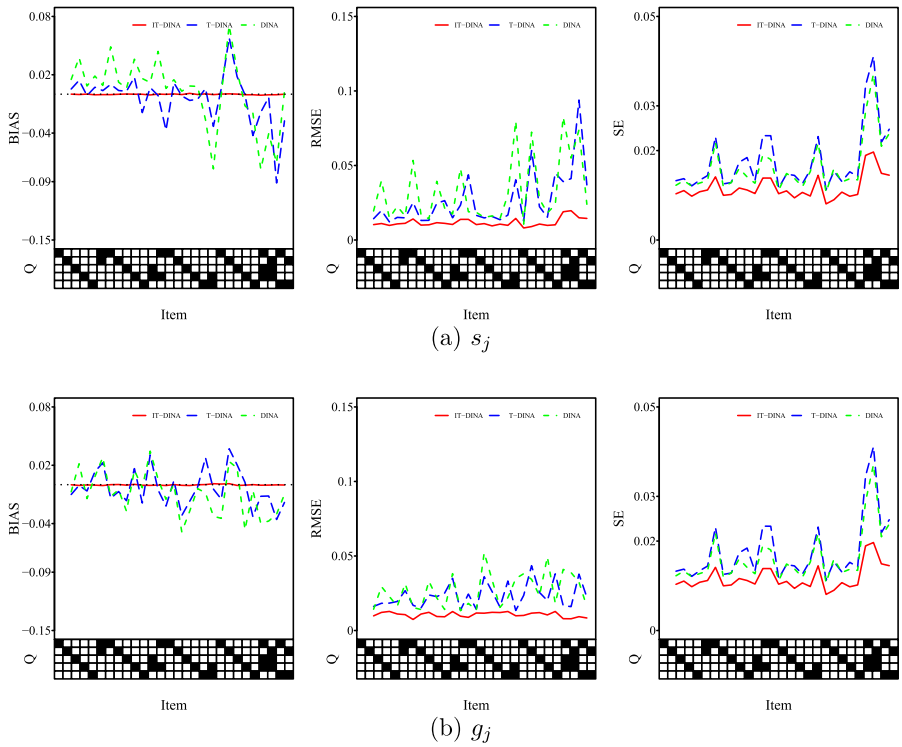


FIGURE 3.
BIAS, RMSE, and SE of slipping and guessing parameters under Case 3 when $N = 4000$. X-axis is the Q-matrix, where a black square denotes the requirement of the attribute for an item.

Figure 3 illustrates the recovery of recalculated slipping and guessing parameters under Case 3 when $N = 4000$. A similar trend is observed as in Cases 1 and 2, wherein the T-DINA and DINA models exhibit a greater bias compared to the proposed model. Additionally, Fig. 3 indicates that the variability in slipping and guessing parameters is larger for items requiring three skills compared to those requiring only one skill.

Figures 4 and 5 present the recovery of recalculated slipping and guessing parameters under Cases 4 and 5 when $N = 4000$. Under Case 4, parameter estimates under the IT-DINA, T-DINA, and DINA models were nearly identical, suggesting that the more complex IT-DINA and T-DINA models could accurately recover the parameters when the true model is the DINA. Under Case 5, when the T-DINA was the data-generating model, the recovery of parameters under the IT-DINA and T-DINA model was highly similar, both better than the recovery of parameters under the DINA model.

To take a closer look at the parameter estimates, Table 12 showed recovery of s_j and g_j of all the items. Under Cases 1–3, as the sample size increased, RMSE and SE of parameters s_j and g_j decreased for the proposed model. However, the RMSE under the T-DINA model remained large for s_j and g_j, especially under Case 2, indicating the interaction between the testlet effect and latent profiles cannot be ignored. Further, RMSEs of parameters s_j and g_j under the DINA model were larger than those under the T-DINA model, indicating the need to model the testlet effect when it is present. Overfitting a more complex model (e.g., the IT-DINA) to data generated under simpler models did not compromise parameter recovery. Specifically, under Case 4 (DINA),

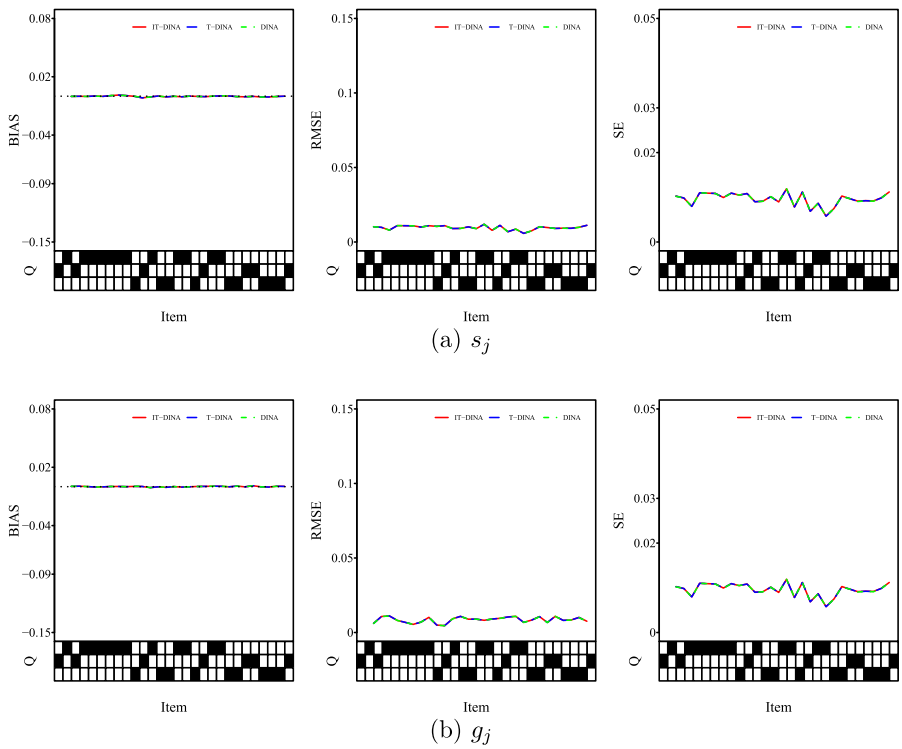


FIGURE 4.
BIAS, RMSE and SE of slipping and guessing parameters under Case 4 considering $N = 4000$. X-axis is the Q-matrix, where a black square denotes the presence of the attribute.

results for fitting the IT-DINA, T-DINA, and DINA models were nearly identical across different sample sizes. Further, under Case 5 (T-DINA), the IT-DINA and T-DINA models performed similarly and better than the DINA model across different sample sizes.

Table 13 shows the recovery of $\boldsymbol{\pi}$ under Case 1. As the sample size increased, BIAS reduced to near 0 under the proposed model. Under the T-DINA and DINA models, estimates of $\boldsymbol{\pi}$ were biased, especially for $\pi_{(0,0,0)}$, and the values of RMSE and BR were remarkably larger.

Table 14 shows the recovery of parameters $\boldsymbol{\pi}$ under Case 2. In line with Case 1, absolute values of bias under IT-DINA were near 0 when $N = 4000$, whereas estimators under the T-DINA and DINA models were biased. Further, with stronger true interactions, the T-DINA and DINA model parameter estimates under Case 2 were much less accurate than those under Case 1.

5. Discussion

This paper aimed to provide an in-depth inspection of the model formulation and identifiability of testlet-effect DCMs. In particular, we discussed the need for modeling potential dependencies between attributes and testlet effects. When the attribute-response relationship could be described by the DINA and DINO models, an approach to account for this dependency via an interaction term is proposed. The proposed IT-DINA model generalizes the testlet DINA model, with potential dependency modeled through $\gamma(2\eta - 1)\xi_2$, an interaction term between the attribute profile and

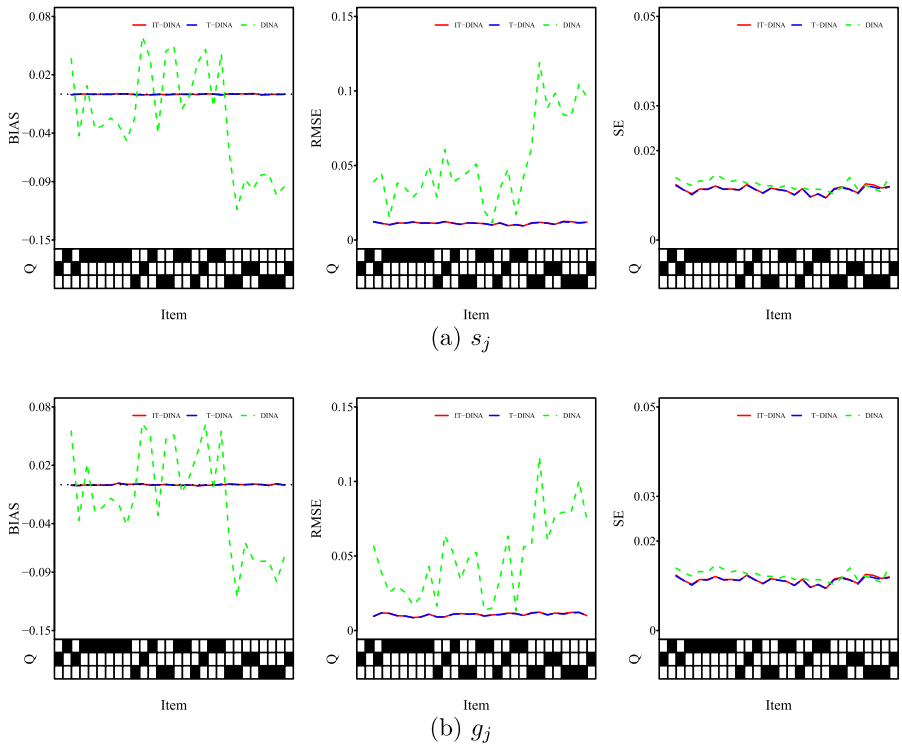


FIGURE 5.
BIAS, RMSE and SE of slipping and guessing parameters under Case 5 considering $N = 4000$. X-axis is the Q-matrix, where a black square denotes the presence of the attribute.

the testlet effect. This specific form of interaction facilitated model comparisons (i.e., testing for the independence between ideal response and testlet effect). Model selection based on BIC demonstrated good performance in selecting the data-generating model under most simulation conditions. In turn, simulation and empirical results from the PISA 2015 CPS assessment showed the utility of adopting the more general IT-DINA model in capturing the potential additional dependency between ideal response and testlet effect.

Another contribution of the current work is the identifiability conditions for testlet-effect DINA and DINO models. The formulation of the dependency as an interaction effect facilitated the derivation of the identifiability results for IT-DINA. In addition, the probit link allowed derivation of an explicit formula of conditional probability $P(Y|\eta^{\boldsymbol{\alpha}})$ after integrating out the normally distributed testlet effect $\boldsymbol{\xi}$. This allowed us to establish easy-to-check sufficient conditions for the identifiability of the IT-DINA model, as well as the T-DINA model as a special case.

While the proposed IT-DINA model extends the T-DINA to account for potential structural dependency between the testlet effects and attribute patterns, one should note that the IT-DINA could still remain an oversimplification: The DINA and DINO models are very special cases of DCMs where the attribute and ideal response relationship is fully conjunctive or disjunctive. When the T-DINA model does not fit the observed data well, the presence of interaction effects (i.e., dependency) is one potential explanation, but it could also be due to a misspecified model, for instance, when the DINA model or the Q-matrix does not fully account for the actual attribute-response relationship (see the discussion of Eq. (6)). In empirical analyses, adopting

TABLE 12.
The recovery of transformed slipping and guessing parameters under the DINA, T-DINA and IT-DINA model.

	Case	N	IT-DINA			T-DINA			DINA		
			BIAS	RMSE	SE	BIAS	RMSE	SE	BIAS	RMSE	SE
s_j	1	500	−0.0000	0.0312	0.0311	−0.0033	0.0322	0.0315	−0.0110	0.0386	0.0310
		1000	−0.0001	0.0222	0.0221	−0.0037	0.0235	0.0226	−0.0113	0.0320	0.0222
		2000	0.0001	0.0153	0.0152	−0.0032	0.0169	0.0156	−0.0109	0.0275	0.0154
		4000	−0.0002	0.0109	0.0109	−0.0035	0.0128	0.0112	−0.0112	0.0253	0.0111
	2	500	0.0001	0.0326	0.0325	−0.0110	0.0630	0.0540	−0.0300	0.0809	0.0379
		1000	−0.0003	0.0229	0.0228	−0.0101	0.0534	0.0327	−0.0307	0.0767	0.0262
		2000	−0.0001	0.0158	0.0157	−0.0095	0.0487	0.0216	−0.0305	0.0741	0.0182
		4000	−0.0003	0.0112	0.0112	−0.0097	0.0463	0.0151	−0.0309	0.0731	0.0129
	3	500	−0.0009	0.0352	0.0342	−0.0059	0.0618	0.0523	0.0016	0.0622	0.0487
		1000	0.0000	0.0245	0.0238	−0.0053	0.0463	0.0357	0.0017	0.0497	0.0330
		2000	0.0000	0.0170	0.0165	−0.0042	0.0374	0.0251	0.0030	0.0430	0.0231
		4000	−0.0002	0.0121	0.0118	−0.0042	0.0321	0.0177	0.0028	0.0389	0.0163
	4	500	−0.0000	0.0276	0.0273	−0.0000	0.0275	0.0271	−0.0000	0.0273	0.0270
		1000	−0.0001	0.0194	0.0192	−0.0001	0.0193	0.0191	−0.0001	0.0193	0.0191
		2000	−0.0000	0.0137	0.0135	−0.0001	0.0137	0.0135	−0.0001	0.0136	0.0135
		4000	−0.0002	0.0097	0.0096	−0.0002	0.0097	0.0096	−0.0002	0.0097	0.0096
	5	500	0.0003	0.0326	0.0325	0.0000	0.0317	0.0316	−0.0247	0.0683	0.0355
		1000	0.0002	0.0228	0.0227	−0.0001	0.0224	0.0223	−0.0243	0.0630	0.0247
		2000	−0.0000	0.0158	0.0158	−0.0003	0.0156	0.0156	−0.0244	0.0611	0.0175
		4000	−0.0001	0.0113	0.0113	−0.0002	0.0112	0.0112	−0.0244	0.0595	0.0124
g_j	1	500	−0.0012	0.0296	0.0293	−0.0036	0.0307	0.0298	−0.0080	0.0366	0.0288
		1000	−0.0005	0.0205	0.0203	−0.0028	0.0217	0.0208	−0.0074	0.0303	0.0203
		2000	−0.0004	0.0144	0.0143	−0.0029	0.0159	0.0147	−0.0075	0.0261	0.0143
		4000	−0.0002	0.0101	0.0100	−0.0027	0.0118	0.0104	−0.0071	0.0241	0.0101
	2	500	−0.0006	0.0314	0.0313	−0.0051	0.0600	0.0529	−0.0203	0.0758	0.0353
		1000	−0.0006	0.0219	0.0218	−0.0016	0.0490	0.0311	−0.0200	0.0717	0.0244
		2000	−0.0002	0.0153	0.0152	−0.0007	0.0445	0.0206	−0.0195	0.0688	0.0170
		4000	−0.0001	0.0107	0.0107	−0.0004	0.0424	0.0144	−0.0195	0.0678	0.0120
	3	500	−0.0002	0.0302	0.0299	−0.0053	0.0509	0.0459	−0.0107	0.0532	0.0451
		1000	−0.0002	0.0214	0.0212	−0.0040	0.0371	0.0311	−0.0090	0.0399	0.0307
		2000	−0.0001	0.0149	0.0147	−0.0039	0.0293	0.0216	−0.0093	0.0331	0.0213
		4000	−0.0002	0.0108	0.0106	−0.0038	0.0249	0.0154	−0.0091	0.0290	0.0152
	4	500	−0.0000	0.0249	0.0244	−0.0001	0.0248	0.0242	−0.0001	0.0246	0.0241
		1000	0.0001	0.0176	0.0172	0.0001	0.0175	0.0172	0.0001	0.0175	0.0171
		2000	−0.0000	0.0124	0.0121	−0.0000	0.0123	0.0121	−0.0000	0.0123	0.0120
		4000	0.0001	0.0087	0.0085	0.0001	0.0087	0.0085	0.0001	0.0087	0.0085
	5	500	−0.0001	0.0314	0.0312	−0.0007	0.0304	0.0303	−0.0152	0.0640	0.0331
		1000	0.0000	0.0218	0.0216	−0.0002	0.0214	0.0213	−0.0150	0.0594	0.0232
		2000	0.0002	0.0153	0.0152	0.0001	0.0151	0.0150	−0.0145	0.0569	0.0163
		4000	−0.0001	0.0108	0.0107	−0.0001	0.0107	0.0106	−0.0147	0.0556	0.0115

the DINA/DINO models requires researchers to thoroughly evaluate the appropriateness of this assumption, for instance, by performing model comparison under general DCMs (e.g., de la Torre, 2011).

TABLE 13.
BIAS and RMSE of $\boldsymbol{\pi}$ under Case 1.

		N	$\boldsymbol{\pi}_{(0,0,0)}$	$\boldsymbol{\pi}_{(1,0,0)}$	$\boldsymbol{\pi}_{(0,1,0)}$	$\boldsymbol{\pi}_{(0,0,1)}$	$\boldsymbol{\pi}_{(1,1,0)}$	$\boldsymbol{\pi}_{(1,0,1)}$	$\boldsymbol{\pi}_{(0,1,1)}$	$\boldsymbol{\pi}_{(1,1,1)}$
BIAS	IT-DINA	500	0.0006	−0.0013	−0.0008	−0.0019	−0.0008	0.0001	−0.0001	0.0041
		1000	0.0007	−0.0009	−0.0004	−0.0008	−0.0007	0.0003	0.0002	0.0016
		2000	−0.0005	−0.0003	−0.0003	−0.0009	0.0003	0.0001	0.0002	0.0014
		4000	−0.0004	−0.0004	0.0001	−0.0007	0.0003	0.0002	0.0001	0.0007
	T-DINA	500	0.0088	−0.0019	−0.0041	−0.0050	−0.0044	−0.0023	−0.0019	0.0108
		1000	0.0101	−0.0012	−0.0039	−0.0047	−0.0048	−0.0029	−0.0021	0.0094
		2000	0.0090	−0.0008	−0.0043	−0.0045	−0.0045	−0.0029	−0.0020	0.0101
		4000	0.0095	−0.0012	−0.0040	−0.0045	−0.0044	−0.0030	−0.0020	0.0097
	DINA	500	0.0115	−0.0002	−0.0057	−0.0032	−0.0067	−0.0041	0.0011	0.0073
		1000	0.0120	0.0004	−0.0054	−0.0019	−0.0078	−0.0049	0.0012	0.0064
		2000	0.0113	0.0006	−0.0059	−0.0023	−0.0077	−0.0048	0.0017	0.0071
		4000	0.0116	0.0004	−0.0056	−0.0021	−0.0073	−0.0051	0.0014	0.0066
RMSE	IT-DINA	500	0.0179	0.0192	0.0172	0.0184	0.0184	0.0175	0.0180	0.0205
		1000	0.0134	0.0131	0.0126	0.0126	0.0127	0.0123	0.0121	0.0127
		2000	0.0094	0.0090	0.0090	0.0086	0.0089	0.0089	0.0088	0.0097
		4000	0.0066	0.0064	0.0065	0.0063	0.0063	0.0064	0.0063	0.0066
	T-DINA	500	0.0209	0.0190	0.0176	0.0188	0.0187	0.0179	0.0181	0.0230
		1000	0.0178	0.0130	0.0133	0.0135	0.0134	0.0131	0.0121	0.0166
		2000	0.0133	0.0091	0.0099	0.0097	0.0099	0.0094	0.0090	0.0143
		4000	0.0117	0.0067	0.0077	0.0078	0.0075	0.0072	0.0066	0.0120
	DINA	500	0.0215	0.0181	0.0171	0.0183	0.0190	0.0177	0.0176	0.0203
		1000	0.0182	0.0123	0.0133	0.0123	0.0147	0.0137	0.0117	0.0143
		2000	0.0145	0.0087	0.0106	0.0091	0.0114	0.0099	0.0088	0.0118
		4000	0.0133	0.0064	0.0086	0.0066	0.0095	0.0082	0.0065	0.0092

TABLE 14.
BIAS and RMSE of $\boldsymbol{\pi}$ under Case 2.

		N	$\boldsymbol{\pi}_{(0,0,0)}$	$\boldsymbol{\pi}_{(1,0,0)}$	$\boldsymbol{\pi}_{(0,1,0)}$	$\boldsymbol{\pi}_{(0,0,1)}$	$\boldsymbol{\pi}_{(1,1,0)}$	$\boldsymbol{\pi}_{(1,0,1)}$	$\boldsymbol{\pi}_{(0,1,1)}$	$\boldsymbol{\pi}_{(1,1,1)}$
BIAS	IT-DINA	500	0.0029	−0.0011	−0.0020	−0.0021	−0.0003	−0.0007	−0.0012	0.0044
		1000	0.0015	−0.0012	−0.0008	−0.0005	−0.0007	−0.0003	−0.0002	0.0020
		2000	−0.0003	−0.0003	−0.0005	−0.0005	0.0002	0.0004	−0.0001	0.0011
		4000	−0.0001	−0.0004	0.0001	−0.0006	0.0003	0.0003	−0.0002	0.0006
	T-DINA	500	0.0266	0.0012	−0.0076	−0.0103	−0.0080	−0.0044	−0.0117	0.0143
		1000	0.0286	0.0020	−0.0066	−0.0085	−0.0086	−0.0041	−0.0137	0.0108
		2000	0.0283	0.0025	−0.0069	−0.0089	−0.0085	−0.0042	−0.0130	0.0107
		14000	0.0281	0.0025	−0.0057	−0.0088	−0.0084	−0.0045	−0.0135	0.0104
	DINA	500	0.0252	0.0074	−0.0134	−0.0041	−0.0169	−0.0081	0.0024	0.0075
		1000	0.0261	0.0083	−0.0148	−0.0024	−0.0172	−0.0092	0.0030	0.0060
		2000	0.0256	0.0095	−0.0141	−0.0035	−0.0166	−0.0094	0.0022	0.0063
		4000	0.0254	0.0089	−0.0134	−0.0031	−0.0171	−0.0095	0.0026	0.0062
RMSE	IT-DINA	500	0.0192	0.0186	0.0174	0.0185	0.0183	0.0179	0.0174	0.0203
		1000	0.0137	0.0131	0.0125	0.0124	0.0126	0.0124	0.0119	0.0131
		2000	0.0093	0.0091	0.0091	0.0084	0.0087	0.0091	0.0091	0.0096
		4000	0.0067	0.0064	0.0065	0.0063	0.0060	0.0065	0.0064	0.0067
	T-DINA	500	0.0394	0.0256	0.0247	0.0257	0.0239	0.0236	0.0251	0.0323
		1000	0.0348	0.0169	0.0175	0.0190	0.0187	0.0182	0.0210	0.0223
		2000	0.0317	0.0120	0.0137	0.0137	0.0138	0.0133	0.0170	0.0177
		4000	0.0300	0.0091	0.0104	0.0118	0.0114	0.0096	0.0158	0.0147
	DINA	500	0.0338	0.0239	0.0238	0.0217	0.0256	0.0220	0.0197	0.0216
		1000	0.0303	0.0169	0.0195	0.0148	0.0218	0.0172	0.0142	0.0160
		2000	0.0278	0.0144	0.0172	0.0097	0.0188	0.0135	0.0105	0.0123
		4000	0.0265	0.0117	0.0150	0.0079	0.0182	0.0118	0.0078	0.0094

As the DINA/DINO are special cases to general DCMs, there are clear limitations to the scope of the current work. In particular, the proposed IT-DINA model only considers the situation where the attribute pattern enters the measurement model and influences testlet effect distribution via the conjunctive ideal response. The addition of a single interaction term between the ideal response and a testlet effect likely will not be adequate under a more complex DCM. How to parameterize the potential dependency between testlet effects and attribute patterns under general DCMs remains an open question. Further, the current identifiability results for the IT-DINA model and the T-DINA model are only applicable to cases when the DINA and DINO DCMs are appropriate. Although more general testlet-effect DCMs have been proposed (Hansen, 2013; Ma et al., 2023) , identifiability conditions for these general models are not yet available. We hope the current discussion of the simple case provides initial results for addressing these open questions related to testlet-effect DCMs.

Aside from extension to more general DCMs, there are several directions for future work. Firstly, testlet-effect DINA and its identifiability conditions may be extended to model polytomous responses. In the literature, there exist several different parameterizations for polytomous DCMs, e.g., Fang et al. (2019), Culpepper (2019), and Ma (2019). We expect the incorporation of testlet effect and potential latent dependency to differ depending on the adopted parameterization. Secondly, the current identifiability results assumed the Q-matrix to be known. Future research may consider extension of the results to situations of unknown Q. Thirdly, in the current paper, the dependence is modeled through a specific interaction term. Alternative formulation of dependency may be of interest. Lastly, with the development of computer-based test, we have access to not only the response data but also additional data such as the response times. How to incorporate these more complicated data may require more sophisticated modeling techniques.

Funding This project is supported in part by the National Science Foundation (DMS-2015417), China Postdoctoral Science Foundation (2021M700466), and the China National Natural Science Foundation (12301376, 12371263).

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.

Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64*(2), 153–168.

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika, 75*(4), 581–612.

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.

Chen, Y., Liu, Y., & Xu, S. (2018). Mutual information reliability for latent class analysis. *Applied Psychological Measurement, 42*(6), 460–477.

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

Coelho, P. S., & Pereira, L. N. (2011). A spatial unit level model for small area estimation. *REVSTAT-Statistical Journal, 9*(2), 155–180.

Culpepper, S. A. (2015). Bayesian estimation of the DINA model with Gibbs sampling. *Journal of Educational and Behavioral Statistics, 40*(5), 454–476.

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

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.
de La Torre, J., & Karelitz, T. M. (2009). Impact of diagnosticity on the adequacy of models for cognitive diagnosis under a linear attribute structure: A simulation study. *Journal of Educational Measurement, 46*(4), 450–469.
DeMars, C. E. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. *Journal of Educational Measurement, 43*(2), 145–168.
Fang, G., Liu, J., & Ying, Z. (2019). On the identifiability of diagnostic classification models. *Psychometrika, 84*(1), 19–40.
Fang, G., Guo, J., Xu, X., Ying, Z., & Zhang, S. (2021). Identifiability of bifactor models. *Statistica Sinica, 31*, 2309–2330.
Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika, 57*(3), 423–436.
Gu, Y. (2020). *Statistical analysis of structured latent attribute models* (Unpublished doctoral dissertation).
Gu, Y., & Xu, G. (2019b). 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. *The 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*.
Gu, Y., & Xu, G. (2022). Generic identifiability of the DINA model and blessing of latent dependence. *Psychometrika*.
Hansen, M. (2013). *Hierarchical item response models for cognitive diagnosis*. University of California.
Hansen, M., Cai, L., Monroe, S., & Li, Z. (2016). Limited-information goodness-of-fit testing of diagnostic classification item response models. *British Journal of Mathematical and Statistical Psychology, 69*(3), 225–252.
Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74*(2), 191–210.
Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika, 77*(3), 442–454.
Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258–272.
Kang, H.-A., Liu, J., & Ying, Z. (2017). A graphical diagnostic classification model. Preprint at arXiv:1707.06318
Kish, L., Namboodiri, N. K., & Pillai, R. K. (1962). The ratio bias in surveys. *Journal of the American Statistical Association, 57*(300), 863–876.
Köhn, H.-F., & Chiu, C.-Y. (2016). A proof of the duality of the DINA model and the DINO model. *Journal of Classification, 33*, 171–184.
Ma, W. (2019). A diagnostic tree model for polytomous responses with multiple strategies. *British Journal of Mathematical and Statistical Psychology, 72*(1), 61–82.
Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology, 69*(3), 253–275.
Ma, W., Wang, C., & Xiao, J. (2023). A testlet diagnostic classification model with attribute hierarchies. *Applied Psychological Measurement*, 01466216231165315.
Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics, 2*(2), 99–120.
Meng, X.-L. (1993). On the absolute bias ratio of ratio estimators. *Statistics & Probability Letters, 18*(5), 345–348.
OECD. (2016). *Pisa 2015 assessment and analytical framework: Science, reading, mathematic and financial literacy*. Author Paris.
Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.
Sha, S. (2016). *Nonparametric diagnostic classification analysis for testlet-based tests (Unpublished doctoral dissertation)*. The University of North Carolina at Greensboro.
Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement, 28*(3), 237–247.
Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of educational measurement, 20*(4), 345–354.
Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), 287.
Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge University Press.
Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *The Annals of Statistics, 45*(2), 675–707.
Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika, 81*(3), 625–649.
Yavuz, E., & Atar, H. Y. (2020). An examination of Turkish students' PISA 2015 collaborative problem-solving competencies. *International Journal of Assessment Tools in Education, 7*(4), 588–606.
Zhan, P., Li, X., Wang, W.-C., Bian, Y., & Wang, L. (2015). The multidimensional testlet-effect cognitive diagnostic models. *Acta Psychologica Sinica*.
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.

Manuscript Received: 1 AUG 2023