A Generalized Nonparametric Classification Method for Small Samples Using Qc Matrix

Authors: Chen, Xingyu, Yan, Yakun, Kang,Chunhua, Kang,Chunhua

Date: 2025-12-04T00:00:00+00:00

Abstract

Objective: This study integrates the concept of the QC matrix with the generalized nonparametric classification method (GNPC) and extends GNPC to the generalized sequential nonparametric classification method (seq-GNPC) for application to polytomously scored items. The primary focus is on addressing the challenges of saturated models that exhibit poor discriminability under small-sample conditions, wherein parametric models cannot be estimated.

Methods: Both simulation studies and empirical research were conducted. The simulation conditions were as follows: the number of attributes (K) was set to 3 or 5; the number of items (J) was set to 20 or 40; the proportion of polytomous items was fixed at 50%. Two types of QC matrices were also considered: restricted and unrestricted.

Results: Simulation results indicate that seq-GNPC outperforms parametric methods in small samples when the data pattern conforms to a saturated model. When the data pattern conforms to a reduced model, parametric methods exhibit severe convergence failure under small-sample conditions, whereas seq-GNPC guarantees a 100% estimation success rate. Empirical results further demonstrate that seq-GNPC is more stable than seq-GDINA in small-sample cases, with higher attribute pattern recovery rates and smaller standard deviations.

Limitations: While this study documented the number of replications in which the parametric model could be successfully estimated under each condition, it did not thoroughly investigate the specific sample sizes or conditions under which parametric models would overcome convergence failure, which warrants further exploration.

Conclusions: The seq-GNPC method proposed in this paper demonstrates strong applicability in small-sample contexts and effectively addresses the

challenges of diagnostic assessment for polytomous or mixed-format scoring instruments.

Full Text

Preamble

A Generalized Nonparametric Classification Method for Small Samples Using the Qc Matrix

Xingyu CHEN[1], Yakun YAN[2] & Chunhua KANG[1*]

[1]Zhejiang Philosophy and Social Science Laboratory for the Mental Health and Crisis Intervention of Children and Adolescents, Zhejiang Normal University, Jinhua, 321004, China
[2]Jinhua Education College, Jinhua, 321004, China

Author Contribution Statement
Xingyu CHEN: Responsible for generating simulated data, collecting empirical data, analyzing data, and drafting the paper.
Yakun YAN: Responsible for revising the final version.
Chunhua KANG: Proposed research ideas and designed research schemes.

This research was supported by the Humanities and Social Sciences Fund of the Ministry of Education (22YJA190005) and the Key Open Fund from Zhejiang Philosophy and Social Science Laboratory for the Mental Health and Crisis Intervention of Children and Adolescents, PR China (No. 23MHCICAZD04).

Abstract

Objective: This study integrates the Qc matrix concept with the generalized nonparametric classification method (GNPC) and extends GNPC to a generalized sequential nonparametric classification method (seq-GNPC) applicable to graded scoring items. The primary focus addresses two key issues: the poor discrimination of saturated models under small samples and the inability to estimate parametric models in small-sample contexts.

Methods: The study employs both simulation studies and empirical research. Simulation conditions include: number of attributes K set to 3 or 5; number of items J set to 20 or 40; proportion of graded scoring items fixed at 50%; and two types of Qc matrices (restricted and unrestricted).

Results: Simulation results demonstrate that seq-GNPC outperforms parametric methods in small samples when data patterns conform to saturated models. When data patterns conform to reduced models, parametric models exhibit significant convergence failures in small-sample situations, whereas seq-GNPC guarantees a 100% estimation rate. Empirical results further show that seq-GNPC is more stable than seq-GDINA in small-sample cases, with higher attribute pattern retest rates and smaller standard deviations.

Limitations: This study counted the number of replicates where parametric models could be estimated under each condition but did not deeply examine the specific circumstances (e.g., number of attributes, number of graded scoring items) under which parametric models would cease to fail in diagnostic classification at each sample size—an area requiring further exploration.

Conclusions: The seq-GNPC method proposed in this paper demonstrates good applicability in small samples and can effectively solve diagnostic evaluation problems for graded or mixed scoring programs in small-sample contexts.

Keywords: generalized nonparametric classification method; nonparametric classification method; graded scoring; Qc matrix; seq-GDINA

1. Introduction

In recent years, educational assessment has become an increasingly significant area of research, with diagnostic educational assessment modes such as cognitive diagnostic assessment (CDA) playing a vital role in this evolution. CDA represents a key component of the new generation of psychometric theories, focusing on understanding individuals' cognitive processing (Wang & Gierl, 2011). Through CDA, researchers can gain deep insights into individuals' cognitive structures and identify their strengths and weaknesses in cognitive attribute or skill mastery, making the approach particularly well-suited for diagnostic educational assessment (Guo et al., 2024).

Early cognitive diagnostic assessment methods were primarily parametric, with parametric models containing item and structural parameters that follow particular distributions and require parameter estimation methods for computation. For dichotomously scored items, examples include the deterministic inputs, noisy "and" gate (DINA) model (Junker & Sijtsma, 2001), the deterministic inputs, noisy "or"gate (DINO) model (Templin & Henson, 2006), and the generalized DINA (GDINA) model (de la Torre, 2011). Dichotomous items are limited to scored (1) and not scored (0) responses, providing restricted information. To meet practical assessment needs, the importance of graded scoring items has gradually increased. Graded scoring items refer to items with three or more scoring levels (e.g., items with 0, 1, and 2 scoring levels). For graded scoring programs, models include the General Diagnostic Model (GDM; Davier, 2008), sequential generalized DINA (seq-GDINA; Ma & Torre, 2016), the DINA model for graded data (DINA-GD; Tu et al., 2018), and polytomous CDMs (Gao et al., 2021), among others. Recent studies have also proposed cognitive diagnostic models for multiple strategies in individual problem solving (Liao & Jiao, 2023; Ma, 2019; Ma & Guo, 2019; Wei et al., 2023). Additionally, researchers have applied machine learning to cognitive diagnosis to construct machine learning-based cognitive diagnostic assessment methods (Cuerda et al., 2022; Gao et al., 2022; Li et al., 2022; Zhang et al., 2023).

In the aforementioned graded scoring cognitive diagnostic assessment methods, most models rely primarily on the Q-matrix for diagnostic assessment, considering only the relationship between items and attributes while neglecting the relationship between each scoring level and attributes. The seq-GDINA model was proposed to address sequential response data discrimination, introducing a Qc matrix that can indicate the sequential relationship between attributes and item categories. The Qc matrix can specify the relationship between categories and attributes in graded scoring items. By configuring the Qc matrix, the model can distinguish between sequential and non-sequential graded scoring data. Consequently, the model has been applied in various contexts, including second-language assessment construction (Yuan et al., 2022), English writing proficiency diagnosis (Shi et al., 2024), and mathematical arithmetic examination diagnosis (Saso et al., 2023). The seq-GDINA model requires responses at all categories of all graded scoring items to be present for practical application (Ma & Torre, 2020), a condition easily met in large samples but not necessarily in small samples. It is well known that the purpose of cognitive diagnostic assessment is to identify or diagnose strengths or deficiencies in a student' s (individual' s) knowledge structure, providing information and a basis for teachers' instruction or students' self-study. Thus, daily educational assessment occurs primarily in small-sample contexts such as schools and classrooms (Chiu et al., 2018). However, in small-sample situations, commonly used parametric estimation methods such as Marginal Maximum Likelihood Estimation-Expectation Maximization (MMLE-EM) may encounter boundary problems (Sorrel et al., 2023), reducing estimation effectiveness. Nonparametric cognitive diagnostic methods are less affected by sample size and can maintain better diagnostic classification effects in small samples, such as the nonparametric classification method (NPC) suitable for 0-1 scoring proposed by Chiu and Douglas (2013) and the subsequent generalized nonparametric classification method (GNPC; Chiu et al., 2018). However, these methods do not consider sequential response data or discrimination of generalized data under graded scoring. Therefore, this study proposes applying the Qc matrix from seq-GDINA to GNPC to construct a generalized sequential nonparametric classification method (seq-GNPC) suitable for small-sample sequential response data, thereby meeting practical needs for diversified data types and scoring modes in small-sample assessment contexts.

Methodological details are presented in the Technical Background and Methodology Presentation sections, followed by Simulation and Empirical Studies sections where we test and compare these models using simulated and empirical data. Finally, in the Discussion and Conclusion section, we summarize the study results and their practical implications.

2.1 Cognitive Diagnosis Models

The GDINA model is a generalized dichotomous scoring model proposed by de la Torre (2011) that considers both main effects of attributes and interaction effects between attributes in its formulation:

$$p_j(\alpha_{lj}^*) = \phi_{j0} + \sum \phi_{jk}\alpha_{lk} + \sum\sum \phi_{jkk'}\alpha_{lk}\alpha_{lk'}^{k'=k+1} + \cdots + \phi_{j12\ldots K_j^*}\prod \alpha_{lk}$$

Equation 1

Here, ϕ_{j0} is the intercept term for item j, representing the guessing probability of item j; ϕ_{jk} is the main effect of attribute α_k on item j, representing the change in correct response probability after mastering this attribute alone; $\phi_{jkk'}$ is the interaction between attributes α_k and $\alpha_{k'}$, representing the change in correct response probability when both attributes are simultaneously mastered; and $\phi_{j12\ldots K_j^*}$ is the interaction among attributes α_k to $\alpha_{K_j^*}$, representing the change in response probability when mastering all attributes examined by item j.

Generalized models, also known as saturated models, account for the full effects between attributes. In contrast, reduced models constrain some effects and have simpler structures, with the DINA model being the classical reduced model. The DINA model can be obtained by setting all parameters in the GDINA model to zero except ϕ_{j0} and $\phi_{j12\ldots K_j^*}$. The DINA model is a conjunctive model where an individual must master all attributes examined by an item to score. Its item response function (IRF) is:

$\xi_{ij}^{g_j}\xi_{ij}$ is the latent response variable, Q_{jk} is set to be the kth attribute in item j (i.e., row j and column k in t

$$P(X_{ij} = 1|\alpha, s, g) = (1 - s_j)^{1-\xi_{ij}} \equiv P_j(\alpha_{i\bullet})$$

Equation 2

$$\xi_{ij} = \prod_{k:Q_{jk}=1}\alpha_{ik} = \prod \alpha_{ik}$$

Equation 3

Here, s_j is the slip probability of item j, and g_j is the guess probability of item j, calculated by combining the individual' s potential response variable on item j with observed response values:

$$s_j = P(X_{ij} = 0|\xi_{ij} = 1)$$

Equation 4

$$g_j = P(X_{ij} = 1|\xi_{ij} = 0)$$

Equation 5

In contrast to the conjunctive model, the disjunctive model allows an individual to score by mastering at least one of the attributes examined by the item. The DINO model is typically a disjunctive model where latent response variables are defined as:

$$\omega_{ij} = 1 - \prod (1 - \alpha_{ik})^{Q_{jk}}$$

Equation 6

The computation of IRF, s_j, and g_j for the DINO model is identical to that for the DINA model except for the computation of latent category variables. The DINO model can also be obtained by setting parameters in the GDINA model.

Ma & Torre (2016) proposed the sequential generalized deterministic inputs, noisy "and" gate (seq-GDINA) based on GDINA for graded scoring data to consider sequential responses. seq-GDINA modifies the traditional Q-matrix. The traditional Q-matrix is a $J \times K$ matrix indicating which attributes were examined by each item (1 for examined, 0 for not examined). For graded responses, the category Q-matrix, called the Qc matrix, is proposed as a $\sum_{j=1}^{H_j} \times K$ matrix, with H_j being the highest category for item j. Each row of the Qc matrix has K elements indicating which attributes were examined for that category. When the relationship between attributes and categories is clear, a restricted Qc matrix can be established, with attributes examined in each category of the item clearly labeled (see Table 1 for an example). In practical applications, it is not guaranteed that the relationship between categories and attributes can be understood, so when this relationship is unclear, an unrestricted Qc matrix can be used, setting all categories of the same item to examine identical attributes (see Items 2 and 3 in Table 2 for specific examples).

If all items are 0-1 scoring items, the Qc matrix is equivalent to the traditional Q-matrix. Although it is assumed that response categories are obtained sequentially, it is not necessary for different categories to measure different attributes, nor do the attributes need to follow any particular structure.

The seq-GDINA design based on the Qc matrix can split graded scoring items into several 0-1 scoring categories, using the GDINA model function as a process function:

$$S_j(b|\alpha^*_{ljb}) = \phi_{jb0} + \sum \phi_{jbk}\alpha_{lk} + \sum\sum \phi_{jbkk'}\alpha_{lk}\alpha_{lk'}^{k'=k+1} + \cdots + \phi_{jb12\ldots K_j^*}\prod \alpha_{lk}$$

Equation 7

The following assumptions are included:

$$S_j(b|\alpha_c) = \begin{cases} 1, & \text{if } b = 0 \\ 0, & \text{if } b = H_j + 1 \end{cases}$$

Equation 8

This is because an individual can always reach category 0 and never reach category $H_j + 1$ (maximum grade of H_j). The process function is key to the seq-GDINA model. Based on the process of category b, the IRF of item j is:

$$P(X_j = b|\alpha_c) = [1 - S_j(b+1|\alpha_c)] \prod S_j(x|\alpha_c)$$

Equation 9

Simultaneously, it is constrained as follows:

$$\sum_{b=0}^{H_j} P(X_j = b|\alpha_c) = 1 \quad \forall c$$

Equation 10

$P(X_j = b|\alpha_c)$ is the probability that an individual with knowledge state α_c scores b on item j. The sum of probabilities that an individual scores in categories 0 through H_j is 1. The seq-GDINA model allows different cognitive processes to be modeled in different categories within a single item and can parameterize each category separately. As with the GDINA model, seq-GDINA and seq-DINO models can be obtained by constraining parameters. The seq-DINA model is obtained by setting all process function parameters to zero except ϕ_{jb0} and $\phi_{jb12\ldots K_j^*}$.

2.2 Nonparametric Method

Chiu and Douglas (2013) proposed a nonparametric discriminant (NPC) based on Hamming distance (HD), which utilizes the distance between observed response patterns (ORP) and ideal response patterns (IRP) to determine an individual' s attribute mastery pattern for 0-1 scoring items.

First, define the ideal response of individual i on item j as $\eta_{ij} = \prod \alpha_{ik}$, then $\eta_i = (\eta_{i1}, \eta_{i2}, \ldots, \eta_{iJ})$. Since there are 2^K attribute mastery patterns α_c, there are correspondingly 2^K ideal responses for each item: $\eta^{(1)}, \eta^{(2)}, \ldots, \eta^{(2^K)}$. Let $d(\mathbf{y}_i, \eta^{(m)})$ be the distance between individual i' s observed response pattern $\mathbf{y}_i$ across all items and the ideal response pattern $\eta^{(m)}$ under the mth attribute mastery pattern. For 0-1 scoring items, a widely used and natural distance metric for clustering is Hamming distance, which counts only the number of times two vectors disagree:

$$d_h(\mathbf{y},\eta)=\sum|y_j-\eta_j|$$

Equation 11

The individual' s attribute mastery pattern is determined by the pattern that minimizes the $d_h(y,\eta)$ value for their observed response pattern.

However, since the variation produced by each item response is not identical, a weighted Hamming distance method can be employed:

$$d_{wh}(\mathbf{y},\eta)=\sum\bar{p}_j(1-\bar{p}_j)|y_j-\eta_j|$$

Equation 12

Here, $\bar{p}_j$ is the score rate for item j.

The penalized Hamming distance method adds a penalty term to the original Hamming distance:

$$d_{gs}(\mathbf{y},\eta)=\sum\omega_g I[y_j=1]|y_j-\eta_j|+\sum\omega_s I[y_j=0]|y_j-\eta_j|$$

Equation 13

Where ω_g is the assigned guessing weight and ω_s is the assigned slipping weight. Considering differences between items, ω_g and ω_s can be replaced by ω_{gj} and ω_{sj}, giving each item its corresponding guess weight and slip weight. When $\omega_g=\omega_s=1$, the penalized Hamming distance method is equivalent to the Hamming distance method. More weight is given to guessing when $g<s$, and more weight to slipping when $g>s$. NPC can be applied to data conforming to conjunctive or disjunctive models such as DINA and DINO.

With the development of generalized models, Chiu et al. (2018) extended NPC to GNPC to address the poor classification performance of generalized parametric models under small samples. This method maintains good discrimination in small samples (e.g., classroom settings) and can accommodate both conjunctive and disjunctive response data:

$$\eta_{lj}^{(w)}=w_{lj}\eta_{lj}^{(c)}+(1-w_{lj})\eta_{lj}^{(d)}=\eta_{lj}^{(d)}+w_{lj}(\eta_{lj}^{(c)}-\eta_{lj}^{(d)})$$

Equation 14

If item j examines K^* attributes ($K^*\le K$), then item j can distinguish 2^{K^*} potential categories instead of 2^K, because attributes beyond the K^* attributes examined by that item provide no information. Thus, the potential categories involved in the item can be collapsed to 2^{K^*}. The collapsed categories are denoted C_l, where $C_l=\cup_{m\in l}C_m$, $C_m=\{i|\alpha_i=\alpha_m\}$, $\alpha_m=(\alpha_l^*,\alpha_\bullet)$, α_l is the ideal attribute mastery pattern for the K_j^* attributes currently examined by item j, and $\alpha_\bullet$ denotes attributes not examined by item j. Suppose item j has $\mathbf{q}_j = (1,1,0)$ when $K = 3$ and $K_j^* = 2$.

Ideally, there are $2^3 = 8$ ideal mastery patterns, but since item j does not examine attribute A_3, the collapsed attribute mastery patterns have $2^2 = 4$ categories. Specifically, $(1,0,0)$ and $(1,0,1)$ belong to the same C_l, $(0,1,0)$ and $(0,1,1)$ belong to the same C_l, $(1,1,0)$ and $(1,1,1)$ belong to the same C_l, and $(0,0,0)$ is a C_l. The attribute mastery patterns in the same C_l share the same w_{lj} for item j. $\hat{\eta}_{lj}^{(w)}$ is the weighted ideal response to item j for an individual with collapsed potential category C_l, $\eta_{lj}^{(c)}$ denotes the ideal response under the conjunctive model (e.g., DINA), $\eta_{lj}^{(d)}$ denotes the ideal response under the disjunctive model (e.g., DINO), and w_{lj} is the weight for item j for an individual with collapsed potential category C_l, given by:

$$w_{lj} = \frac{\sum_{i \in C_l} (y_{ij} - \eta_{lj}^{(d)})}{||C_l||(\eta_{lj}^{(c)} - \eta_{lj}^{(d)})}$$

Equation 15

Here, $||C_l||$ represents the number of individuals whose attribute mastery pattern is C_l. GNPC uses a loss function to represent the total distance between an individual' s observed response and the item' s weighted ideal response, minimizing:

$$d_{lj}^{(w)} = \sum (y_{ij} - \hat{\eta}_{lj}^{(w)})^2 = \sum (y_{ij} - \eta_{lj}^{(d)} - w_{lj}(\eta_{lj}^{(c)} - \eta_{lj}^{(d)}))^2$$

Equation 16

$$d(\mathbf{y}_i, \hat{\eta}_{mj}^{(w)}) = \sum d(y_{ij}, \hat{\eta}_{mj}^{(w)}) = \sum (y_{ij} - \hat{\eta}_{mj}^{(w)})^2$$

Equation 17

The attribute mastery pattern corresponding to the minimum sum of loss function values across all items is taken as the individual' s knowledge state:

$$\hat{\alpha}_i = \arg \min_{m \in \{1, \ldots, 2^K\}} d(\mathbf{y}_i, \hat{\eta}_m^{(w)})$$

Equation 18

Figure 1 [Figure 1: see original paper] presents the basic flow of the GNPC approach for discriminating individual attribute mastery patterns.

3.1 Methodology Presentation: seq-GNPC

GNPC is a 0-1 scoring cognitive diagnostic method based on the Q-matrix. This paper applies the Qc-matrix from seq-GDINA and its splitting concept to GNPC, extending GNPC to a generalized sequential nonparametric classification method (seq-GNPC) suitable for graded scoring. Certain restrictions are imposed on seq-GNPC:

$$\eta_{ljb}^{(c)*} = \begin{cases} 1, & \text{if } b = 0 \\ \prod \alpha_{ik}^{*Q_{jbk}^{*}}, & \text{if } 0 < b < H_j + 1 \\ 0, & \text{if } b = H_j + 1 \end{cases}$$

Equation 19

$$\eta_{ljb}^{(d)*} = \begin{cases} 1, & \text{if } b = 0 \\ 1 - \prod (1 - \alpha_{ik}^{*})^{Q_{jbk}^{*}}, & \text{if } 0 < b < H_j + 1 \\ 0, & \text{if } b = H_j + 1 \end{cases}$$

Equation 20

$$\eta_{ljb}^{(c)} = \begin{cases} \eta_{lj(b-1)}^{(c)*}, & b > 0 \\ 1, & b = 0 \end{cases}$$

Equation 21

$$\eta_{ljb}^{(d)} = \begin{cases} \eta_{lj(b-1)}^{(d)*}, & b > 0 \\ 1, & b = 0 \end{cases}$$

Equation 22

The approach for calculating the total seq-GNPC distance involves first calculating the process distance for each category under each item separately, then summing them:

$$d(\mathbf{y}_i, \eta_m^{(w)}) = \sum d_{hjb}(y_{jb}, \eta_{jb}^{(w)}) = \sum (y_{ljb} - \eta_{ljb}^{(w)})^2$$

Equation 23

Here, y_{ijb} is the 0-1 response of individual i in category b of item j. Assuming the highest category of item j is 3 and individual i scores 2 on item j, then the individual' s responses across the three categories are $y_{ij1} = 1$, $y_{ij2} = 1$, and $y_{ij3} = 0$. If the individual scores 1 on item j, then only $y_{ij1} = 1$, while both y_{ij2} and y_{ij3} are 0. $\eta_{ljb}^{(w)}$ is derived through:

$$w_{ljb} = \frac{\sum_{i \in C_{lb}} (y_{ijb} - \eta_{ljb}^{(d)})}{||C_{lb}||(\eta_{ljb}^{(c)} - \eta_{ljb}^{(d)})}$$

Equation 24

$$\eta_{ljb}^{(w)} = w_{ljb}\eta_{ljb}^{(c)} + (1 - w_{ljb})\eta_{ljb}^{(d)} = \eta_{ljb}^{(d)} + w_{ljb}(\eta_{ljb}^{(c)} - \eta_{ljb}^{(d)})$$

Equation 25

In GNPC, NPC classification results are used as initial input. Therefore, this study similarly extends NPC to a sequential nonparametric classification method (seq-NPC) to serve as the initial attribute mastery model input for seq-GNPC in graded scoring contexts. seq-NPC has the same constraints as seq-GNPC:

$$\eta_{ijb}^{*} = \begin{cases} 1, & \text{if } b = 0 \\ \prod \alpha_{ik}^{*Q_{jbk}^{*}}, & \text{if } 0 < b < H_j + 1 \\ 0, & \text{if } b = H_j + 1 \end{cases}$$

Equation 26

$$\eta_{ijb} = \begin{cases} \eta_{ij(b-1)}\eta_{ljb}^{*}, & \text{if } b > 0 \\ 1, & \text{if } b = H_j + 1 \end{cases}$$

Equation 27

Here, η_{ijb}^{*} is the process ideal response of category b, and η_{ijb} is the ideal response of category b. The seq-NPC method uses sequential splitting to divide graded scoring categories into multiple 0-1 scoring categories, constrains them with certain conditions, then uses NPC to calculate the distance between individual observed responses and ideal responses, and finally sums them:

$$d_h(\mathbf{y}, \eta) = \sum d_{hjb}(y_{jb}, \eta_{jb})$$

Equation 28

The process distance $d_{hjb}(y_{jb}, \eta_{jb})$ references the process function in seq-GDINA. Since NPC proposes three distance computation methods, the process distance is also computed in three ways:

$$d_{hjb}(y_{jb}, \eta_{jb}) = \begin{cases} |y_{jb} - \eta_{jb}|, & \text{Hamming distance} \\ \bar{p}_{jb}(1 - \bar{p}_{jb})|y_{jb} - \eta_{jb}|, & \text{weighted Hamming distance} \\ \omega_g I[y_{jb} = 1]|y_{jb} - \eta_{jb}| + \omega_s I[y_j = 0]|y_{jb} - \eta_{jb}|, & \text{penalized Hamming distance} \end{cases}$$

Equation 29

3.2.1 Simulation Study Design

The simulation study was conducted in two parts: one setting the data pattern to conform to the reduced model (seq-DINA model) and the other to the saturated model (seq-GDINA). The correct response probability is more discretized in reduced models (e.g., DINA model: only g or $1-s$), suggesting that parametric models would experience more convergence failures when estimating small samples following reduced model data patterns. The number of attributes K was set to 3 or 5, and the number of items J was set to 20 or 40.

The proportion of fixed graded scoring items was 50%, meaning that when $J = 20$, there were 10 dichotomous items and 10 graded scoring items, with the maximum graded scoring item category set to 4 categories (0, 1, 2, and 3). Two types of Qc matrices were considered: restricted and unrestricted. Qc matrices were randomly generated using R programs. Under restricted Qc matrix conditions, different categories of graded scoring items would not share the same attributes, while under unrestricted Qc matrices, attributes examined in each category were identical. All Qc matrices were tested for completeness (Köhn & Chiu, 2017). Sample size N was set to 30, 50, 100, and 500.

For data conforming to the seq-DINA model, three item quality levels were established: high, medium, and low. Following parameter settings from Ma & Torre (2016) and Chiu et al. (2018), when item quality was high, slip and guess parameters were randomly generated from a $U(0, 0.1)$ distribution; for moderate quality, parameters were drawn from $U(0, 0.2)$; and for poor quality, from $U(0, 0.3)$. For data conforming to the seq-GDINA model, item parameters were designed following Chiu et al. (2018), as described in the Appendix.

In a uniform distribution, each attribute mastery pattern has equal selection probability ($p = 1/2^K$). For multivariate normal distribution, the latent distribution of attribute mastery patterns follows a multivariate normal distribution with attribute relationships following $MVN(\mathbf{0}_K, \Sigma)$, where Σ is the covariance matrix:

$$\Sigma = \begin{pmatrix} 1 & \cdots & \rho \\ \vdots & \ddots & \vdots \\ \rho & \cdots & 1 \end{pmatrix}$$

where ρ was set to 0.5 ($\rho = 0.5$).

Potential continuous scores $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{iK})$ conforming to $MVN(\mathbf{0}_K, 0.5)$ were randomly generated, and knowledge states $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iK})$ were determined by:

$$\alpha_{ik} = \begin{cases} 1, & \text{if } \theta_{ik} \geq \Phi^{-1}(\cdot) \\ 0, & \text{otherwise} \end{cases}$$

This study used the GDINA package in R to call the seq-GDINA model (GDINA function), while the seq-GNPC method called a self-programmed R program.

The pattern matching rate (PMR) and average attribute matching rate (AAMR) served as evaluation metrics, calculated as follows:

$$PMR = \frac{\sum N_{i\_correct}}{N}$$

$$AAMR = \frac{\sum\sum N_{ik\_correct}}{N \times K}$$

PMR calculates classification accuracy of cognitive diagnostic methods for attribute mastery patterns on each dataset, while AAMR calculates classification accuracy for attribute levels. Additionally, this study counted the number of convergences across 100 replications for different cognitive diagnostic assessment methods, indicating non-estimable conditions (e.g., failure to converge or lack of response for a category in graded scoring items), which were excluded from PMR and AAMR calculations for parametric models.

3.2.2 Empirical Study Design

This study used 12 items from the TIMSS 2007 Mathematics Assessment (M041052, M041281, M041275, M031303, M031309, M031245, M031242A, M031242B, M031242C, M031247, M031173, M031172) as empirical items. These items examined eight attributes total, with specific Qc matrices and attribute meanings available in Ma & Torre (2016). Response data from 10 regions (Singapore, England, Hungary, Sweden, Armenia, Norway, Colombia, Morocco, Qatar, and Yemen) were selected, retaining only those who responded to all items, yielding 1,399 responses.

4. Results

4.1 Simulation Study Results

The main findings are presented using double Y-axis bar charts and line graphs. The line graph represents the number of replicates analyzable by each method under corresponding conditions (convergence count), corresponding to the right Y-axis. Bar charts display mean PMR values for estimated replicates, corresponding to the left Y-axis. "Uni" indicates uniformly distributed student attribute mastery patterns, while "mvn" indicates multivariate normal distribution. Complete results are presented in tables in the Appendix.

4.1.1 Data Patterns Conform to the seq-GDINA Model Overall, when data patterns follow the saturated model (seq-GDINA), seq-GNPC' s PMR and AAMR are generally better than seq-GDINA' s in small samples, showing comparative advantages. Under the simulation conditions, seq-GDINA could estimate all 100 replications in only 55 of 64 conditions, failing to estimate all replications in 9 conditions.

For PMR, differences between seq-GNPC and seq-GDINA were greater with unrestricted versus restricted Qc matrices (Figure 2 [Figure 2: see original paper]) and larger when $K = 5$ versus $K = 3$ (Figure 4 [Figure 4: see original paper]). The same pattern held for AAMR (Figure 3 [Figure 3: see original paper] & Figure 5 [Figure 5: see original paper]). Across all simulation conditions, seq-GNPC' s PMR ranged from 0.44 to 0.99 with standard deviations from 0.003 to 0.101, while AAMR ranged from 0.82 to 1.00 with standard deviations from 0.002 to 0.049. seq-GDINA' s PMR ranged from 0.13 to 1.00 with standard deviations from 0.003 to 0.146, while AAMR ranged from 0.69 to 1.00 with standard deviations from 0.001 to 0.044. Standard deviations for both methods decreased with increasing sample size, with seq-GNPC showing greater decreases.

Item number and Qc matrix type had greater impact on seq-GDINA than on seq-GNPC, while individual attribute pattern distribution had smaller effects on both methods. To compare classification differences between seq-GNPC and seq-GDINA in small samples, ANOVA was performed on PMR and AAMR means for $N = 30, 50$, and 100 conditions. With unrestricted Qc matrices, seq-GNPC performed significantly better than seq-GDINA.

4.1.2 Data Patterns Conform to the seq-DINA Model When data patterns followed the reduced model (seq-DINA), both parametric models (seq-DINA, seq-GDINA) showed higher estimation failure rates, with 77 of 192 conditions unable to estimate all 100 replications. The two nonparametric methods proposed (seq-NPC, seq-GNPC) estimated all replications successfully. Parametric methods had convergence counts ranging from 64 to 100 across conditions, achieving full estimation only at $N = 500$. Overall, the nonparametric methods estimated all replicates and maintained classification effects comparable to parametric methods under small-sample conditions.

In this data model, item number J had relatively small effects on the two nonparametric methods, while item quality had greater effects on all four methods. seq-GNPC' s PMR ranged from 0.57 to 1.00 with standard deviations from 0 to 0.108; seq-GDINA' s PMR ranged from 0.19 to 1.00 with standard deviations from 0 to 0.187; seq-NPC' s PMR ranged from 0.59 to 1.00 with standard deviations from 0 to 0.103; and seq-DINA' s PMR ranged from 0.60 to 1.00 with standard deviations from 0 to 0.144. The estimated standard deviations of both nonparametric methods were smaller than those of parametric methods in small samples, showing greater stability. Figure 6 [Figure 6: see original paper] and Figure 7 [Figure 7: see original paper] present partial results, with Tables 3 and 4 showing corresponding ANOVA results.

4.2 Empirical Study Results

Data were analyzed using both seq-GNPC and seq-GDINA. Results for attribute proportions are shown in Figure 8 [Figure 8: see original paper]. Of eight attributes, four (A3, A5, A7, and A8) showed essentially identical proportions, while three (A1, A2, and A4) varied widely. To better assess seq-GNPC' s practical effectiveness, its sampling retest reliability was evaluated. Random samples of sizes $N = 30, 50$, and 100 were selected from all data for analysis, calculating attribute mastery pattern consistency between these results and those obtained using all data. Each sample size was replicated 100 times, with results shown in Table 5 . In small samples, seq-GNPC was more stable than seq-GDINA, with higher attribute model retest rates and smaller standard deviations. For attribute retest rates, seq-GNPC and seq-GDINA were essentially consistent, though seq-GNPC showed lower standard deviations.

7. Discussion

This paper extends GNPC' s item evaluation scope based on the Qc matrix, obtaining the seq-GNPC method, and also extends GNPC' s pre-input method, NPC, to the seq-NPC method. Simulation results show that both seq-GNPC and seq-NPC maintain good performance under small-sample conditions.

When data conform to saturated models, seq-GNPC' s PMR can outperform seq-GDINA in small samples, consistent with GNPC research findings (Chiu et al., 2018). This addresses the primary research purpose: solving the problem of poor discriminative classification of saturated models in small samples. Combined with research results, this main objective has been achieved. Notably, seq-GDINA' s PMR is very low when $K = 5$ and the Qc matrix is unrestricted, because it parameterizes attribute parameters (Yamaguchi & Okada, 2020) requiring increasingly more parameters as attribute numbers increase. Under identical conditions, unrestricted Qc matrices require more parameter estimation than restricted Qc matrices. According to Equation 7, more parameters must be estimated using unrestricted Qc matrices. With restricted Qc matrices, seq-GDINA is only viable for small samples when fewer attributes are examined and item numbers are large. In other cases, seq-GNPC is preferred for small samples. With unrestricted Qc matrices, seq-GNPC is optimal for small samples. It should be noted that when more attributes are examined with fewer items, seq-GDINA remains inferior to seq-GNPC even at $N = 500$, requiring larger samples for seq-GDINA application.

When data conform to reduced models (conjunctive or disjunctive), the PMRs of the two nonparametric methods proposed are essentially similar to parametric models in small samples, consistent with Sorrel et al. (2023). However, parametric models show more significant diagnostic classification inability in small samples. Even with high item quality, nearly half of the data cannot be discriminated, whereas nonparametric methods not only achieve discriminative classification but also maintain good accuracy. This represents the secondary objective: addressing situations where parametric models cannot be estimated in small samples.

In conclusion, the seq-GNPC method proposed demonstrates good applicability in small samples and can effectively solve diagnostic evaluation problems for graded or mixed scoring programs without the labeling problem (determining the knowledge state corresponding to clustering categories) of clustering methods (Guo & Zhou, 2022). The method is more applicable in most small-sample situations. In actual educational assessment, attribute mastery patterns of students in the same class may be more similar than theoretical distributions, their ORPs will be more uniform, and the probability of scoring in each category of graded scoring items will be smaller than in simulation conditions. Additionally, the number of attributes evaluated in actual assessments may exceed 3 or 5, sometimes examining more attributes. In these cases, saturated parametric models would be unsatisfactory for class assessment, while seq-GNPC consistently produces good results. Furthermore, seq-GNPC can be applied to small samples in psychological assessments where items are mostly Likert-scaled and relationships between cognitive attributes and item categories are mostly uncertain (using unrestricted Qc matrix), conditions under which seq-GNPC is superior to parametric modeling. Nonparametric cognitive diagnostic methods for small samples have wide practical applications, and further research on nonparametric methods needs advancement (Sessoms & Henson, 2018). Nonparametric methods have limitations: they cannot estimate item parameters or structural parameters (Chiu et al., 2018). When sufficient conditions exist to collect large samples, parametric methods remain preferred due to higher stability and greater information provision.

This study counted parametric model estimation replicates under each condition but did not deeply examine specific circumstances (number of attributes, number of graded scoring items, etc.) under which parametric models would cease classification failures at each sample size—an area needing further exploration. Additionally, seq-GNPC is a single-strategy method, whereas Wang et al. (2024) proposed a multi-strategy GNPC method based on GNPC, suggesting that extending seq-GNPC to multi-strategy evaluation represents a worthwhile future research direction. Currently, cognitive diagnostic methods are developing rapidly, and improving their application rate and applicability in practice (Zhang et al., 2023) remains a topic deserving further exploration.

Acknowledgements

This research was supported by the Humanities and Social Sciences Fund of the Ministry of Education (22YJA190005) and the Key Open Fund from Zhejiang Philosophy and Social Science Laboratory for the Mental Health and Crisis Intervention of Children and Adolescents, PR China (No. 23MHCICAZD04).

References

Chiu, C.-Y., & Douglas, J. (2013). A Nonparametric Approach to Cognitive Diagnosis by Proximity to Ideal Response Patterns. *Journal of Classification*, 30(2), 225–250. https://doi.org/10.1007/s00357-013-9132-9

Chiu, C.-Y., Sun, Y., & Bian, Y. (2018). Cognitive Diagnosis for Small Educational Programs: The General Nonparametric Classification Method. *Psychometrika*, 83(2), 355–375. https://doi.org/10.1007/s11336-017-9595-4

Cuerda, C., Zornoza, A., Gallud, J. A., Tesoriero, R., & Ayuso, D. R. (2022). Deep learning assisted cognitive diagnosis for the D-Riska application. *Soft Computing*, 26(2), 665–676. https://doi.org/10.1007/s00500-021-06510-w

Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61(2), 287–307. https://doi.org/10.1348/000711007X193957

de la Torre, J. (2011). The Generalized DINA Model Framework. *Psychometrika*, 76(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7

Gao, L., Zhao, Z., Li, C., Zhao, J., & Zeng, Q. (2022). Deep cognitive diagnosis model for predicting students' performance. *Future Generation Computer Systems*, 126, 252–262. https://doi.org/10.1016/j.future.2021.08.019

Gao, X., Ma, W., Wang, D., Cai, Y., & Tu, D. (2021). A Class of Cognitive Diagnosis Models for Polytomous Data. *Journal of Educational and Behavioral Statistics*, 46(3), 297–322. https://doi.org/10.3102/1076998620951986

Guo, L., & Zhou, W. (2022). Nonparametric methods for cognitive diagnosis to multiple-choice test items. *Acta Psychologica Sinica*, 53(9), 1032–1043. https://doi.org/10.3724/SP.J.1041.2021.01032

Guo, L., Zhou, W., & Li, X. (2024). Cognitive Diagnosis Testlet Model for Multiple-Choice Items. *Journal of Educational and Behavioral Statistics*, 49(1), 32–60. https://doi.org/10.3102/10769986231165622

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. https://doi.org/10.1177/01466210122032064

Köhn, H.-F., & Chiu, C.-Y. (2017). A Procedure for Assessing the Completeness of the Q-Matrices of Cognitively Diagnostic Tests. *Psychometrika*, 82(1), 112–132. https://doi.org/10.1007/s11336-016-9536-7

Li, G., Hu, Y., Shuai, J., Yang, T., Zhang, Y., Dai, S., & Xiong, N. (2022). NeuralNCD: A Neural Network Cognitive Diagnosis Model Based on Multi-Dimensional Features. *Applied Sciences*, 12(19), Article 19. https://doi.org/10.3390/app12199806

Liao, M., & Jiao, H. (2023). Modelling multiple problem-solving strategies and strategy shift in cognitive diagnosis for growth. *British Journal of Mathematical and Statistical Psychology*, 76(1), 20–51. https://doi.org/10.1111/bmsp.12280

Ma, W. (2019). A diagnostic tree model for polytomous responses with multiple strategies. *British Journal of Mathematical and Statistical Psychology*, 72(1), 61–82. https://doi.org/10.1111/bmsp.12137

Ma, W., & Guo, W. (2019). Cognitive diagnosis models for multiple strategies. *British Journal of Mathematical and Statistical Psychology*, 72(2), 370–392. https://doi.org/10.1111/bmsp.12155

Ma, W., & Torre, J. de la. (2020). GDINA: An R Package for Cognitive Diagnosis Modeling. *Journal of Statistical Software*, 93, 1–26. https://doi.org/10.18637/jss.v093.i14

Ma, W., & Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology*, 69(3), 253–275. https://doi.org/10.1111/bmsp.12070

Saso, S., Oka, M., & Uesaka, Y. (2023). Development of Assessment Tools for Depth of Understanding Quantitatively with Cognitive Diagnostic Models. In K. Arai (Ed.), *Advances in Information and Communication* (pp. 766–774). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-28076-4_55

Sessoms, J., & Henson, R. (2018). Applications of Diagnostic Classification Models: A Literature Review and Critical Commentary. *Measurement: Interdisciplinary Research and Perspectives*, 16, 1–17. https://doi.org/10.1080/15366367.2018.1435104

Shi, X., Ma, X., Du, W., & Gao, X. (2024). Diagnosing Chinese EFL learners' writing ability using polytomous cognitive diagnostic models. *Language Testing*, 41(1), 109–134. https://doi.org/10.1177/02655322231162840

Sorrel, M. A., Escudero, S., Nájera, P., Kreitchmann, R. S., & Vázquez-Lira, R. (2023). Exploring Approaches for Estimating Parameters in Cognitive Diagnosis Models with Small Sample Sizes. *Psych*, 5(2), Article 2. https://doi.org/10.3390/psych5020023

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11(3), 287–305. https://doi.org/10.1037/1082-989X.11.3.287

Tu, D., Zheng, C., Cai, Y., Gao, X., & Wang, D. (2018). A Polytomous Model of Cognitive Diagnostic Assessment for Graded Data. *International Journal of Testing*, 18(3), 231–252. https://doi.org/10.1080/15305058.2017.1396465

Wang, C., & Gierl, M. J. (2011). Using the Attribute Hierarchy Method to Make Diagnostic Inferences about Examinees' Cognitive Skills in Critical Reading. *Journal of Educational Measurement*, 48(2), 165–187. https://doi.org/10.1111/j.1745-3984.2011.00142.x

Wang, D., Ma, W., Cai, Y., & Tu, D. (2024). A general nonparametric classification method for multiple strategies in cognitive diagnostic assessment. *Behavior Research Methods*, 56(2), 723–735. https://doi.org/10.3758/s13428-023-02075-8

Wei, J., Luo, L., Cai, Y., & Tu, D. (2023). A Multistrategy Cognitive Diagnosis Model Incorporating Item Response Times Based on Strategy Selection Theories. *Journal of Educational and Behavioral Statistics*, 10769986231200469. https://doi.org/10.3102/10769986231200469

Yamaguchi, K., & Okada, K. (2020). Variational Bayes Inference Algorithm for the Saturated Diagnostic Classification Model. *Psychometrika*, 85(4), 973–995. https://doi.org/10.1007/s11336-020-09739-w

Yuan, L., Liu, Y., Chen, P., & Xin, T. (2022). Development of a New Learning Progression Verification Method based on the Hierarchical Diagnostic Classification Model: Taking Grade 5 Students' Fractional Operations as an Example. *Educational Measurement: Issues and Practice*, 41(3), 69–82. https://doi.org/10.1111/emip.12488

Zhang, S., Liu, J., & Ying, Z. (2023). Statistical Applications to Cognitive Diagnostic Testing. *Annual Review of Statistics and Its Application*, 10(Volume 10, 2023), 651–675. https://doi.org/10.1146/annurev-statistics-033021-111803

Zhang, W., Meng, L., & Liang, B. (2023). EW-KNN: Evaluating information technology courses in high school with a non-parametric cognitive diagnosis method. *Interactive Learning Environments*, 31(10), 6783–6798. https://doi.org/10.1080/10494820.2022.2043912

Note: Figure translations are in progress. See original paper for figures.