ChinaRxiv 中X

Validation and Estimation of Hierarchical Attribute Q-Matrix

Authors: Qin Chunying, Yu Xiaofeng, Yu Xiaofeng

Date: 2022-05-26T15:25:31+00:00

Abstract

Polytomous attributes extend the traditional dichotomous attributes (i.e., two levels, typically defined as 0 and 1) in diagnostic assessments to multiple values (multiple levels can be 0, 1, ⋯). They can describe not only whether examinees have mastered knowledge attributes but also the degree of mastery on these attributes, thereby enabling diagnostic assessments to provide richer details regarding knowledge mastery for test takers. This paper extends the statistic applicable to Q-matrices with dichotomous attributes (S statistic) to Q-matrix validation and estimation under polytomous attributes. Two estimation algorithms are designed under two common conditions: a joint estimation algorithm and an online estimation algorithm. Simulation study results indicate that the joint estimation algorithm is suitable for validating expert-defined initial Q-matrices; when the initial Q-matrix contains few errors, the joint estimation algorithm has a high probability of recovering the correct Q-matrix. The online estimation algorithm is suitable for the online calibration of attribute vectors and item parameters for "new items"; based on a certain number of "base items", the online estimation algorithm can also achieve satisfactory success rates for estimating new items. Empirical data analysis further demonstrates the application of this method.

Full Text

Validation and Estimation of the Polytomous-Attribute Q-Matrix

QIN Chunying[1],[2], **YU Xiaofeng**[1]
[1] School of Psychology, Jiangxi Normal University, Nanchang, 330022, China
[2] School of Mathematics and Information Science, Nanchang Normal University, Nanchang, 330032, China

Abstract

Polytomous attributes extend the traditional binary definition of attributes (i.e., two levels, typically defined as 0 and 1) in cognitive diagnostic assessment to multiple levels (0, 1, ⋯). This extension not only describes whether examinees have mastered knowledge attributes but also characterizes the degree of mastery, thereby providing richer diagnostic information about knowledge acquisition. This study extends the S statistic—originally developed for binary-attribute Q-matrix validation and estimation—to polytomous-attribute contexts. Under two common practical conditions, we propose two estimation algorithms: a joint estimation algorithm and an online estimation algorithm. Simulation results demonstrate that the joint estimation algorithm is suitable for validating an expert-defined initial Q-matrix; when the initial Q-matrix contains few errors, the algorithm can recover the correct Q-matrix with high probability. The online estimation algorithm is appropriate for calibrating attribute vectors and item parameters for "new items" incrementally. Based on a sufficient number of "anchor items," this algorithm achieves satisfactory success rates for new item estimation. An empirical data analysis further illustrates the practical application of these methods.

Keywords: polytomous attributes, Q-matrix, pG-DINA model, S statistic

1 Introduction

With societal development, educational and psychological assessments have evolved beyond overall evaluation. Cognitive Diagnosis Assessment (CDA) provides detailed profiles of students' knowledge mastery and has attracted widespread attention (Leighton & Gierl, 2007; Tatsuoka, 2009; Rupp et al., 2010; Luo, 2019; von Davier & Lee, 2019). Traditional assessments based on Classical Test Theory (CTT) or Item Response Theory (IRT) only provide overall scores or ability estimates. In contrast, CDA offers knowledge states (KS) that can guide student learning, inform teaching practices, and evaluate instructional effectiveness.

In conventional CDA, knowledge mastery is typically described using binary values (0 or 1), where 1 indicates mastery and 0 indicates non-mastery. While this dichotomous approach is simple and interpretable, it is relatively coarse and cannot accurately characterize the degree of mastery. Two students both coded as 0 on an attribute may differ substantially in their actual knowledge level. Consequently, many researchers have considered extending attributes to multiple levels (Karelitz, 2004; von Davier, 2005; Chen & de la Torre, 2013; Sun et al., 2013; Cai & Tu, 2015; Tu & Cai, 2015; Zhan et al., 2016; Zhan et al., 2020; Shang et al., 2021). In practice, many educational contexts require multi-level assessment of knowledge attributes. For example, the *Full-time Compulsory Education Mathematics Curriculum Standards (Revised Draft)* uses four ordered categorical terms— "know (recognize)," "understand," "master," and "apply" — to describe different levels of knowledge and skill objectives. Polytomous at-

tributes enable finer-grained classification of students, making diagnostic tests with polytomous attributes both practically valuable and theoretically significant.

Researchers have developed diagnostic models specifically for polytomous attributes, including the OCAC-DINA model based on ordered-category attribute coding (Karelitz, 2004), polytomous extensions of the RRUM (Templin, 2004), LCDM (Templin & Bradshaw, 2004), and GDM (Haberman, von Davier, & Lee, 2008; von Davier, 2005). Zhan et al. (2020) developed higher-order diagnostic models for polytomous attributes, while Shang et al. (2021) defined continuous polytomous attributes and constructed corresponding diagnostic models drawing on multidimensional IRT. As in traditional CDA, the Q-matrix plays a critical role in polytomous-attribute CDA. Its accuracy directly affects model parameter identifiability, examinee classification, and overall test reliability and validity. Moreover, Q-matrices defined solely by experts are prone to errors and inconsistencies (de la Torre, 2008; Tu et al., 2012; DeCarlo, 2012; Liu et al., 2012; Yu et al., 2015; Yu & Cheng, 2020). Existing studies on polytomous-attribute Q-matrices have primarily relied on expert definition or simulation, typically assuming correctness without validating appropriateness. Objective methods for validating or estimating polytomous-attribute Q-matrices remain underexplored. This study extends objective Q-matrix validation and estimation methods from binary to polytomous attributes, aiming to advance the development of polytomous-attribute CDA.

2 Polytomous-Attribute Q-Matrix and Diagnostic Models

Before presenting estimation algorithms for polytomous-attribute Q-matrices, we first introduce the Q-matrix structure and corresponding diagnostic models.

2.1 Polytomous-Attribute Q-Matrix

For clarity, we define a *binary attribute* as one taking only values 0 and 1, and a *binary-attribute Q-matrix* (BQM) as a matrix composed solely of binary attributes, denoted by $\mathbf{Q}_b$. The corresponding CDA is denoted as BCDA. A *polytomous attribute* can take values $0, 1, 2, ...,$ and a *polytomous-attribute Q-matrix* (PQM) containing such attributes is denoted by $\mathbf{Q}_p$, with the corresponding CDA denoted as PCDA. $\mathbf{Q}_p$ is a $J \times K$ matrix where J and K represent the number of items and attributes, respectively. Unlike binary attributes, each element q_{jk} in $\mathbf{Q}_p$ has $L + 1$ possible levels with value space $\{0, 1, ..., L\}$.

Consider a simple polytomous-attribute Q-matrix example (Karelitz, 2004) with 4 items assessing 2 attributes, where both attributes have 5 levels: $\{0, 1, 2, 3, 4\}$. If attributes are dichotomized using 0 as the cutoff point (traditional binary approach), the corresponding binary Q-matrix would be as shown in Equation (2). For a test assessing K attributes, binary attributes can classify examinees into at most 2^K groups, whereas polytomous attributes (with $L + 1$ levels per attribute) can classify examinees into $(L+1)^K$ groups, where $(L+1)^K > 2^K$. For

example, with 2 attributes, binary classification yields 4 groups, while 5-level attributes yield 25 groups.

2.2 Diagnostic Models for Polytomous Attributes

Existing polytomous-attribute diagnostic models include OCAC-DINA (Karelitz, 2004), polytomous extensions of LCDM (Templin & Bradshaw, 2004), GDM (Haberman et al., 2008; von Davier, 2005), G-DINA framework extensions (Chen & de la Torre, 2013; Cai & Tu, 2015), higher-order models (Zhan et al., 2020), and continuous polytomous-attribute models (Shang et al., 2021). For brevity, we focus on the pG-DINA and p-DINA models relevant to this study.

The **pG-DINA** (polytomous generalized deterministic inputs, noisy, "and" gate) model is the polytomous extension of the G-DINA model (Chen & de la Torre, 2013). Assuming all attributes have the same number of levels L and following the notation of Chen and de la Torre (2013) and de la Torre (2011), let K_j^* denote the number of attributes measured by item j. For simplicity, assume item j measures the first K_j^* attributes. The required attribute vector can be expressed as a simplified vector $\mathbf{q}_j^* = (q_{j1}, q_{j2}, \ldots, q_{jK_j^*})$, where each element ranges from 0 to L. This simplification reduces the number of attribute vectors to consider from $(L+1)^K$ to $(L+1)^{K_j^*}$, improving computational efficiency.

Under the **p-DINA** model, each item classifies examinees into two groups: those who have mastered the item (having mastered all required attributes at levels no lower than those required) and those who have not. For item j, if $q_{jk} > 0$, the examinee's mastery status on attribute k can be compressed into a binary state: $\alpha_{jk}^* = I(\alpha_{jk} \geq q_{jk})$. This yields a compressed attribute mastery vector $\alpha_j^* = (\alpha_{j1}^*, \ldots, \alpha_{jK_j^*}^*)$, reducing the number of examinee parameters from $(L+1)^K$ to $2^{K_j^*}$. See Chen and de la Torre (2013, Table 2) for details.

In the saturated pG-DINA model, the probability of a correct response for an examinee with attribute vector α_j^* is:

$$P(X_j = 1|\alpha_j^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{jk}^* + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\alpha_{jk}^*\alpha_{jk'}^* + \cdots + \delta_{j12\ldots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{jk}^*$$

where δ_{j0} is the intercept (probability of correct response when no attributes are mastered), δ_{jk} are main effects, and higher-order terms represent interaction effects. The pG-DINA model reduces to p-DINA when only intercept and K_j^*-order interaction are considered, and to pA-CDM when only intercept and main effects are retained. Other models like p-DINO and pR-RUM can be obtained through parameter constraints. Due to its relative simplicity, this study uses the p-DINA model for polytomous-attribute Q-matrix estimation and validation.

3 Estimation Methods for Polytomous-Attribute Q-Matrix

Before introducing estimation methods, we briefly review binary-attribute Q-matrix estimation. Numerous studies have investigated Q-matrix validation and estimation in BCDA, including methods by de la Torre (2008), Tu et al. (2012), DeCarlo (2012), Liu et al. (2012), Xiang (2013), Chung (2014), Yu et al. (2015), de la Torre and Chiu (2016), Wang et al. (2020), and Yu and Cheng (2020). Among these, the S-statistic method is entirely data-driven, model-independent, and theoretically rigorous (Liu et al., 2013; Xu, 2013), offering excellent generalizability. This study extends the S statistic to polytomous-attribute Q-matrix estimation.

We consider two practical scenarios: (1) an expert-defined Q-matrix $\mathbf{Q}_0$ exists but may contain errors (i.e., some item attribute vectors are misspecified), requiring validation; and (2) only a small set of items have defined attribute vectors, with many "new items" requiring attribute definition. We denote the binary-attribute S statistic as S_b and its polytomous extension as S_p.

3.1 S-Statistic-Based Estimation for Polytomous-Attribute Q-Matrix

The core of the S statistic is the T-matrix, whose elements describe expected correct response probabilities for different ability groups on individual items or item combinations. The T-matrix links expected response distributions to model structure, reflecting Q-matrix definitions and establishing linear dependence between attribute and response distributions (Liu et al., 2012, 2013; Qin et al., 2015).

For a test assessing K attributes each with $L+1$ levels, examinees have $(L+1)^K$ possible attribute mastery patterns. The T-matrix has $(L+1)^K$ columns and rows corresponding to correct response probabilities on single items, item pairs, ⋯, and all J items combined, as shown in Equation (4). Row $\mathbf{t}_{12}$ represents the probability of correctly answering both items 1 and 2; columns represent all possible examinee classes.

Let π denote the population distribution of attribute patterns. Without prior information, π can be treated as uniform and updated using empirical Bayes methods (de la Torre, 2009). The expected response distribution τ on single items and their combinations is obtained via $\tau = \mathbf{T}\pi$, where τ_1 represents the expected probability of correctly answering item 1, calculated as in Equation (6). The observed response distribution $\hat{\tau}$ is derived from response data, with item parameters $\hat{\delta}$ estimated via EM algorithm (de la Torre, 2011) and examinee knowledge states $\hat{\alpha}$ via MAP algorithm (de la Torre, 2009).

When the Q-matrix is correctly specified and parameter estimation errors are small, the law of large numbers ensures that as sample size $N \to \infty$, $\hat{\tau}$ converges in probability to τ (Liu et al., 2012, 2013; Xu, 2013). With guessing and slipping present, fewer Q-matrix errors yield smaller distance between $\hat{\tau}$ and τ. Therefore, the objective function for polytomous-attribute Q-matrix estimation is:

$$\hat{\mathbf{Q}} = \arg \inf_{\mathbf{Q} \in \mathcal{Q}} \|\hat{\tau} - \mathbf{T}(\mathbf{Q})\pi\|$$

where $\mathbf{Q}$ is a candidate Q-matrix, $\hat{\mathbf{Q}}$ is the estimate, and "arg inf" denotes the Q-matrix minimizing the function across all possible Q-matrices.

3.2 Joint Estimation Algorithm (JE)

The JE algorithm starts from an expert-defined initial Q-matrix $\mathbf{Q}_0$ that may contain errors. Using $\mathbf{Q}_0$ as input, the algorithm produces estimates $\hat{\mathbf{Q}}$, $\hat{\delta}$, and $\hat{\alpha}$. Comparing $\hat{\mathbf{Q}}$ with the true $\mathbf{Q}$ assesses estimation success. The algorithm proceeds as follows:

1. Estimate item parameters $\hat{\delta}$ and examinee parameters $\hat{\alpha}$ using EM and MAP algorithms (Chen & de la Torre, 2013) based on $\mathbf{Q}_0$ and response data $\mathbf{X}$, then compute $\hat{\tau}$.

2. For each item j, while fixing other items, iterate through all possible attribute vectors $\mathbf{q}_j^*$ (from space $\mathcal{Q}_j$ with $(L+1)^{K_j^*}$ possibilities) to obtain $\mathbf{Q}_j^*$, estimate parameters, and compute the corresponding $\hat{\tau}$. Select the attribute vector minimizing the distance as item j' s estimate:

$$\hat{\mathbf{q}}_j = \arg \inf_{\mathbf{q}^{j*} \in \mathcal{Q}_j} \|\hat{\tau} - \mathbf{T}(\mathbf{Q}_j^*)\pi\|$$

3. After processing all items, one iteration completes, yielding $\mathbf{Q}^{(t)}$. If $\mathbf{Q}^{(t)}$ matches the true $\mathbf{Q}$, proceed to step 5; otherwise, increment iteration count and continue.

4. Set $\mathbf{Q}_0 = \mathbf{Q}^{(t)}$ and repeat step 2.

5. Terminate and output $\hat{\mathbf{Q}}$ and parameter estimates.

3.3 Online Estimation Algorithm (OE)

Unlike JE, which requires all items to have initial attribute definitions, OE only needs a small set of anchor items with known attributes. The remaining "new items" require attribute vector definition. Let $\mathbf{Q}_a$ denote anchor items and $\mathbf{Q}_n$ denote new items. The algorithm incrementally adds one new item at a time from $\mathbf{Q}_n$ to $\mathbf{Q}_a$, estimating its attribute vector before proceeding.

The OE algorithm proceeds as follows:

1. Randomly select one item from $\mathbf{Q}_n$ without replacement and add it to $\mathbf{Q}_a$, placing it in the first row.

2. Based on the expanded $\mathbf{Q}_a$, estimate item and examinee parameters from response data, then compute $\hat{\tau}$.

3. For the newly added item, iterate through all possible attribute vectors $\mathbf{q}_j^*$, estimate parameters, and compute $\hat{\tau}$. Select the attribute vector minimizing the distance as the item's estimate.

4. If $\mathbf{Q}_n$ is not empty, repeat step 1; otherwise proceed to step 5.

5. Terminate and output the estimated $\hat{\mathbf{Q}}$.

When the initial anchor set is correct and sufficiently large, this incremental approach avoids the "masking effect" (Fung, 1993; Yuan & Zhong, 2008) caused by introducing multiple erroneous items simultaneously. However, if anchor items contain errors or are too few, OE may produce incorrect estimates for some items, necessitating a "second-stage correction" using JE on the full Q-matrix.

4 Research Design

We evaluate the two algorithms' performance in recovering correct polytomous-attribute Q-matrices across various conditions via simulation studies. We assume an expert-defined initial Q-matrix containing few errors. Two error types are considered: (I) attribute level mis-specification (over- or under-estimation, excluding changes to/from 0), where the true value should be 2 but is specified as 1 or 3; and (II) both level mis-specification and inclusion/exclusion of measured attributes (e.g., incorrectly setting $q_{jk} > 0$ when it should be 0, or vice versa). Error Type II represents general test development scenarios, with Error Type I as a special case.

Given $(L + 1)^K$ possible attribute patterns, a sample of 500 examinees would yield only 2.06 examinees per pattern on average, which is insufficient. Therefore, the minimum sample size is set to 1,000.

4.1 Joint Estimation Algorithm Conditions

For JE, we manipulate four factors: number of items (15 or 30, following Chen & de la Torre, 2013), sample size (1,000, 2,000, or 4,000), number of misspecified items (3, 4, or 5, following Liu et al., 2012), and error type (I or II). This yields $2 \times 3 \times 3 \times 2 = 36$ experimental conditions.

4.2 Online Estimation Algorithm Conditions

For OE, we manipulate: number of items (15 or 30), sample size (1,000, 2,000, or 4,000), and number of anchor items. Following Qin et al. (2015, 2020), for 30-item tests we use 8–15 anchor items (8 levels), and for 15-item tests we use 5–10 anchor items (6 levels). This yields $2 \times 3 \times 8 + 2 \times 3 \times 6 = 84$ conditions.

4.3 Experimental Design Details

Q-matrix: True Q-matrices for 30-item ($\mathbf{Q}_{30}$) and 15-item ($\mathbf{Q}_{15}$) tests are provided in Appendix Tables A1 and A2 (Chen & de la Torre, 2013; Yu &

Cheng, 2020).

Item Parameters: Guessing and slipping parameters are simulated from uniform distributions.

Examinee Parameters: Attribute mastery patterns follow a uniform distribution.

Response Data: Generated using the p-DINA model based on true Q-matrices and parameters.

Initial Q-matrix: For JE, randomly select items from the true Q-matrix and modify their attribute vectors according to error types I or II (excluding all-zero vectors or correct values). For OE, randomly select anchor items; new items' initial attribute vectors are randomly generated (excluding zero vectors or correct vectors).

Parameter Estimation: Implemented in MATLAB with 100 replications per condition.

Evaluation Metrics: We use three criteria: (1) Q-matrix recovery rate—the proportion of 100 replications where the estimated Q-matrix exactly matches the true Q-matrix; (2) average iterations; and (3) average execution time. Recovery rate indicates estimation accuracy, while iterations and time reflect computational efficiency.

4.4 Study 1: Joint Estimation of Polytomous-Attribute Q-Matrix and Parameters

JE is suitable when experts have defined attributes for all items but some definitions are uncertain or disputed. We examine two error conditions.

Error Type I (Level Mis-specification Only): This simpler scenario occurs when experts disagree on attribute levels. We investigate JE' s performance when the initial Q-matrix contains only under- or over-estimated attribute levels (excluding changes to/from 0).

Error Type II (Level and Attribute Inclusion/Exclusion Errors): This more severe scenario involves both level mis-specification and incorrect inclusion/exclusion of attributes, which may occur in practice. Error Type I is a special case of Error Type II.

4.5 Study 2: Online Estimation of Polytomous-Attribute Q-Matrix and Parameters

OE is suitable when only a small set of items is correctly defined and many new items require attribute definition, such as when developing new test items. New items' attribute vectors can be randomly initialized. The algorithm leverages information from anchor items to define new items incrementally, fixing anchor attributes while estimating one new item at a time. After all new items are

processed, JE is applied to the full Q-matrix to improve accuracy and reduce negative effects from "masking." Estimation success is determined by comparing the final Q-matrix to the true Q-matrix.

4.6 Results

JE Algorithm Results: Tables 1–4 present JE results for 30- and 15-item tests under error types I and II. Performance is influenced by sample size, test length, and number of misspecified items. Simulations were run on cloud servers with dual Xeon E5-2697 CPUs, 64GB DDR5 RAM, and 512GB SSD. Despite reduced search spaces, execution times remain substantial (minimum ~24 hours). Test length dramatically affects accuracy: reducing items from 30 to 15 decreases success rates by an average of 61.67%.

Key Findings: Success rates increase with sample size and test length, but decrease with more misspecified items. For 30-item tests, success rates exceed 80% across conditions; for 15-item tests, maximum success rates are below 60%. Iterations increase with test length (average <2.5 for 15 items, >3 for 30 items). Error Type II yields slightly lower success rates and more iterations than Error Type I due to larger attribute vector search spaces.

OE Algorithm Results: Tables 5–6 show OE performance. Required anchor item counts vary by sample size: for Q_{30}, 1,000 examinees need 10 anchor items for 90% success, while 2,000–4,000 need only 8; for 95% success, 1,000–2,000 need $ $13 anchors, while 4,000 need 12. For Q_{15}, all sample sizes require $ $9 anchors for 80% success. OE performs better with 30 items than 15 items because longer tests improve attribute pattern estimation accuracy. Execution time decreases with more anchor items (e.g., for Q_{30} with 1,000 examinees: 8 anchors = 176,481.88s; 15 anchors = 23,545.31s), as most time is consumed by joint estimation iterations (1.78 vs. 0.22 average iterations).

5 Empirical Data Analysis

We applied both algorithms to data from a high school monthly mathematics exam focusing on probability. The test assessed four attributes: random events, sample space, classical probability, and frequency-based probability estimation. Each attribute had five ordered categories: unaware, aware, understand, master, and apply (coded 0–4). Twenty items were administered to 1,960 examinees.

Using the expert-defined initial Q-matrix (Table 7) as input, JE terminated after 4 iterations (more than in simulations, reflecting real-data complexity). The suggested Q-matrix (Appendix Table A3) modified 6 items involving 7 attributes, all representing level mis-specifications (Error Type I). Parameter estimation revealed 76 distinct attribute mastery patterns, indicating non-uniform distribution.

For OE, we selected 5 anchor items (highlighted in Appendix Table A4) with unanimous expert agreement and JE validation. The remaining 15 items were

treated as "new" and estimated incrementally. After OE completion, JE was applied to the full Q-matrix, yielding the suggested Q-matrix in Appendix Table A4. OE recommended modifying 6 items involving 6 attributes. Except for item 19, JE and OE produced identical suggestions. For item 19, the initial vector was $(2, 0, 0, 0)$; JE suggested $(3, 0, 0, 0)$ while OE suggested $(0, 0, 0, 2)$. After discussion with five practicing teachers, four favored OE's suggestion of changing attribute 4 from level 2 to level 3.

6 Discussion and Future Directions

This study extends S-statistic-based Q-matrix estimation from binary to polytomous attributes, enabling objective validation and estimation. Two algorithms —JE and OE—address different practical scenarios. Simulations show that despite larger search spaces for polytomous-attribute Q-matrices, both algorithms achieve high success rates under appropriate conditions.

However, several limitations warrant future research. For JE, we only considered cases with few errors; performance with more extensive misspecification and the maximum tolerable error rate require investigation. For OE, we randomly selected anchor items without considering their quality. Future research should explore optimal anchor item design, such as incorporating reachability matrices (Chen et al., 2015; Ding et al., 2019; Peng et al., 2016, 2018; Gu et al., 2018; Gu & Xu, 2021) to facilitate new item estimation. We also limited error types to two categories; other error patterns need examination. Realistic testing scenarios often involve multiple solution strategies (Huang et al., 2019) and attribute hierarchies (Yu et al., 2021); Q-matrix estimation under these complex conditions merits further study.

A notable limitation of S-statistic-based methods is computational time, which may hinder practical application. Future work should improve efficiency or develop faster alternatives. Yu and Cheng (2020) showed residual-based statistics outperform S statistics in binary CDA; extending residual-based methods to polytomous attributes is promising. Nonparametric methods requiring smaller samples and offering computational advantages (Liu et al., 2021) and deep learning approaches (Zhang et al., 2021; Li et al., 2022) also warrant exploration.

Empirical analysis suggests experts more commonly mis-specify attribute levels (over- or under-estimation) than omit or add attributes. A valuable byproduct of OE is simultaneous parameter estimation for new items on the same scale as anchor items. Future research should incorporate attribute relationships and apply these algorithms to other diagnostic models (Zhan et al., 2020).

References

Cai, Y., & Tu, D. B. (2015). Extension of cognitive diagnosis models based on the polytomous attributes framework and their Q-matrices designs. *Acta Psychologica Sinica*, 47(10), 1300–1308.

Chen, J. S., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. *Applied Psychological Measurement*, 37(6), 419-437.

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

Chung, M. T. (2014). *Estimating the Q-matrix for cognitive diagnosis models in a Bayesian framework* (Unpublished doctoral dissertation). Columbia University, New York.

de la Torre, J. (2009). DINA Model and Parameter Estimation: A Didactic. *Journal of Educational and Behavioral Statistics*, 34(1), 115–130.

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

DeCarlo, L. T. (2012). Recognizing Uncertainty in the Q-Matrix via a Bayesian Extension of the DINA Model. *Applied Psychological Measurement*, 36(6), 447–468.

Ding, S. L., Luo, F., Wang, W. Y., & Xiong, J. H. (2019). The designing cognitive diagnostic test with dichotomous scoring. *Journal of Jiangxi Normal University (Natural Science)*, 43(5), 441–447.

Fung, W. K. (1993). Unmasking outliers and leverage points: A confirmation. *Journal of the American Statistical Association*, 88(422), 515–519.

Gu, Y. Q., Liu, J. C., Xu, G. J., & Ying, Z. L. (2018). Hypothesis testing of the Q-matrix. *Psychometrika*, 83(3), 515–537.

Gu, Y. Q., & Xu, G. J. (2021). Sufficient and Necessary Conditions for the Identifiability of the Q-matrix. *Statistica Sinica*, 31, 449–472.

Haberman, S. J., von Davier, M., & Lee, Y.-H. (2008). Comparison of multidimensional item response models: Multivariate normal ability distributions versus multivariate polytomous ability distributions (ETS Research Report No. RR-08-45). Princeton, NJ: Educational Testing Service.

Huang, Y., Luo, F., Xiong, J. H., Ding, S. L., & Gan, D. W. (2019). The multiple-strategy cognitive diagnosis method with polytomous scoring. *Journal of Jiangxi Normal University (Natural Science)*, 43(4), 376–381.

Karelitz, T. M. (2004). *Ordered category attribute coding framework for cognitive assessments* (Unpublished doctoral dissertation). University of Illinois at Urbana–Champaign.

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, 41(3), 205–236.

Li, C. C., Ma, C. C., & Xu, G. J. (2022). Learning large Q-matrix by restricted Boltzmann machines. *Psychometrika.* https://doi.org/10.1007/s11336-021-09828-4.

Liu, J. C., Xu, G. J., & Ying, Z. L. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement*, 36(7), 548–564.

Liu, J. C., Xu, G. J., & Ying, Z. L. (2013). Theory of the self-learning Q-matrix. *Bernoulli*, 19(5A), 1790–1817.

Liu, N., Liu, X. L., Li, J. J., Zeng, P. F., Yu, X. J., & Kang, C. H. (2021). Constructing a non-parametric Q-matrix correction method based on Manhattan distance. *Journal of Jiangxi Normal University (Natural Science)*, 45(6), 634–641.

Ma, W., & Torre, J. (2019). An empirical Q-matrix validation method for the sequential generalized DINA model. *British Journal of Mathematical and Statistical Psychology*, 73(1), 142–163.

Peng, Y. F., Luo, Z. S., Li, Y. J., & Gao, C. L. (2018). Optimization of test design for examinees with different cognitive structures. *Acta Psychologica Sinica*, 50(1), 130–140.

Peng, Y. F., Luo, Z. S., Yu, X. F., Gao, C. L., & Li, Y. J. (2016). The optimization of test design in Cognitive Diagnostic Assessment. *Acta Psychologica Sinica*, 48(12), 1600–1611.

Qin, C. Y., Zhang, L., Qiu, D., Huang, L., Geng, T., Jiang, H., & Zhou, J. (2015). Model identification and Q-matrix incremental inference in cognitive diagnosis. *Knowledge-Based Systems*, 86, 66–76.

Qin, C. Y., Jia, S., Fang, X. W., & Yu, X. F. (2020). Relationship validation among items and attributes. *Journal of Statistical Computation and Simulation*, 90(18), 3360–3375.

Shang, Z. R., Erosheva, E. A., & Xu, G. J. (2021). Partial-mastery cognitive diagnosis models. *The Annals of Applied Statistics*, 15(3), 1529–1555.

Sun, J., Xin, T., & Zhang, S. (2013). The polytomous extension of the cognitive diagnosis model. *Acta Psychologica Sinica*, 45(10), 1095–1102.

Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, 30(2), 251–275.

Templin, J. L. (2005). *Generalized linear mixed proficiency models for cognitive diagnosis* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Tu, D. B., & Cai, Y. (2015). The development of CD-CAT with polytomous attributes. *Acta Psychologica Sinica*, 47(11), 1405–1414.

von Davier, M. (2005). A general diagnostic model applied to language testing data (ETS Research Report No. RR-05-16). Princeton, NJ: Educational Testing Service.

Wang, D. X., Cai, Y., & Tu, D. B. (2020). Q-matrix estimation methods for cognitive diagnosis models: Based on partial known Q-matrix. *Multivariate Behavioral Research*, 1–13.

Xiang, R. (2013). *Nonlinear penalized estimation of true Q-Matrix in cognitive diagnostic models* (Unpublished doctoral dissertation). Columbia University, New York.

Xu, G. J. (2013). *Statistical inference for diagnostic classification models* (Unpublished doctoral dissertation). Columbia University, New York.

Yu, X. F., & Cheng, Y. (2020). Data-driven Q matrix validation using a residual based statistic in cognitive diagnostic assessment. *British Journal of Mathematical and Statistical Psychology*, 73(1), 145–179.

Yu, X. F., Luo, Z. S., Gao, C. L., Li, Y. J., Wang, R., & Wang, Y. T. (2015). Joint estimation of model parameters and Q-matrix based on response data. *Acta Psychologica Sinica*, 47(2), 273–282.

Yu, X. F., Luo, Z. S., Gao, C. L., Li, Y. J., Wang, R., & Wang, Y. T. (2015). An item attribute specification method based on the likelihood D2 statistic. *Acta Psychologica Sinica*, 47(3), 417–426.

Yu, X. F., Ma, Y. F., Luo, Z. S., & Qin, C. Y. (2021). The attribute hierarchical structure learning based on K2 algorithm. *Journal of Jiangxi Normal University (Natural Science)*, 45(4), 376–383.

Yuan, K. H., & Zhong, X. (2008). Outliers, leverage observations, and influential cases in factor analysis: Using robust procedures to minimize their effect. *Sociological Methodology*, 38(1), 329–368.

Zhan, P. D., Bian, Y. F., & Wang, L. J. (2016). Factors affecting the classification accuracy of reparametrized diagnostic classification models for expert-defined polytomous attributes. *Acta Psychologica Sinica*, 48(3), 318–330.

Zhan, P. D., Wang, W., & Li, X. M. (2020). A partial mastery, higher-order latent structural model for polytomous attributes in cognitive diagnostic assessments. *Journal of Classification*, 37, 328–351.

Zhang, Y. L., Zhao, B., & Tao, J. H. (2021). The study on students' cognitive state based on fuzzy cognitive diagnostic framework. *Journal of Jiangxi Normal University (Natural Science)*, 45(5), 452–459.

Appendix

Table A1. Q-matrix for 30 items
Table A2. Q-matrix for 15 items

Table A3. Suggested Q-matrix for probability data from JE algorithm
Table A4. Suggested Q-matrix for probability data from OE algorithm

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

Source: ChinaXiv −Machine translation. Verify with original.