# A Unified Design Method for Minimal Complete Q-Matrix in Cognitive Diagnostic Testing

**Authors:** Tang Xiaojuan, Mao Mengmeng, Li Yu, Ding Shuliang, Peng Zhixia, Tang Xiaojuan, Li Yu

**Date:** 2025-09-08T10:33:46+00:00

## Abstract

Attribute levels (dichotomous attributes and polytomous attributes) and item ideal scoring methods (0-1 scoring and polytomous scoring) constitute two important dimensions in cognitive diagnostic test design. Polytomous attribute tests can provide more detailed diagnostic information, while polytomous scoring tests can enhance classification accuracy; however, existing cognitive diagnostic tests lack integrated designs for polytomous attributes and polytomous scoring. Drawing upon the concept of structured/unstructured simplest complete Q-matrix (SSCQM/USCQM) for dichotomous attributes with polytomous scoring, this paper proposes a unified simplest complete Q-matrix design methodology for cognitive diagnostic tests, addressing test design challenges across various combinations of attribute levels and item ideal scoring methods. Under both long-test and short-test conditions, and using the (quasi-)reachability matrix as a benchmark, simulation studies were conducted to compare the accuracy rates of various SSCQM and USCQM configurations. The results demonstrate that, overall, SSCQM and USCQM exhibit higher classification accuracy. Empirical data further corroborate the advantages of SSCQM and USCQM-based tests.

## Full Text

## A Unified Design Method for the Simplest Complete Q-Matrix in Cognitive Diagnostic Testing

**Tang Xiaojuan[1], Mao Mengmeng[2], Li Yu[3], Ding Shuliang[4], Peng Zhixia[5]**
([1] School of Education, Jiangxi Normal University, Nanchang 330022, China)
([2] School of Public Policy and Administration, Nanchang University, Nanchang 330036, China)
([3] Mental Health Education Center, School of Marxism, Zhejiang Gongshang

University, Hangzhou 310018, China)
([4] College of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022, China)
([5] School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, China)

## Abstract

Attribute levels (dichotomous and polytomous attributes) and item scoring schemes (0-1 scoring versus polytomous scoring) represent two critical dimensions in cognitive diagnostic test design. Polytomous attribute tests provide more detailed diagnostic information, while polytomously scored tests yield higher classification accuracy. However, existing cognitive diagnostic tests lack integrated designs that combine polytomous attributes with polytomous scoring. Drawing upon concepts of structured/unstructured simplest complete Q-matrices (SSCQM/USCQM) for dichotomous attributes with polytomous scoring, this paper proposes a unified design method for the simplest complete Q-matrix in cognitive diagnostic testing. This approach addresses test design challenges across various combinations of attribute levels and scoring schemes, using (quasi-)reachability matrices as benchmarks. Through simulation studies under both long and short test conditions, we compare the classification accuracy of various SSCQM and USCQM designs. Results demonstrate that SSCQM and USCQM generally achieve higher classification accuracy, with empirical data further validating the advantages of these test designs.

**Keywords:** cognitive diagnostic testing, test design, simplest complete Q-matrix, unified design method

## 1 Introduction

Educational assessment fundamentally shapes educational development, serving as the guiding compass for instructional direction. The *Overall Plan for Deepening Education Evaluation Reform in the New Era* issued by the Central Committee of the Communist Party of China and the State Council reflects the government's emphasis on formative assessment while imposing higher demands on the diagnostic functions of educational evaluation. Multiple intelligences theory similarly advocates strengthening diagnostic capabilities to provide empirical foundations for teachers to guide students. Cognitive diagnostic testing (CDT), grounded in cognitive psychology theory, offers more granular diagnostic information than standardized tests [?]. Since CDT quality directly impacts the precision of diagnostic information and subsequently its remedial functions, test design plays a pivotal role in the cognitive diagnostic process. Gorin [?] proposed that an effective test should elicit specific behaviors from examinees while remaining diagnostically tractable. Consequently, ideal test designs can trigger differentiated response patterns across different knowledge states (KS) by establishing one-to-one correspondences between knowledge states and ideal response patterns (IRP) or observed response patterns (ORP), thereby better distinguishing among examinees. Research indicates that under certain conditions, (quasi-)reachability matrices facilitate establishing these one-to-one correspondences [?, ?, ?]. However, when the number of attributes (or attribute levels) is large, (quasi-)reachability matrices contain numerous columns, making them impractical for short-test scenarios. A key challenge in formative assessment is effectively distinguishing examinees through minimal test length. In practice, CDT applications involve complex combinations of attribute levels and scoring schemes, yet relevant design research remains fragmented and lacks integrated methodologies, hindering both implementation and theoretical advancement.

Attribute levels and item scoring schemes constitute two essential dimensions of CDT design. The Q-matrix represents CDT, where row $k$ denotes attribute $k$ and column $\mathbf{q}_j$ represents item $j$. Attributes describe knowledge or skills, while items are represented by attribute vectors. Most existing research addresses dichotomous attributes (levels 0 or 1), where 0 indicates an item does not assess an attribute and 1 indicates it does, yielding a binary Q-matrix or Boolean matrix. As CDT theory and applications have evolved, dichotomous-attribute items have proven limited in precisely describing attribute mastery levels. Polytomous-attribute items provide more detailed diagnostic information about examinees' attribute mastery levels, leading to polytomous Q-matrices. For $K$ attributes, dichotomous-attribute knowledge states (also represented as attribute vectors) number $2^K$, while polytomous-attribute knowledge states number $\prod M_k$ (where $M_k$ represents the number of levels for attribute $k$, $k = 1, 2, \dots, K$, with some attributes having $M_k \geq 2$). Clearly, polytomous attributes enable finer-grained classification. Researchers have developed numerous polytomous-attribute cognitive diagnostic models (CDM), including five models mentioned by Chen and de la Torre [?], plus GDD-P [?], PA-rRUM and PA-DINA [?], RPa-DINA, RPa-DINO and RPa-LLM [?], and GRPa-DINA [?].

Item scoring schemes can extend beyond 0-1 scoring to polytomous scoring, which provides more detailed diagnostic information [?, ?]. Most polytomously scored CDMs extend from 0-1 scoring models [?].

Based on different combinations of attribute levels and scoring schemes, CDT can be categorized into four types: dichotomous-attribute 0-1 scoring, polytomous-attribute 0-1 scoring, dichotomous-attribute polytomous scoring, and polytomous-attribute polytomous scoring. These test types exhibit substantial differences in classification accuracy. Although polytomous-attribute 0-1 scoring tests provide richer diagnostic information, their classification accuracy is lower than dichotomous-attribute 0-1 scoring tests [?]. Dichotomous-attribute polytomous scoring tests achieve higher classification accuracy than their 0-1 scoring counterparts, particularly when attribute counts are large [?], but offer less diagnostic richness than polytomous-attribute tests. Polytomous-attribute polytomous scoring tests not only cover diverse educational assessment scenarios but also yield rich and precise diagnostic results, demonstrating broad application prospects [?]. Existing research primarily focuses on dichotomous-attribute 0-1 scoring, constructing tests through item structure [?, ?, ?, ?, ?] and test assembly indices [?, ?, ?, ?, ?]. Research on other test types is scarce or nonexistent, let alone unified methods applicable to all four CDT types. Specifically, the multiple attribute levels in polytomous-attribute tests and diverse scoring schemes in polytomously scored tests complicate test design. Despite limited research, test design' s importance and practical necessity make this a critical problem to solve. A complete Q-matrix can identify all examinees, whereas an incomplete Q-matrix misclassifies them [?]. Therefore, Q-matrix completeness is essential for CDT design [?, ?, ?, ?, ?]. Building upon complete Q-matrices, this study proposes a unified design method for the simplest complete Q-matrix applicable to various combinations of attribute levels (dichotomous and polytomous) and scoring schemes (0-1 and polytomous), providing theoretical support for addressing test design challenges.

## 2 Complete Q-Matrix

As a high-quality test Q-matrix capable of identifying all examinees, the complete Q-matrix can be applied to various test types, with (quasi-)reachability matrices being widely used [?, ?]. For dichotomous attributes with $K$ attributes, the reachability matrix (generally denoted as $R$) is a $K$-order matrix. For polytomous attributes, the matrix with this property is called the quasi-reachability matrix (generally denoted as $R_P$). Using three attributes as an example (Figure 1 [Figure 1: see original paper]):

**Example 1:** The reachability matrix for Figure 1(a) is $R = ()$. For Figure 1(b), based on the dichotomous-attribute reachability matrix $R$, columns are added on the main diagonal for attributes where the $K$-th attribute level count exceeds 1, yielding the polytomous-attribute quasi-reachability matrix. For instance, if attributes $A_1$ to $A_3$ have level counts of 2, 3, and 4 respectively, the quasi-reachability matrix is $R_P = ()$.

Current research on complete Q-matrices primarily focuses on two design perspectives: (1) ideal response patterns, and (2) cognitive diagnostic models. Regarding ideal response patterns, Ding et al. [?] consider the relationship among knowledge states, ideal response patterns, and observed response patterns as the core of cognitive diagnosis. Ideal response patterns are unaffected by item properties, motivation, or random factors, whereas observed response patterns are. Under equivalent conditions, complete Q-matrices diagnose ideal (or observed) response patterns with higher accuracy than incomplete Q-matrices. Based on ideal response patterns, Ding and Luo [?] proposed that for non-compensatory attributes, a 0-1 scoring test Q-matrix containing the reachability matrix as a submatrix is complete. Ding et al. [?, ?] developed construction methods and proofs for polytomous scoring test complete Q-matrices for several basic attribute hierarchy structures. Tang et al. [?] further proposed a construction method for the simplest complete Q-matrix (SCQM) applicable to various attribute hierarchy structures in polytomous scoring tests.

Regarding cognitive diagnostic models, some scholars argue that completeness is model-dependent—a Q-matrix complete for one model may not be complete for another [?]. For example, Köhn and Chiu [?, ?] proposed that Q-matrices complete for the DINA model are matrices between the reachability matrix and identity matrix. For DINA, DINO, GPDINA, sequential DINA, and RegLCMs models, a necessary and/or sufficient condition for a complete Q-matrix is that the test contains an identity matrix [?, ?, ?, ?, ?, ?, ?]. The distinction and connection between these two perspectives: examining test completeness based on ideal response patterns actually underpins model-based completeness assessment and allows evaluating test blueprint quality before implementation. The former requires only a one-to-one correspondence between knowledge states and ideal response patterns without considering random error, while the latter incorporates random error (requiring correspondence between knowledge states and observed response patterns). When item scoring schemes and maximum scores are identical and no random error exists, both approaches yield identical results. With minimal random error (e.g., $s, g \sim U(0, t)$ where $0 < t < 0.1$), estimated knowledge states differ negligibly, though this demands meticulously crafted items, high motivation, and ideal testing conditions.

Based on whether all items in the Q-matrix fully conform to attribute hierarchy structures, complete Q-matrices are classified as structured complete Q-matrices or unstructured complete Q-matrices. Existing research is scattered across different combinations of attribute levels and scoring schemes, lacking an integrated framework. First, for dichotomous-attribute 0-1 scoring tests, the reachability matrix serves as the structured complete Q-matrix, while the identity matrix is the unstructured complete Q-matrix (except for independent structures), with matrices between them remaining unstructured [?]. Furthermore, Ding et al. [?] demonstrated that test Q-matrices containing Köhn and Chiu' s [?] unstructured complete Q-matrix submatrices retain completeness. Second, for dichotomous-attribute polytomous scoring tests, Ding and Luo [?] and Ding et al. [?] proposed structured complete Q-matrices for several basic attribute hierarchy structures. Tang et al. [?] extracted structured simplest complete Q-matrices (SSCQM) from reachability matrices and extended Köhn and Chiu' s [?] method to construct unstructured simplest complete Q-matrices (USCQM) for polytomous scoring. Third, for polytomous-attribute 0-1 scoring tests, the structured complete Q-matrix is the quasi-reachability matrix [?, ?], though research on other structured and unstructured complete Q-matrices remains scarce. Finally, for polytomous-attribute polytomous scoring tests, the structured complete Q-matrix is the quasi-reachability matrix or matrices formed by merging its columns [?]. While detailed research on unstructured complete Q-matrices is lacking, some scholars mention they can comprise items satisfying T-matrix conditions [?, ?]. The T-matrix, an alternative Q-matrix representation, describes the linear dependency between attribute distributions and observed response distributions (see Liu et al. [?]).

In summary, Q-matrix completeness design methods in CDT exhibit spontaneous and fragmented characteristics, lacking a unified perspective. Notably, the simplest complete Q-matrix [?] is the complete Q-matrix with the fewest items in the reachability matrix, offering advantages of brevity and strong classification power, particularly in short tests. This holds significant practical value under the "double reduction" policy and warrants investigation. In fact, the reachability matrix is the simplest complete Q-matrix for dichotomous-attribute 0-1 scoring tests, while the quasi-reachability matrix serves this role for polytomous-attribute 0-1 scoring tests. However, only one such simplest complete Q-matrix exists, compromising test security. For more complex scenarios, Tang et al. [?] proposed SSCQM and USCQM for dichotomous-attribute polytomous scoring tests, where the design methods yield non-unique simplest complete Q-matrices, enhancing test security. Yet for the most complex case—polytomous-attribute polytomous scoring tests—corresponding simplest complete Q-matrix research remains absent. The three simpler test types can all be viewed as special cases of polytomous-attribute polytomous scoring tests. Therefore, studying the simplest complete Q-matrix design for polytomous-attribute polytomous scoring tests is theoretically significant from both classification capability and test security perspectives. This study proposes a unified design method for the simplest complete Q-matrix applicable to all CDT types from the perspective of item attribute total scores. Based on the two aforementioned design perspectives and item structures, this research comprises: (1) proposing unified design methods for structured and unstructured simplest complete Q-matrices based on ideal response patterns; (2) examining whether completeness changes when combining these matrices with cognitive diagnostic models; and (3) validating the classification capability of simplest complete Q-matrices through simulation and empirical studies.

## 3.1 Cognitive Diagnostic Test Design Elements

Cognitive diagnostic test design involves attributes, attribute hierarchy structures, item scoring schemes, and item maximum scores. This study examines five basic attribute hierarchy structures: linear, convergent, divergent, unstructured, and independent structures. Other hierarchy structures can be formed by combining these basic types. Regarding attributes, we investigate both dichotomous and polytomous attributes. Item scoring schemes include 0-1 and polytomous scoring. The proposed scoring scheme satisfies the attribute-score correspondence assumption [?] with equal attribute weights (relative importance of attributes in overall evaluation).

**(1) The proposed item ideal scoring function $\eta_{ij}$ is defined as:**

$$\eta_{ij} = \begin{cases} (I\{\alpha_i \geq \mathbf{q}_j\})^{\gamma}, & \text{when } \gamma = 1 \\ \sum_{k=1}^{K} \beta_k I\{\alpha_{ik} \geq q_{kj}\} q_{kj}, & \text{when } \gamma = 0 \end{cases}$$

where examinee $i$ has attribute vector $\alpha_i = \{\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{iK}\}$ and item $j$ has attribute vector $\mathbf{q}_j = \{q_{1j}, q_{2j}, \cdots, q_{Kj}\}$. When $\gamma = 1$, the scoring is 0-1: if $\forall k \leq K, \alpha_{ik} \geq q_{kj}$, the score is 1, otherwise 0; each item' s maximum score is

$m_j = 1$. When $\gamma = 0$, scoring is polytomous: if $\alpha_{ik} \geq q_{kj}$, examinee $\alpha_i$ gains $\beta_k q_{kj}$ points on item $j$ (for both dichotomous and polytomous attributes). Each item' s maximum score is:

$$m_j = \sum_{k=1}^{K} \beta_k q_{kj}$$

Specifically, when attribute weights $\beta_k = 1$, the examinee' s ideal score increases by $q_{kj}$ for each additional mastered attribute $k$ in item $j$.

**(2) Item Attribute Total Score (IATS)**

The unified design method is based on the Item Attribute Total Score concept, defined as:

$$IATS_j = \sum_{k=1}^{K} \beta_k q_{kj}$$

For 0-1 scoring with $\beta_k = 1$, $IATS_j$ equals 1 divided by the number of attributes assessed by the item. For polytomous scoring using the proposed ideal scoring and maximum scores, $IATS_j = 1$ for all items. In practice, the ratio of average examinee score to $IATS_j$ reflects item attribute difficulty.

**(3) Vector-Related Definitions**

When selecting columns (vectors) from (quasi-)reachability matrices, comparison requires these concepts:

**Definition 1:** For two $K$-dimensional vectors $\mathbf{X} = \{x_1, x_2, \cdots, x_K\}^T$ and $\mathbf{Y} = \{y_1, y_2, \cdots, y_K\}^T$, if $x_k \leq y_k$ $(\forall k \leq K)$, then $\mathbf{X} \leq \mathbf{Y}$, making $\mathbf{X}$ and $\mathbf{Y}$ comparable, with $\mathbf{Y}$ greater than or equal to $\mathbf{X}$. $\mathbf{X} = \mathbf{Y}$ if and only if $x_k = y_k$ $(\forall k \leq K)$, where "$\leq$" is a partial order relation. If no partial order relation exists, $\mathbf{X}$ and $\mathbf{Y}$ are incomparable.

**Definition 2:** Among all comparable different vectors, if there exists $\mathbf{Y}$ such that $\mathbf{X} \leq \mathbf{Y}$ for all $\mathbf{X}$, then $\mathbf{Y}$ is called the maximum column.

**Example 2:** Given matrix $R = (\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3) = ()$, since $q_{k1} \leq q_{k2}$ $(\forall k \leq 3)$, $\mathbf{q}_1$ and $\mathbf{q}_2$ are comparable, with $\mathbf{q}_2$ greater than $\mathbf{q}_1$. Similarly, $\mathbf{q}_1 \leq \mathbf{q}_3$, making $\mathbf{q}_3$ greater than $\mathbf{q}_1$. However, because $q_{22} \geq q_{23}$ while $q_{32} \leq q_{33}$, $\mathbf{q}_2$ and $\mathbf{q}_3$ are incomparable. Based on comparability, $R$ can be partitioned into two groups where vectors within each group are comparable and vectors across groups are incomparable. For instance, $(\mathbf{q}_1, \mathbf{q}_2)$ and $\mathbf{q}_3$, where $\mathbf{q}_2$ is the maximum column of $(\mathbf{q}_1, \mathbf{q}_2)$. Alternatively, $R$ can be partitioned as $(\mathbf{q}_1, \mathbf{q}_3)$ and $\mathbf{q}_2$, with $\mathbf{q}_3$ as the maximum column of $(\mathbf{q}_1, \mathbf{q}_3)$.

The following sections present unified design methods for structured simplest complete Q-matrices (SSCQM) and unstructured simplest complete Q-matrices (USCQM) across all CDT types.

## 3.2 Unified Design Method for Structured Simplest Complete Q-Matrix (SSCQM)

Unlike previous SSCQM design methods [?], this study proposes a unified SSCQM design method based on item attribute total scores.

### 3.2.1 SSCQM Unified Design Procedure

The SSCQM unified design method for cognitive diagnostic testing proceeds as follows:

**Step 1:** Partition comparable column vectors in the dichotomous-attribute reachability matrix into one or multiple groups based on partial order relations.

**Step 2:** Retain the maximum column from each group to achieve Q-matrix minimization while maintaining completeness.

**Step 3:** Within each group, retain columns with item attribute total scores different from the maximum column, then combine these with the retained maximum columns to form the dichotomous-attribute SSCQM.

**Step 4:** For polytomous-attribute tests, add all columns where attribute level counts exceed 1 to the dichotomous-attribute SSCQM.

### 3.2.2 SSCQM Design Example

SSCQM design applies to all attribute hierarchy structures. Using a three-attribute divergent structure (Figure 1) with attribute weights $\beta_k = 1$, and polytomous attributes $A_1$ to $A_3$ having level counts of 2, 3, and 4 respectively, we illustrate the unified SSCQM design method.

**Example 3 (continuing Examples 1 and 2):** The divergent structure' s reachability matrix is $R = (\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$. First partition: $(\mathbf{q}_1, \mathbf{q}_2)$ and $\mathbf{q}_3$, where $\mathbf{q}_2$ is the maximum column of $(\mathbf{q}_1, \mathbf{q}_2)$. Since $\mathbf{q}_3$ has only one column, it is deemed that group' s maximum column. Second partition: $(\mathbf{q}_1, \mathbf{q}_3)$ and $\mathbf{q}_2$, with maximum columns $\mathbf{q}_3$ and $\mathbf{q}_2$ respectively.

**(1) 0-1 Scoring Tests**

**Dichotomous-attribute 0-1 scoring test:** With $\gamma = 1$ and maximum item scores $m_j = \sum \beta_k q_{kj} = 1$, each attribute level is $q_{kj} = 0$ or 1. For partition $(\mathbf{q}_1, \mathbf{q}_2)$ and $\mathbf{q}_3$, retain maximum columns $\mathbf{q}_2$ and $\mathbf{q}_3$. $\mathbf{q}_1$ and $\mathbf{q}_2$ assess 1 and 2 attributes respectively. $\mathbf{q}_1$' s maximum score $m_1 = 1$ with $q_{11} = 1, q_{21} = q_{31} = 0$, giving $IATS_1 = m_1/(\beta_1 q_{11}) = 1$. $\mathbf{q}_2$' s maximum score $m_2 = 1$ with $q_{12} = q_{22} = 1, q_{32} = 0$, giving $IATS_2 = m_2/(\beta_1 q_{11} + \beta_2 q_{12}) = 1/2$. Since $IATS_1 \neq IATS_2$, retain $\mathbf{q}_1$ and combine with maximum columns $\mathbf{q}_2$ and $\mathbf{q}_3$, yielding the dichotomous-attribute 0-1 scoring SSCQM as the reachability matrix $R = (\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3)$. Similarly, partition $(\mathbf{q}_1, \mathbf{q}_3)$ and $\mathbf{q}_2$ yields the same SSCQM.

**Polytomous-attribute 0-1 scoring test:** For polytomous items, each attribute has level $q_{kj}$. Starting from the dichotomous-attribute SSCQM, add columns on the main diagonal for attributes where level counts exceed 1. With $A_2$ having level 2 and $A_3$ having levels 2 and 3, add corresponding column vectors to obtain the polytomous-attribute 0-1 scoring SSCQM—the quasi-reachability matrix $R_P$ in Figure 1(b).

**(2) Polytomous Scoring Tests**

**Dichotomous-attribute polytomous scoring test:** With $\gamma = 0$ and maximum item scores $m_j = \sum \beta_k q_{kj}$. For partition $(\mathbf{q}_1, \mathbf{q}_2)$ and $\mathbf{q}_3$, retain maximum columns $\mathbf{q}_2$ and $\mathbf{q}_3$. Under the proposed scoring scheme, all items have $IATS = 1$. Since $\mathbf{q}_1$ and $\mathbf{q}_2$ share identical $IATS$, $\mathbf{q}_1$ is not retained. Combining maximum columns $\mathbf{q}_2$ and $\mathbf{q}_3$ yields the dichotomous-attribute polytomous scoring SSCQM (denoted $Q_{RS}$) as $Q_{RS} = ()$. The alternative partition yields identical results.

**Polytomous-attribute polytomous scoring test:** Adding all columns where attribute level counts exceed 1 to the dichotomous-attribute polytomous scoring SSCQM (as in (1) ) yields the polytomous-attribute polytomous scoring SSCQM (denoted $Q_{RPS}$) as $Q_{RPS} = ()$.

Changing the scoring scheme or maximum scores alters column retention in Step 3, modifying SSCQM composition. Adding columns where attribute level counts exceed 1 to dichotomous-attribute SSCQM yields polytomous-attribute SSCQM. This $IATS$-based design method covers all CDT types across all attribute hierarchy structures, including dichotomous-attribute 0-1 scoring, dichotomous-attribute polytomous scoring, polytomous-attribute 0-1 scoring, and polytomous-attribute polytomous scoring tests, demonstrating broad applicability. Theoretical proofs for SSCQM appear in the Appendix. Multiple partitioning possibilities generate numerous SSCQMs (especially with many attributes and levels), allowing selection of different SSCQMs in practice to reduce exposure rates and enhance test security.

## 3.3 Unified Design Method for Unstructured Simplest Complete Q-Matrix (USCQM)

Köhn and Chiu [?] demonstrated that unstructured complete Q-matrices for dichotomous-attribute 0-1 scoring tests lie between the reachability matrix (upper bound) and identity matrix (lower bound). Tang et al. [?] extended this to construct USCQM for dichotomous-attribute polytomous scoring tests: using SSCQM as the upper bound with columns corresponding to reachability matrix $R$ at positions $(j_1, j_2, \cdots, j_t)$, extract columns at these positions from identity matrix $E$ to form lower bound $E'$, ensuring each USCQM row contains at least one 1 (representing all attributes being assessed). Using Figure 1' s divergent structure, Example 3' s dichotomous-attribute polytomous scoring SSCQM is $Q_{RS} = ()$ corresponding to reachability matrix $R$ columns (2, 3). Extracting these columns from the same-order identity matrix yields $E' = ()$. The dichotomous-attribute polytomous scoring USCQM (denoted $Q_{RU1}$) satisfies $E' < Q_{RU1} < Q_{RS}$, with each row containing at least one 1, such as $Q_{RU1} = ()$.

The unified USCQM design method (Figure 2 [Figure 2: see original paper]) is: obtain the upper bound SSCQM (i.e., (quasi-)reachability matrix or its complete submatrix) from Section 3.2, identify its corresponding (quasi-)reachability matrix column positions $(j_1, j_2, \cdots, j_t)$, extract these columns from the same-order (quasi-)identity matrix $E$ ($E_P$) to form lower bound $E'$ ($E'_P$), with USCQM lying between $E'$ ($E'_P$) and SSCQM, and ensuring each USCQM row contains at least one 1. Using Figure 1(b)' s divergent structure:

**(1) 0-1 Scoring Tests**

**Dichotomous-attribute 0-1 scoring test:** From Section 3.2(1) , the upper bound SSCQM is reachability matrix $R$, lower bound is same-order identity matrix $E$, giving USCQM (denoted $Q_{RU2}$) satisfying $E < Q_{RU2} < R$, such as $Q_{RU2} = ()$—the Köhn and Chiu [?] method.

**Polytomous-attribute 0-1 scoring test:** From Section 3.2(1) , upper bound SSCQM is quasi-reachability matrix $R_P$, lower bound is same-order quasi-identity matrix $E_P = ()$, giving USCQM (denoted $Q_{RPU1}$) satisfying $E_P < Q_{RPU1} < R_P$, such as $Q_{RPU1} = ()$.

**(2) Polytomous Scoring Tests**

**Dichotomous-attribute polytomous scoring test:** As described in Section 3.3' s first paragraph.

**Polytomous-attribute polytomous scoring test:** From Section 3.2(2) , upper bound SSCQM is $Q_{RPS}$, corresponding to quasi-reachability matrix columns (2,3,4,5,6). Extracting these columns from the quasi-identity matrix yields $E'_P = ()$, giving USCQM (denoted $Q_{RPU2}$) satisfying $E'_P < Q_{RPU2} < Q_{RPS}$, with each row containing at least one 1, such as $Q_{RPU2} = ()$.

Notably, for independent structures, SSCQM equals the (quasi-)identity matrix where upper and lower bounds coincide, so no USCQM exists. Both bounds for 0-1 scoring USCQM are complete, yielding complete USCQMs. However, polytomous scoring USCQM lower bounds are submatrices of identity or quasi-identity matrices that lack completeness, so resulting unstructured Q-matrices may not be complete and require completeness verification. This method generates numerous USCQMs, enriching test diversity.

## 4 Integrating Simplest Complete Q-Matrices with Cognitive Diagnostic Models

As previously discussed, based on the two design perspectives, to achieve high classification accuracy in practice, we further integrate SSCQM and USCQM with cognitive diagnostic models to examine their classification capability. For dichotomous-attribute 0-1 scoring tests, Köhn and Chiu [?] used the DINA model for completeness verification. For dichotomous-attribute polytomous scoring tests, Tang [?] used a modified RP-DINA model [?], which we won't reiterate. Below, we apply SSCQM and USCQM from Section 3 to corresponding models for polytomous-attribute 0-1 scoring and polytomous-attribute polytomous scoring tests.

## 4.1 Polytomous-Attribute 0-1 Scoring Cognitive Diagnostic Model

Chen and de la Torre [?] extended the DINA model to the Pa-DINA model, also called RPa-DINA [?], with probability model:

$$P_{ij} = P(Y_{ij} = 1|\alpha_i) = (1 - s_j - g_j)\eta_{ij} + g_j$$

$$\eta_{ij} = \prod_{k=1}^{K} \omega_{ijk}^{q_{kj}{}^*}$$

$$\omega_{ijk} = I\{\alpha_{ik} \geq q_{kj}\}$$

$$q_{kj}^* = I\{q_{kj} > 0\}$$

where $P_{ij}$ is the probability of examinee $i$ correctly answering item $j$; $s_j$ and $g_j$ are slipping and guessing parameters; $K$ is the total number of attributes; $q_{kj}$ are Q-matrix elements; $\alpha_{ik}$ is examinee $i$' s mastery of attribute $k$; $\eta_{ij}$ is the ideal response; $q_{kj}^*$ are elements of the collapsed Q-matrix (collapsing Q-matrix elements greater than 0 to 1 to indicate whether an attribute is assessed); and $\omega_{ijk}$ represents the latent response of examinee $i$ to attribute $k$ in item $j$ [?, ?, ?, ?].

## 4.2 Polytomous-Attribute Polytomous Scoring Cognitive Diagnostic Model

Current research on polytomous-attribute polytomous scoring CDMs is limited, primarily including GDD-P [?] and GRPa-DINA [?]. Since the DINA model is widely used and has been extended to polytomous attributes and polytomous scoring, this study focuses on the extended GRPa-DINA model.

### 4.2.1 RP-DINA Model

The P-DINA model extends DINA to polytomous scoring [?, ?]:

$$P(Y_{ij} = t|\alpha_i) = P^*(Y_{ij} \geq t|\alpha_i) - P^*(Y_{ij} \geq t + 1|\alpha_i)$$

$$P^*(Y_{ij} \geq t|\alpha_i) = (1 - s_{jt})^{\eta_{ij}} g_{jt}^{1-\eta_{ij}}$$

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{kj}}$$

Since P-DINA's ideal scoring only yields 0 and 1, limiting classification precision, Cai et al. [?] modified P-DINA's scoring scheme, proposing the RP-DINA model:

$$P^*(Y_{ij} \geq t|\alpha_i) = (1 - s_{jt})^{\delta_{ijt}} g_{jt}^{1-\delta_{ijt}}$$

$$\delta_{ijt} = \begin{cases} 1, & \text{if } \eta_{ij} \geq t \\ 0, & \text{if } \eta_{ij} < t \end{cases}$$

where $\eta_{ij} = \text{fix}[\alpha_i' \times \mathbf{q}_j / m_j]$, fix is an integer function, and $m_j$ is item $j$' s maximum score. Tang et al. [?] noted this scoring might alter Q-matrix completeness, so $\eta_{ij}$ was changed to ideal scoring, i.e., formula (1).

### 4.2.2 GRPa-DINA Model

Wang et al. [?] extended RPa-DINA to polytomous scoring using cumulative category response functions, proposing the GRPa-DINA model:

$$P(Y_{ij} = t|\alpha_i) = P^*(Y_{ij} \geq t|\alpha_i) - P^*(Y_{ij} \geq t + 1|\alpha_i)$$

$$P^*(Y_{ij} \geq t|\alpha_i) = (1 - s_{jt} - g_{jt})\eta_{ij} + g_{jt}$$

where $\eta_{ij}$, $\omega_{ijk}$, and $q_{kj}^*$ are as in equations (5), (6), and (7). Since $\eta_{ij}$ in (14) may prevent reachability matrices from being complete [?], this study uses the modified GRPa-DINA model with $\delta_{ijt}$ from (12).

## 4.3 Integration of SSCQM and USCQM with Cognitive Diagnostic Models

Köhn and Chiu [?] proposed that if $S(\alpha) = S(\alpha') \to \alpha = \alpha'$, then the Q-matrix is complete, where $S(\alpha) = E(\mathbf{Y}|\alpha)$ represents the expected observed response pattern $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_J)$ for examinee $\alpha \in KS$ across all items, and $S_j(\alpha) = E(Y_j|\alpha) = \sum_{t=0}^{m_j} tP(Y_j = t|\alpha)$ $(j \in \{1, 2, \cdots, J\})$ is the expected response on item $j$. The $S(\alpha)$ values for polytomous-attribute 0-1 scoring SSCQM and USCQM applied to RPa-DINA appear in Appendix Table 1 , while those for polytomous-attribute polytomous scoring SSCQM and USCQM applied to modified GRPa-DINA appear in Appendix Tables 2 and 3. These tables show that for both

SSCQM and USCQM, all $S(\alpha)$ values differ, indicating that completeness is preserved when integrated with cognitive diagnostic models.

## 5 Simulation Studies

Generally, summative assessments use long tests while formative assessments use short tests, imposing higher demands on CDT. Theoretically, SSCQM and USCQM offer completeness with fewer items, and long tests containing them as submatrices also achieve completeness. Their validity for both assessment types warrants investigation. Classification accuracy serves as a key validity indicator for CDT [?]. Simulation Studies 1-4 examine how attribute hierarchy structure, attribute count, attribute levels, and complete Q-matrix count affect SSCQM and USCQM classification accuracy under long and short test conditions, comparing them against (quasi-)reachability and incomplete Q-matrices. Since dichotomous-attribute tests can be viewed as polytomous-attribute tests with level count 2, simulations focus on polytomous-attribute 0-1 scoring and polytomous-attribute polytomous scoring tests.

Let $H$, $K$, $L$, and $M$ represent attribute hierarchy structure, attribute count, attribute level count, and number of complete Q-matrices in the test, respectively.

### 5.1 Monte Carlo Simulation

**(1) Simulation conditions** are shown in Table 1 (where UC denotes incomplete Q-matrix):

**Table 1: Simulation Conditions**

| Simulation Object | Quantity | Polytomous-attribute polytomous scoring: Examinees follow normal distribution; Polytomous-attribute 0-1 scoring: Examinees follow uniform distribution. All knowledge states obtained via expansion algorithm [?] and generalized expansion algorithm [?]. |
|---|---|---|
| 0-1 scoring | Test length = columns in quasi-reachability matrix (denoted $R_P$) | Test length = 50. Tests contain SSCQM, USCQM, and UC. Item parameters $s_j, g_j \sim U(0, 0.25)$. |

| Simulation Object | Quantity | Polytomous-attribute polytomous scoring: Examinees follow normal distribution; Polytomous-attribute 0-1 scoring: Examinees follow uniform distribution. All knowledge states obtained via expansion algorithm [?] and generalized expansion algorithm [?]. |
|---|---|---|
| Polytomous scoring | Test length = columns in $R_P$ $(N^*)$ | Item parameters $s_j, g_j \sim U(0, 0.25)$. Tests contain $R_P$, UC, and one each of SSCQM and USCQM. |
| 0-1 scoring | Test length = columns in SSCQM $(n^*)$ | Item parameters $s_j, g_j \sim U(0, 0.35)$. Tests contain SSCQM, USCQM, and UC. |
| Polytomous scoring | Test length = 35 | Item parameters $s_j, g_j \sim U(0, 0.35)$. Tests contain quasi-reachability matrix, SSCQM, USCQM, and UC. |

Polytomous-attribute 0-1 scoring uses the RPa-DINA model; polytomous-attribute polytomous scoring uses modified GRPa-DINA. Examinee response patterns are simulated based on true values, Q-matrix, and CDM. Knowledge states are estimated using Maximum A Posteriori (MAP) method.

Short test length $N^*$ examines classification accuracy when tests containing SSCQM and USCQM reach the quasi-reachability matrix column count, comparing them against quasi-reachability and UC. Short test length $n^*$ examines accuracy when quasi-reachability column count reduces to SSCQM column count (making it incomplete UC), comparing against SSCQM and USCQM.

Theoretically, polytomous scoring tests yield higher accuracy, so we set slightly poorer item quality ($s_j, g_j \sim U(0, 0.35)$) and shorter long test length (35 items) versus 0-1 scoring (50 items, $s_j, g_j \sim U(0, 0.25)$) to explore whether accuracy advantages persist.

**(2) Evaluation metrics:** Pattern Match Ratio (PMR) and Marginal Match Ratio (MMR):

$$PMR = \frac{\sum_{i=1}^{N} I\{\hat{\alpha}_i = \alpha_i\}}{N}$$

$$MMR = \frac{\sum_{i=1}^{N}\sum_{k=1}^{K} I\{\hat{\alpha}_{ik} = \alpha_{ik}\}}{NK}$$

where $N$ is total examinees, $I\{\hat{\alpha}_i = \alpha_i\}$ indicates correct attribute pattern classification (1 if correct, 0 otherwise); $K$ is attribute count, $I\{\hat{\alpha}_{ik} = \alpha_{ik}\}$ indicates correct classification of attribute $k$ for examinee $i$.

Simulations repeated 100 times (randomly selecting from available USCQMs), averaging PMR and MMR.

### 5.2 Study 1: Effect of Attribute Hierarchy Structure

**5.2.1 Conditions** Attribute count fixed at 5. Independent structure' s quasi-reachability matrix equals quasi-identity matrix, yielding no USCQM, so we examine four hierarchy structures: Linear (L), Convergent (C), Divergent (D), and Unstructured (U). Attribute level counts are shown in Table 2:

**Table 2: Attribute Level Counts**

| Attribute | Level Count | Levels |
|---|---|---|
| 1-3 | 3 | 0,1,2 |
| 4-5 | 4 | 0,1,2,3 |
| 6-7 | 5 | 0,1,2,3,4 |

*Note: 5 attributes for Study 1, 7 attributes for Study 2.*

**5.2.2 Results** Long and short test results appear in Tables 3 and 4 :

**(1)** For polytomous-attribute 0-1 scoring tests, long tests show higher accuracy than short tests. Both long and short test accuracy decreases with more complex attribute hierarchies. For short tests, quasi-reachability (i.e., SSCQM) achieves highest accuracy, USCQM is second and close to SSCQM, while incomplete Q-matrix (UC) is lowest. For long tests, accuracy ranks: USCQM-containing tests highest, quasi-reachability/SSCQM second, UC lowest.

**(2)** For polytomous-attribute polytomous scoring tests, short test accuracy decreases with hierarchy complexity (MMR decrease smaller than PMR decrease), while long test accuracy generally increases (MMR increase ~4%, PMR increase \$ 6N^), *SSCQM−andUSCQM−containingtestsoutperformquasi−reachabilityandUC, withUSCQM−containingtestshighest.WhenshorttestlengthequalsSSCQMcolumncour* accuracy ranks: USCQM highest, SSCQM second, UC lowest.

### 5.3 Study 2: Effect of Attribute Count

**5.3.1 Conditions** Divergent structure examined. Attribute count $K$ varies from 4 to 7, with level counts from Table 2.

**5.3.2 Results** Results appear in Tables 5 and 6 :

**(1)** For polytomous-attribute 0-1 scoring tests, accuracy decreases as attribute count increases for both long and short tests, with short test decreases smaller than long test decreases. Short test accuracy ranks: quasi-reachability/SSCQM highest, USCQM second, UC lowest. Long test accuracy ranks: USCQM-containing highest, SSCQM-containing second, UC lowest. SSCQM-USCQM differences are \$ \$0.02 for MMR and \$ \$0.04 for PMR.

**(2)** For polytomous-attribute polytomous scoring tests, accuracy decreases with attribute count, with short test decreases larger than long test decreases. Long tests outperform short tests. Polytomous scoring accuracy exceeds 0-1 scoring accuracy. For long tests, complete Q-matrices outperform incomplete ones. For short tests at length $N^*$, accuracy ranks: USCQM-containing, SSCQM-containing, quasi-reachability, UC. At length $n^*$, USCQM highest, SSCQM second, UC lowest.

### 5.4 Study 3: Effect of Attribute Level Count

**5.4.1 Conditions** Divergent structure with 4 attributes, all attributes having 2-5 levels.

**5.4.2 Results** Results appear in Tables 7 and 8 , replicating completeness effects from Studies 1-2: SSCQM and USCQM far outperform UC. Attribute level effects show:

**(1)** For polytomous-attribute 0-1 scoring tests, accuracy is highest at 2 levels. Generally, accuracy decreases with level count for both long and short tests, with long test decreases larger. USCQM-containing long tests achieve highest accuracy.

**(2)** For polytomous-attribute polytomous scoring tests, short test accuracy decreases with level count. At length $N^*$ and levels 2-4, SSCQM-containing tests are most accurate, UC least, with USCQM- and quasi-reachability-containing tests intermediate. At level 5, USCQM-containing tests are most accurate. At length $n^*$, SSCQM and USCQM outperform UC. For long tests, accuracy first increases then decreases with level count, peaking at 3-4 levels.

### 5.5 Study 4: Effect of Complete Q-Matrix Count

**5.5.1 Conditions** Divergent structure with 5 attributes, examining only long tests containing $M$ complete Q-matrices.

**5.5.2 Results** Tables 9 and 10 show accuracy for tests containing $M$ complete Q-matrices versus UC. As $M$ increases:

**(1)** All test accuracies increase. Polytomous-attribute 0-1 scoring tests start lower and gain more, but don’ t reach polytomous scoring upper limits.

**(2)** For polytomous-attribute 0-1 scoring tests, SSCQM-containing tests benefit most from increased complete Q-matrix count, with complete Q-matrices always outperforming UC.

**(3)** For polytomous-attribute polytomous scoring tests, SSCQM- and USCQM-containing tests outperform quasi-reachability-containing tests, with UC lowest.

## 6 Empirical Study

Using real data, we further examine SSCQM and USCQM classification capability. Unlike simulations, real data lack true knowledge state information. Under this study' s scoring scheme and maximum scores, reachability and quasi-reachability matrices serve as complete Q-matrices with high accuracy, so we use them as benchmarks to compute attribute rate (AR) and pattern rate (PR) for SSCQM and USCQM:

$$AR = \frac{\sum_{i=1}^{N} I\{\hat{\alpha}_{ik}^{R} = \hat{\alpha}_{ik}\}}{N}$$

$$PR = \frac{\sum_{i=1}^{N} I\{\hat{\alpha}_{i}^{R} = \hat{\alpha}_{i}\}}{N}$$

where $\hat{\alpha}_{ik}^{R}$ is attribute $k$ estimate from (quasi-)reachability matrix for examinee $i$; $I\{\hat{\alpha}_{ik}^{R} = \hat{\alpha}_{ik}\}$ equals 1 when other test estimates match, 0 otherwise; $\hat{\alpha}_{i}^{R}$ is pattern estimate from (quasi-)reachability matrix; $I\{\hat{\alpha}_{i}^{R} = \hat{\alpha}_{i}\}$ equals 1 when patterns match, 0 otherwise.

### 6.1 Conditions

Data from a carry notation CDT [**?**] with 750 students (705 valid responses). Scoring scheme and maximum scores follow equations (1) and (2). Content involves 5 dichotomous attributes: base conversion concept ($A_1$), decimal to other bases ($A_2$), other bases to decimal ($A_3$), binary to octal/hexadecimal ($A_4$), and octal/hexadecimal to binary ($A_5$), with unstructured hierarchy where $A_1$ is prerequisite. Expert opinion suggests compressing these into 3 polytomous attributes: $A_1'$ (base concept, 2 levels: not assessed/assessed), $A_2'$ (merging $A_2$ and $A_3$, 3 levels: no conversion assessed, other-to-decimal assessed, decimal-to-other assessed), and $A_3'$ (merging $A_4$ and $A_5$, 3 levels: no inter-base conversion assessed, binary-to-octal/hex assessed, octal/hex-to-binary assessed). Hierarchy remains unstructured with $A_1$ as prerequisite. Table 11 shows test items for both representations.

To explore USCQM classification capability, structured items were modified to unstructured (independent structure) by setting all $A_1$ attribute levels to 0 except items 2-4 in Table 11.

From Section 3' s unified design methods: - Dichotomous-attribute 0-1 scoring SSCQM is reachability matrix $R = ()$, with USCQM between identity $E$ and SSCQM: randomly selected $USCQM = ()$ - Dichotomous-attribute polytomous scoring SSCQM = (), with USCQM between identity submatrix $E'$ and SSCQM: randomly selected $USCQM = ()$ - Polytomous-attribute 0-1 scoring SSCQM is quasi-reachability matrix $R_P = ()$, with USCQM between quasi-identity $E_P$ and SSCQM: randomly selected $USCQM = ()$ - Polytomous-attribute polytomous scoring SSCQM = (), with USCQM between quasi-identity submatrix $E'_P$ and SSCQM: randomly selected $USCQM = ()$

Each test type uses one SSCQM, one USCQM, and one UC. We examine classification capability across all four test types. Theoretically, reachability matrix $R$ and quasi-reachability matrix $R_P$ yield high accuracy, so we compare against two $R$ and $R_P$ accuracy estimates.

Using MCMC, we estimate item parameters from examinee response patterns on each test (SSCQM, USCQM, UC, $R$, and $R_P$) using CDMs from Section 5.1(3), then estimate knowledge states and compute AR and PR. Since multiple SSCQMs and USCQMs can be constructed, we average their accuracy values to evaluate classification capability.

### 6.2 Results

Table 12 shows $R_1$ and $R_2$ (reachability matrices) for dichotomous-attribute tests; Table 13 shows $R_{P1}$ and $R_{P2}$ (quasi-reachability matrices) for polytomous-attribute tests. Setting $R_1$ and $R_{P1}$ AR and PR as 1.0, $R_2$ and $R_{P2}$ serve as comparison benchmarks.

**Table 12** shows dichotomous-attribute test AR. For 0-1 scoring, USCQM AR exceeds $R_2$ (i.e., SSCQM), all above 0.90, while UC AR is lower at 0.8377. For polytomous scoring, complete Q-matrix AR are all above 0.84, with SSCQM and USCQM above 0.87, exceeding $R_2$, while UC is lowest at 0.7401. For 0-1 scoring, complete Q-matrix PR are all above 0.61, with USCQM PR exceeding $R_2$ (SSCQM), while UC PR is 0.3844. For polytomous scoring, PR ranks: SSCQM, USCQM, $R$, UC, with UC far below complete Q-matrices.

**Table 13** shows polytomous-attribute AR and PR, with conclusions similar to dichotomous tests: complete Q-matrices outperform incomplete ones, and SSCQM and USCQM accuracy nearly exceeds quasi-reachability $R_{P2}$.

## 7 Discussion and Conclusion

Achieving precise cognitive diagnostic assessment with minimal items is the ultimate goal of CDT. Toward this end, Ding and Luo [**?**] and Ding et al. [**?**] proposed basic complete Q-matrix design concepts for dichotomous-attribute polytomous scoring tests for several basic hierarchy structures. Tang et al. [**?**] further proposed structured and unstructured simplest complete Q-matrices for dichotomous-attribute polytomous scoring tests applicable to all hierarchy structures. Theoretically, polytomous-attribute tests provide richer information and polytomous scoring yields higher accuracy, making research on their design significant. This paper proposes unified design methods for structured and unstructured simplest complete Q-matrices across all combinations of attribute levels and scoring schemes. Four simulation studies and one empirical study systematically examined influencing factors (hierarchy structure, attribute count, level count, complete Q-matrix count) and validity (MMR, PMR, AR, PR).

### 7.1 Discussion and Future Directions

**(1)** Simplest complete Q-matrices are closely related to scoring schemes and maximum scores. While reachability matrices are generally considered complete, they require specific scoring schemes and maximum scores to be truly complete [**?**]. The proposed polytomous scoring scheme is widely used, satisfies attribute-score correspondence [**?**], and assumes equal attribute weights. If weights differ, this method cannot be used. When $\beta_k = 1$, maximum score equals the sum of attribute level counts. If maximum score exceeds this sum, the unified design method still applies. When attributes and scores are independent, the method also works for 0-1 scoring. Future research should examine other scoring schemes (e.g., attribute manifestation assumptions, conjunctive condensation rules [**?**]) and maximum score configurations (e.g., $m_j \leq \sum \beta_k q_{kj}$ or $m_j \geq \sum \beta_k q_{kj}$ with unequal weights). Future work should also investigate simplest complete Q-matrix designs for models like DINO, A-CDM, and G-DINA.

**(2)** Explore constructing SSCQM and USCQM from different item types. This study uses reachability and quasi-reachability matrices. Future research should investigate constructing these matrices beyond reachability matrices or from all possible item types. For all item types, SSCQM is a submatrix of (quasi-)reachability matrices, but is it the minimal-column complete Q-matrix? Not necessarily. For example, with 6 unstructured attributes, dichotomous-attribute polytomous scoring SSCQM from reachability matrix is $Q_{21} = ()$, while SSCQM from non-reachability items is $Q_{22} = ()$. Investigating classification accuracy and influencing factors for simplest complete Q-matrices constructed from different item types is important for test design.

**(3)** USCQMs generated by the unified method require completeness verification. Since lower bounds are incomplete, matrices between bounds must be verified for establishing one-to-one correspondence between knowledge states and ideal response patterns. With many attributes and levels, the number of intermediate matrices grows substantially. Constructing complete lower bound matrices could reduce verification time. For example, in inverted pyramid structures (Figure 4 [Figure 4: see original paper]), lower bound completeness construction relates closely to dichotomous-attribute polytomous scoring USCQM lower bounds: prerequisite attributes cannot be assessed in the same column (see Appendix example). Whether other special hierarchy structures exist requires further research.

**(4)** Results inform short test type selection. Although SSCQM and USCQM theoretically have equivalent distinguishing power, experiments show each has advantages under different conditions. If attribute hierarchy structure is clear, short tests might use SSCQM; if hierarchy structure is uncertain, USCQM may be preferable.

**(5)** Regarding CDT validity: Test validity measures how well test content matches measured constructs. CDT validity indicators include attribute/pattern classification accuracy and theoretical construct validity (TCV) [**?**]. TCV examines how well the test Q-matrix represents theoretical attributes and hierarchy structures [**?**], assessing consistency between ideal response patterns and true latent classes [**?**]. Currently, TCV mainly applies to dichotomous-attribute test design; polytomous-attribute TCV indicators remain unstudied. Thus, this study uses classification accuracy (MMR, PMR in simulations; AR, PR in empirical studies) as validity indicators. CDT reliability faces similar issues, warranting future research on polytomous-attribute validity and reliability indicators.

### 7.2 Conclusions

Theoretically, multiple partitioning possibilities for comparable items generate more SSCQMs and USCQMs than quasi-reachability matrices, addressing test design uniformity issues, reducing exposure rates, and enhancing test security. Results show: (1) Complete Q-matrices outperform incomplete ones. For long and short tests containing SSCQM, USCQM, and (quasi-)reachability matrices, accuracy decreases with more complex hierarchy structures, more attributes, and more attribute levels (except polytomous-attribute polytomous scoring long tests), while increasing with more complete Q-matrices. (2) For short tests at quasi-reachability column count, SSCQM- and USCQM-containing tests almost always outperform (quasi-)reachability matrices. At SSCQM column count (fewer columns than reachability), SSCQM and USCQM are minimal-column complete Q-matrices outperforming incomplete matrices. Simulations show SSCQM and USCQM superiority over (quasi-)reachability matrices in short tests. (3) For long tests, SSCQM- and USCQM-containing tests achieve classification capability comparable to (quasi-)reachability matrices. (4) Empirical studies demonstrate SSCQM and USCQM achieve high accuracy relative to (quasi-)reachability matrices in short tests.

## References

*Note: References are preserved in their original format as provided in the source document.*

# Appendices

## I. Theoretical Proof of SSCQM Completeness

Ding and Luo [?] and Ding et al. [?] proved completeness for dichotomous-attribute polytomous scoring test Q-matrices (consistent with this study' s designs). For polytomous-attribute polytomous scoring SSCQM proofs, we consider converting polytomous Q-matrices to Boolean matrices (dichotomous), then applying Boolean matrix polytomous scoring completeness proofs.

**Key facts:** 1) Polytomous Q-matrices convert to Boolean matrices via expansion algorithm (P-to-D conversion), and back via compression (D-to-P conversion) [?], establishing a one-to-one correspondence. 2) Using the proposed scoring scheme (1) and maximum scores (2), computer-mined polytomous simplest complete Q-matrices (denoted $B$) have minimal columns, ensuring simplicity. 3) Proof steps: (a) Convert polytomous Q-matrix test to Boolean matrix via expansion; (b) Apply Boolean matrix polytomous scoring completeness theory [?, ?] to obtain Boolean simplest complete Q-matrix for given hierarchy, attribute count, and level count; (c) Compress to polytomous simplest complete Q-matrix $B$, matching our results. 4) Computer verification confirms that for various hierarchy structures, attribute counts, and level counts, the obtained simplest complete Q-matrices establish one-to-one correspondence between knowledge states and ideal response patterns.

## II. Simplest Complete Q-Matrices Applied to Cognitive Diagnostic Models

### (1) Polytomous-attribute 0-1 scoring SSCQM and USCQM with RPa-DINA (Appendix Table 1 )

*Table content preserved exactly as in original document*

### (2) Polytomous-attribute polytomous scoring SSCQM and USCQM with modified GRPa-DINA (Appendix Tables 2 and 3)

*Table content preserved exactly as in original document*

## III. Example from Discussion Section 7.1(3)

Using Figure 1 with attribute level counts 2, 2, 3, 4, and 5, dichotomous-attribute polytomous scoring USCQM lower bounds can be $Q_{11} = ()$, $Q_{12} = ()$, $Q_{13} = ()$, but not $Q_{14} = ()$. For polytomous-attribute polytomous scoring USCQM lower bounds, attributes 1 and 2 cannot appear together in columns expanded from dichotomous identity submatrix $E_1 = ()$ columns 3-5, so the complete lower bound must ensure each $\star$ and $\bullet$ contains at least one 1.

**Figure 4 [Figure 4: see original paper]: Inverted Pyramid Structure**

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

*Source: ChinaXiv —Machine translation. Verify with original.*