Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 8, 5343-5363

A Modified Theory-Based Method For Answer-Correlated Weighting Ordering Theory And Its Application

[1]Hui-Ju Chen , [2*]Shu-Chuan Shih , [3].Tian-Wei Sheu , [4]Boi-Yee Liao

[1]*Graduate Institute of Educational Information and Measurement, National Taichung University, Taichung, Taiwan, R.O.C. cms103106@gm.ntcu.edu.tw*
[2*]*Graduate Institute of Educational Information and Measurement, National Taichung University, Taichung, Taiwan, R.O.C. ssc@mail.ntcu.edu.tw*
[3]*Graduate Institute of Educational Information and Measurement, National Taichung University, Taichung, Taiwan, R.O.C. ssc@mail.ntcu.edu.tw*
[4] *Instituteof Engineering Technology Management, International College, Krirk University, Bangkok, Thailand. Y57082112@gmail.com*

Abstract

Multiple-choice questions have been used as objective tests, and their results have served as optimal materials for analyzing test question response theories and conceptual structures. People's answers to multiple-choice questions can only be used to analyze distracters and trick questions. At present, conceptual structure analyses only assess whether correct or incorrect answers have been given when providing visual descriptions, failing to account for information hidden among different incorrect answers. In practice, question designers can design incorrect answers that reveal some information about the participants' knowledge. Accordingly, this study proposed a new answer-correlated weighted conceptual structure model that enhances the precision of analysis results to more accurately reflect participants' learning statuses.

To verify the validity of the answer-correlated weighting ordering structure, this study used ordering theory and modified the model through answer weighting to develop answer-correlated weighting ordering theory. A simulation study was conducted to assess the estimation accuracy of the model. Participants' responses were simulated using Ozaki's structured deterministic inputs, noisy "and" gate model for multiple-choice items, and a four-part cognitive attribute structure was used to form ideal test question responses. The four-part structure, five sample sizes, four participant answering errors, and three different test question numbers were used to create 240 test scenarios. A total of 100 simulated binary sum answers were generated for each simulation scenario. The results demonstrated that answer-correlated weighting ordering theory exhibited the most effective estimation performance and generated favorable results in all participant "answering error" situations.

Keywords: Q-matrix theory, ordering theory, multiple-choice questions, answer-correlated weighting ordering theory.

Introduction

Knowledge learning is "in order" in that previously learned knowledge or concepts are the prerequisites for the acquisition of subsequent knowledge. During the learning process, the brain stores acquired knowledge structurally. When acquiring new knowledge, two processes occur; the knowledge received is affected by the existing knowledge structure, and the knowledge in turn changes the original knowledge structure. On the basis of the aforementioned knowledge learning characteristics, if students' knowledge structure and misconceptions can be identified early during the teaching process, crucial feedback information can be provided to teachers and students, improving teaching methods and learning results, respectively. Accordingly, determining how to estimate and analyze structural changes before and after knowledge acquisition has become a critical research topic in the field of cognitive diagnostic assessment.

Airasian and Bart (1973) argued that, similar to knowledge learning, test questions exhibit order-like characteristics. Thus, they applied ordering theory (OT) to estimate the order structure of test questions and used said structure to establish students' knowledge learning order (Bart & Krus, 1973) and to develop learning order-based cognitive diagnoses. In 1997, Appleby, Samuels, and Treasure-Jones developed Diagnosis, the first knowledge structure-based computerized adaptive testing system for fundamental mathematics university courses. Later, the knowledge structure-based adaptive testing (KSAT)system, a joint OT and knowledge structure-based computerized adaptive testing system, was introduced. Shih, Kuo, and Liu (2012) indicated that on average, using OT and students' responses to build students' knowledge structure and then applying appropriate questions from question databases to diagnose students' misconceptions can reduce more than half the time required to prepare test questions (Liu, Jeng, Tsai, Yu,& Lu, 2013). The time saved can then be used to provide remedial education. The aforementioned diagnostic results are robust, and this system is effective even when applied to other units or related topics. One study compared students' cognitive and expert structures (Kalyuga, Rikers, & Paas, 2012) for teachers to diagnose students' learning difficulties or misunderstandings in the context of remedial education.

Wu, Kuo, and Yang(2012) conducted a simulation study in which they compared the estimation accuracy of and number of test questions presented using adaptive test systems based on OT, item relational structure(Takeya, 1991), Diagnosis, and domain experts. They observed that OT-based adaptive test systems demonstrated the most favorable performance, with their estimated performance similar to that obtained in pen and paper-based tests. OT is the primary method used to build student knowledge structure in the KSAT system. However, OT is mainly applied to binary scoring data, and cognitive diagnosis models (CDMs) receive limited information from responses. Accordingly, Liu (2007) assigned different scores to the answers to constructed-response items at different steps. Related studies have extended binary scoring to poly to outscoring, expanding the application of OT and constructed-response items that adopt polytomous scoring methods (Liu, Sheu, Tsai, Kniew, and Guo, 2017).Liu(2012a, 2012b) used the correlation coefficients between test questions to exclude the independent effects of test questions on each other and developed test question OT based on the Q-matrix theory. All efficient items (or "efficient test questions") clearly correlated with cognitive attributes were tested to obtain the test question correlation structure of all participants and of each individual participant, facilitating cognitive diagnosis analyses and the subsequent provision of proper remedial education. Through enhancing OT and providing the information necessary for adaptive systems to build individuals' knowledge structure, estimation accuracy can be elevated.

CDM

Cognitive diagnosis assessments assess individuals' cognitive processes, processing skills, or knowledge structure (Gierl, Leighton, & Hunka, 2000; Yang & Embretson, 2007). These assessments measure participants' test responses, with CDMs subsequently applied to estimate participants' potential knowledge state (KS) according to their responses. In cognitive diagnoses, participants' KS refers to the participants' mastery of the cognitive attributes or skills required in the assessed tasks. Therefore, in education assessments, relevant tests are designed and students' responses are obtained to determine individuals' knowledge profiles (Mislevy, Steinberg & Almond, 2003).

Q-Matrix Theory

In 1983, Tatsuoka proposed a CDM based on Q-matrix theory called the "rule-space model." Tatsuoka's Q-matrix theory uses test questions to derive related cognitive attributes and their structural relationships. Assuming that a test that measures K cognitive attributes can generate $2^K - 1$ possible test questions after removal of the test questions without any attributes, the correlation matrix between the cognitive attributes and test questions is Q-matrix($\mathbf{Q} = (q_{kj})_{K \times (2^K - 1)}$, $1 \le k \le K$, $1 \le j \le 2^K - 1$). $q_{kj} = 1$ if test question j tests cognitive attribute A_k; otherwise, $q_{kj} = 0$.

Cognitive attributes are relevant attributes, skills, or strategies that experts or question designers wish to measure through tests. When attribute Ai must be known before attribute Aj, Ai is referred to as the prerequisite attribute (or lower-level attribute) of A_j. By contrast, A_j is referred to as the subsequent attribute (or upper-level attribute) of Ai ($A_i \mapsto A_j$). If the absence of attribute A_k results in $A_i \mapsto A_k$ and $A_k \mapsto A_j$, then A_i is the direct prerequisite attribute of A_j ($A_i \to A_j$).Figure 1 illustrates a cognitive attribute structure with four attributes, where attributes$A_2$and A_3arethe direct prerequisite at tributes of A_4, attribute A_1 is the direct prerequisite attribute of $A_2$and A_3, and attribute$A_1$is the prerequisite attribute of A_4. Thus, $A_2 \to A_4$, $A_3 \to A_4$, $A_1 \to A_2$, $A_1 \to A_3$, and $A_1 \mapsto A_4$.

The direct prerequisite relationships between K cognitive attributes can be expressed as K-order binary matrix $\mathbf{A}_K = (a_{ij})_{K \times K}$, which is referred to as the adjacency matrix. If attribute A_i is a direct prerequisite attribute of A_j (i.e., $A_i \to A_j$), matrix element $a_{ij} = 1$; otherwise, $a_{ij} = 0$. The prerequisite relationships between cognitive attributes can be expressed as K-order binary matrix $\mathbf{R}_K = (r_{ij})_{K \times K}$, which is referred to as the reach ability matrix. If attribute A_i is a prerequisite attribute of A_j (i.e., $A_i \mapsto A_j$), $r_{ij} = 1$; otherwise, $r_{ij} = 0$.Figure2 depicts the adjacency matrix and reachability matrix of the cognitive attribute structure presented in Figure 1.

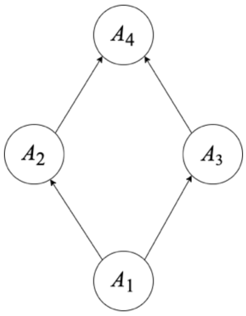

Figure 1

Cognitive AttributeStructure

$$\mathbf{A}_4 = \begin{pmatrix} 0\,1\,1\,0 \\ 0\,0\,0\,1 \\ 0\,0\,0\,1 \\ 0\,0\,0\,1 \end{pmatrix} \qquad \mathbf{R}_4 = \begin{pmatrix} 1\,1\,1\,1 \\ 0\,1\,0\,1 \\ 0\,0\,1\,1 \\ 0\,0\,0\,1 \end{pmatrix}$$

Figure2

Adjacency Matrix and Reachability Matrix

Given cognitive attributes A and B, if A is the prerequisite attribute of B, $A \mapsto B$. In this study, test question j measured only attribute B and not attribute A. Thus, to answer the question correctly, only attribute B was required. Because this was logically flawed and did not conform to the cognitive attribute structure described, such test questions were categorized as inefficient test questions; the other test questions were categorized as efficient test questions. Inefficient test questions in the Q-matrix were removed, producing a reduced Q-matrix (i.e., $\mathbf{Q}_R$) that served as the blueprint for developing tests that satisfied the cognitive attribute structure. As presented inFigure 1, a test question in the cognitive attribute structure has a Q vector of 0100, signifying that the test question tests only attributeA_2. However, according to the cognitive attribute structure, attributeA_1 is a prerequisite attribute of A_2, meaning that the test question tests both attributes$A_2$and A_1; therefore, a test question with a Q vector of 0100 does not conform to the cognitive attribute structure presented in Figure 1. This test question is thus categorized as an inefficient test question and removed from the Q-matrix. A Q-matrix with four cognitive attributes must have 15 (i.e., $2^4 - 1 = 15$) test questions. Figure3depicts a reduced Q-matrix following the removal of inefficient test questions and retention of the remaining five efficient test questions. According to the correlation structure of $\mathbf{Q}_R$, the structure between cognitive attributes can be recertified, and the correlation structure between efficient test questions can be determined (Liu, 2012a). Participants' KS signifies the attributes required for them to answer test questions correctly. Under the assumption that participants do not guess the answers or choose the incorrect answers by mistake, $\mathbf{Q}_R$ plus a 0 vector without attributes is equal to participants' KS matrix α (Figure **4**). The matrix element $a_{ik} = 1$ represents the i-th KS, which can be used to solve the k-th cognitive attribute; otherwise, $a_{ik} = 0$.

$$\mathbf{Q} = \begin{pmatrix} 101010101010101 \\ 011001100110011 \\ 000111100001111 \\ 000000011111111 \end{pmatrix} \rightarrow \mathbf{Q}_R = \begin{pmatrix} 11111 \\ 00111 \\ 01011 \\ 00001 \end{pmatrix}$$

Figure3 Q-Matrix and Reduced Q-Matrix

$$\alpha = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Figure 4 KS Matrix

Ideal Responses

Ideal responses are participants' expected responses to test questions, assuming they neither guess the answers nor mistakenly choose the incorrect answers. For example, according to the reduced Q-matrix and KS matrix illustrated in Figure3and Fig. 4, respectively, with the assumption that the KS of Participant i is $\alpha_i = (1,0,1,0)$, the participant only has cognitive attributes A_1 and A_3 and can only answer Questions 1 and 2 in the $\mathbf{Q}_R$ correctly; thus, their ideal responses are $(1,1,0,0,0)$. If a test has K cognitive attributes, n test questions, and N dissimilar KSs, the participant's ideal response for test question j would be $h_{ij} = \prod_{k=1}^{K} a_{ik}^{q_{kj}}$. Here $h_{ij} = 1$ signifies that the participant has all the cognitive attributes required to answer the test question correctly.Table1 lists the corresponding ideal responses for all the KS in Figure **4**, respectively.

Table1

Knowledge State and Ideal Responses

Knowledge state(α)	Ideal response(η)
(0, 0, 0, 0)	(0, 0, 0, 0, 0)
(1, 0, 0, 0)	(1, 0, 0, 0, 0)
(1, 0, 1, 0)	(1, 1, 0, 0, 0)
(1, 1, 0, 0)	(1, 0, 1, 0, 0)
(1, 1, 1, 0)	(1, 1, 1, 1, 0)
(1, 1, 1, 1)	(1, 1, 1, 1, 1)

Multiple-Choice Test Questions and Test Question Q-Matrix

Test question Q-matrices present the relationships between test questions and attributes and are commonly used in CDMs with binary scoring. These models estimate participants' KS based on whether they have answered the test questions correctly. If test questions are multiple-choice questions that adopt the binary scoring method, the information provided by distractors is overlooked. De la Torre (2009) combined answers and Q-matrices to create answer Q-matrices, joining distractors and participants' KS and enhancing KS estimation accuracy.Table2 presents the test question Q-matrix used in de la Torre's study; the number 0 indicates a test question without any attribute, with the other numbers denoting the number of times a test question was linked to an attribute. For example, the Q vector of Question 21is (3,2,1,0,0), signifying that three, two, and one of the four answers are linked to one, two, and three attributes, respectively.

Based on this method, the test question answer Q-matrix was constructed, as detailed in **Table3**. Assuming that the Q vectors of the four answers were $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$, and $\mathbf{q}_4$, then $\mathbf{q}_4 \subset \mathbf{q}_3 \subset \mathbf{q}_2 \subset \mathbf{q}_1$. Here, $\mathbf{q}_1$is a correct answer (Answer1), and the Q vector of any subsequent answer is a subset of the previous answer.

Q-Matrix as a Basic CDM

To diagnose participants' learning mistakes and provide them with effective remedial education according to these diagnoses, national and international scholars have continually revised and advanced CDMs. Today, CD Ms with Q-matrices all use the deterministic inputs, noisy "and" gate (DINA) and deterministic inputs, noisy "or" gate (DINO) models as their basis.

Table2

Multiple-Choice Test Question Q-Matrix

	Attribute		Attribute
Test question	1 2 3 4 5	Test question	1 2 3 4 5
1	1 0 0 0 0	16	0 1 0 2 0
2	0 1 0 0 0	17	0 1 0 0 2
3	0 0 1 0 0	18	0 0 1 2 0
4	0 0 0 1 0	19	0 0 1 0 2
5	0 0 0 0 1	20	0 0 0 1 2
6	1 0 0 0 0	21	3 2 1 0 0
7	0 1 0 0 0	22	1 2 0 3 0
8	0 0 1 0 0	23	3 2 0 0 1
9	0 0 0 1 0	24	1 0 3 2 0
10	0 0 0 0 1	25	2 0 3 0 1
11	2 1 0 0 0	26	1 0 0 2 3
12	2 0 1 0 0	27	0 2 3 1 0
13	1 0 0 2 0	28	0 3 1 0 2
14	2 0 0 0 1	29	0 2 0 3 1
15	0 2 1 0 0	30	0 0 1 3 2

Source: de la Torre (2009)

Table3

Answer Q-Matrix for Test Question 21

	Attribute				
	1	2	3	4	5
Answer1	1	1	1	0	0
Answer2	1	1	0	0	0
Answer3	1	0	0	0	0
Answer4	0	0	0	0	0

Following numerous corrections and enhancements by various researchers, the DINA, DINO, and generalized deterministic inputs, noisy "and" gate model (G-DINA) are as follows:

DINA Model. Junker and Sijtsma (2001) developed the DINA model to assess whether participants understand their own cognitive attributes. When participants lack the attributes measured through test questions, they are regarded as unable to answer the questions. Accordingly, the probability of Participant i answering question j correctly presented as follows:

$$P(X_{ij}=1|\alpha_i)=(1-s_j)^{\eta_{ij}}g_j^{(1-\eta_{ij})} \tag{1}$$

Here, η_{ij} is Participant i's ideal response to test question j given that they have KS α_i ;s_j is the "mistakenly choose the incorrect answer" parameter; and g_j is the guess parameter.

DINO Model. Templin and Henson (2006) divided participants into two categories, with one category containing participants who had mastered at least one attribute required to answer test questions, and the other containing participants who had not mastered any attribute required to answer test questions. Next, they proposed the DINO model in which the probability of Participant i answering question j correctly is as follows:

$$P(X_{ij}=1|\alpha_i)=(1-s_j)^{\varpi_{ij}}g_j^{(1-\varpi_{ij})} \tag{2}$$

Here, $\varpi_{ij}=1-\prod_{k=1}^{K}(1-\alpha_{ik})^{q_{jk}}$. $\varpi_{ij}=1$ signifies that Participant i has mastered at least one cognitive attribute required to answer test question j correctly, whereas $\varpi_{ij}=0$ signifies that Participant i has not mastered any cognitive attribute required to answer test question j correctly.

G-DINA Model. In the DINA model, unless the participants have mastered all attributes required to answer test questions correctly, the participants are regarded as having guessed the answers to some test questions. Thus, when estimating the probabilities of participants answering test questions correctly, the probabilities of them guessing the correct answers are applied. Similarly, in the DINO model, when participants have not mastered any cognitive attribute required to answer test questions correctly, the probabilities of them guessing the correct answers are used to estimate the probabilities of them answering test questions correctly. Because both the DINA and DINO models are impractical, de la Torre (2011) calculated the correlations between attributes and introduced the G-DINA model. The model divides participants into $2^{K_j^*}$ groups, and defines the probability of Participant i answering test question j correctly as follows:

$$P(\alpha_{ij}^*)=\delta_{j0}+\sum_{k=1}^{K_j^*}\delta_{jk}\alpha_{lk}+\sum_{k'>k}^{K_j^*}\sum^{K_j^*-1}\delta_{jkk'}\alpha_{lk}\alpha_{lk'}+\delta_{j12\ldots K_j^*}\prod_{k=1}^{K_j^*}\alpha_{lk} \quad (3)$$

where,

δ_{j0}:The intercept of test questionj.

δ_{jk}:The main effects on α_k.

$\delta_{jkk'}$:The interactive effects on α_k and $\alpha_{k'}$.

$\delta_{j12\ldots K_j^*}$:The interactive effects from α_k to $\alpha_{k'}$.

δ_{j0} represents the effects produced when a participant does not have any attribute required to answer a given test question, whereas $\delta_{jkk'}$ is the interactive effects generated when a participant has both attributes α_k and $\alpha_{k'}$. $\delta_{j12\ldots K_j^*}$ is the effect size when a participant has all the attributes required to answer a given test question. Thus, in the G-DINA model, varying participant KSs may result in the participant having different probabilities of answering test questions correctly.

Test Question OT-Based Cognitive Diagnoses

Test Question OT

First, assume that N participants are completing a test with n test questions; a value of $x_{ij}=1$ is assigned if Participant I answers test question j correctly, and otherwise, $x_{ij}=0$ ($i=1,2,\ldots,N$, $j=1,2,\ldots,n$). Next, two test questions are presented, namely j and m, with test question j less difficult than test question m. Assuming that Participant I does not guess the answers or choose the incorrect answers by mistake, the lower the probability ratio of answering test question j incorrectly and test question m correctly, the more favorable the result. On the basis of this concept, Airasian and Bart (1973) developed test question OT to determine the order structure between test questions (i.e., the prerequisite relationships between test questions) and define the order coefficients of test questions j and m, which is described as follows:

$$\gamma_{jm}=1-P\left(x_j=0,x_m=1\right) \quad (4)$$

Here, $P\left(X_j=0,X_m=1\right)=\frac{1}{N}\#\left\{i;x_{ij}<x_{im}\right\}$ is the joint probability that a participant answers test questionjincorrectly and test question m correctly. If $\gamma_{jm}>\varepsilon\in\left[0.96,0.98\right]$and $\varepsilon$is the threshold constant of a subjective decision, test questions j and m exhibit a sequential relationship. That is, test question j is the prerequisite test question of test question m, and $X_j\to X_m$ is marked.

Ozaki's Multiple-Choice-S-DINA Model

De la Torre (2009) combined test question answers and the DINACDM (Junker & Sijtsma, 2001; de la Torre, 2008)in developing the multiple-choice deterministic inputs, noisy "and" gate (MC-DINA)model, using distractors to reveal participants' KS. When participants select answers, distractors can induce specific responses from the participants to correctly diagnose their KS. In heriting the characteristics of the DINA model, the MC-DINA model has noncompensatory and connective characteristics. Noncompensatory characteristic refer to cognitive attributes that are absent cannot be compensated by other pre-existing cognitive attributes. By contrast, connective characteristics represent the identical probabilities of participants answering test questions correctly despite not having the attribute required to answer said test questions to those of participants answering questions correctly despite not having two or more attributes required to answer said test questions. On the basis of MC-DINA, Ozaki(2015) proposed the MC-S-DINA model, enabling participants with varying KSs to exhibit different probabilities of answering test questions correctly.

Given that a test measures K cognitive attributes and contains ntest questions, and that each question has l answers that may be selected, if Participant i has a KS of α_i, the probability that they select c as the answer for test question j is expressed as follows:

$$P\left(X_{ij}(c)=1|\alpha_i\right)=\left(1-\prod_{k=1}^{K}\left(1-\alpha_{ik}\right)\right)\left(1-\delta_j(c)\right)^{\eta_{ij}(c)}\omega_{ij}(c)^{1-\eta_{ij}(c)}+\frac{1}{l}\prod_{k=1}^{K}\left(1-\alpha_{ik}\right) \quad (5)$$

where

$$\omega_{ij}(c)=\delta_j(c)\frac{(1-\eta_{ij}(c))(1+r_{ij}(c))}{\sum_{c=1}^{l}(1-\eta_{ij}(c))(1+r_{ij}(c))} \quad (6)$$

If participants do not have any attributes required to answer test questions correctly, then the probability of them answering any test question correctly is one out of the number of possible answers. If participants' KS and the cognitive attributes measured with answer c are identical, $\eta_{ij}(c)=1$; otherwise, $\eta_{ij}(c)=0$. $\delta_j(c)$ is the probability of participants selecting answers that are not identical to their KS. When this occurs, the participants are likely to guess the answers or mistakenly choose the incorrect answers. $1-\delta_j(c)$ signifies the probability of participants selecting answers that are identical to their KS. $\omega_{ij}(c)$ is the probability that Participant i selects the correct answer to test question j when their KS differs from the cognitive attributes measured using said test question by at least one. Here, $r_{ij}(c)=\sum_{k=1}^{K}\alpha_{ik}q_{kj}(c)$ is used to calculate the degree of approximation between Participanti's KS and the cognitive attributes of test questionj's answer c. The higher the $r_{ij}(c)$, the higher the degree of approximation between the participant's KS and the cognitive attributes measured with answer c, and the higher the probability ($\omega_{ij}(c)$) that the participant selects said answer. Conversely, the lower the $r_{ij}(c)$, the wider the gap between the participant's KS and the cognitive attributes measured with answer c, and the lower the $\omega_{ij}(c)$.Thus, when participants lack a varying number (at least one)of attributes measured through certain answers, the probabilities that they answer the test questions correctly differ.

Study Methods

Answer-Correlated Weighting OT

Traditional multiple-choice tests adopt a binary scoring method, that is, participants receive a score of one when they are correct, and zero otherwise. Therefore, they only receive scores for correct answers and not for incorrect answers nor for choosing distractors as answers. In diagnostic tests, if researchers analyze the binary response-based total test scores only (de la Torre, 2009; Liu et al., 2015; Ozaki, 2015), information provided by distractors is lost. OT identifies the sequential relationships between test questions through use of binary responses. Liu et al. (2015) analyzed the information provided by test question answers, converting binary responses into answer indicators and calculating the correlation coefficients between answers and total test scores. The coefficients were then set as normalized weighting coefficients of the answers relative to their test questions, transforming test scores from binary scores(i.e., zero orone)into polytomous scores between zero and one.

However, distractors are negatively correlated with total scores (Gierl & Lai, 2018). The method adopted by Liuet al. (2015) resulted in standardized answer weighting coefficients of distracters being assigned as zero and the weighting coefficients of incorrect answers being assigned a higher value than those of distractors. Therefore, this study used the correlation coefficients between answer indicators and correct answers to recalculate standardized weighting coefficients, subsequently proposing answer-correlated weighting ordering theory (OCWOT).

Assume N participants are completing a test with n test questions, and that each question has lpossible answers. If the ith participant chooses answer c as the answer to test question j, $x_{ij}(c)=1$;otherwise, $x_{ij}(c)=0$ (i = 1,2,…,N; j = 1,2,…, n;c = 1,2,…,l).

Under the assumption that the correct answer to test question j is answer c^*, the answers selected by participants is weighted in correlation to c^*. This study proposed a method of identifying the correlation coefficient $w_j(c)$ between answers c and c^*(the correct answer) for test question j. That is,

$$w_j(c)=\frac{\sum_{i=1}^{N}\left(x_{ij}(c)-\overline{x_j(c)}\right)\left(x_{ij}(c^*)-\overline{x_j(c^*)}\right)}{(N-1)S_j(c)S_j(c^*)} \tag{7}$$

where, $\overline{x_j(c)}=\frac{1}{N}\sum_{i=1}^{N}x_{ij}(c)$;

$\overline{x_j(c^*)}=\frac{1}{N}\sum_{i=1}^{N}x_{ij}(c^*)$;

$S_j(c)=\left\{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_{ij}(c)-\overline{x_j(c)}\right)^2\right\}^{\frac{1}{2}}$;

$S_j(c^*)=\left\{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_{ij}(c^*)-\overline{x_j(c^*)}\right)^2\right\}^{\frac{1}{2}}$

Next, set $W_j(c)=\left\{\left|w_j(c)\right|;c=1,2,...,l\right\}$.In this study, the weighted score of answer c to test question j was $w_j(c)=\frac{W_j(c)-\min W_j}{\max W_j-\min W_j}$.That is, the score of Participant i for test question j in a test was $\sum_{c=1}^{l}x_{ij}(c)w_j(c)$. This signifies that the score of Participant i was $\sum_{j=1}^{n}\sum_{c=1}^{l}x_{ij}(c)w_j(c)$, contrary to that obtained using the previous calculation method $\sum_{j=1}^{n}\sum_{c=1}^{l}x_{ij}(c)$.

When $x_{ij}=\sum_{c=1}^{l}x_{ij}(c)w_j(c)$, researchers can obtain the poly to obtain matrix of $\left(x_{ij}\right)_{N\times n}$, that is, $x_{ij}\in\left[0,1\right]$, which is different from $x_{ij}\in\left\{0,1\right\}$. The order coefficient of test question j and test question m is thus defined as follows:

$$g_{jm}=\begin{cases}1-V_{jm} & ,L_{jm}\neq 0\\ 0 & ,L_{jm}=0\end{cases} \tag{8}$$

$$V_{jm}=\frac{\sum_{c_1=1}^{l}\sum_{c_2=1}^{l}\left[w_m(c_1)-w_j(c_2)\right]^{+}P\left(X_j=w_j(c_1),X_m=w_m(c_2)\right)}{\sum_{c_1=1}^{l}\sum_{c_2=1}^{l}\left[w_m(c_1)-w_j(c_2)\right]^{+}} \tag{9}$$

$$L_{jm}=\sum_{c_1=1}^{l}\sum_{c_2=1}^{l}\left[P\left(X_j=w_j(c_1),X_m=w_m(c_2)\right)-P\left(X_j=w_j(c_1)\right)P\left(X_m=w_m(c_2)\right)\right]^2 \tag{10}$$

Here, $\left[x\right]^{+}=\max\{x,0\}$.

$P\left(X_j=w_j(c_1),X_m=w_m(c_2)\right)$ is the joint probability of receiving a score of $w_j(c_1)$ for test question j and a score of $w_m(c_2)$ for test question m. $P\left(X_j=w_j(c_1)\right)$ and $P\left(X_m=w_m(c_2)\right)$ are the marginal probabilities of receiving a score of $w_j(c_1)$ for test question j and a score of $w_m(c_2)$ for test question m ($j,m=1,2,...,n$), respectively. L_{jm} is used to test the correlation between test questions jand m. When the two test questions are independent, $L_{jm} = 0$, and the two test questions have an order coefficient of $g_{jm} = 0$. When $g_{jm} > e = [0.96, 0.98]$, test questions j and m share a sequential relationship in which test question j is the prerequisite test question of test question m (i.e., $X_j \to X_m$). Here, $e$is a predetermined threshold.

Studying the Simulation and Empirical Data

To compare the accuracy of OT and OCWOT in constructing participants' test question order structure, the simulation study data and empirical data were compared.

Simulation data and procedure. The cognitive attribute structure of the test introduced in this simulation study was built using the sequential relationships of cognitive attributes. Leighton, Gier, and Hunka (2004) maintained that all cognitive attribute structures are constructed using the four basic cognitive attributes presented in Figure 5 (i.e., linear, convergent, divergent, and unstructured cognitive attributes). Accordingly, this study used the four-part basic cognitive attribute structure as a blueprint with which to generate the Q-matrices required for the test questions and their answers.

The test questions all contained four answers. The control factors were as follows:(a)sample size:50,100,150,200, and 1,000participants;(b)percentage of guesses and incorrect answers chosen by mistake:10%, 20%,30%, and 40%;(c) number of test questions:10,20, and 30; and (d)cognitive attribute structure: convergent, divergent, linear, and unstructured structure. The possible simulation scenarios were 240 ($5 \times 4 \times 3 \times 4 = 240$), with 100 binary sum answers generated for each simulation scenario.

The simulation procedure is as follows:

Step 1:

(1) Choose a basic cognitive attribute structure and establish its adjacency and reachability matrices.

(2) Remove test questions in the Q-matrix that do not conform to the test questions in the cognitive attribute structure to derive the reduced Q-matrix $\mathbf{Q}_R$.

(3) Obtain participants' KS (α) through adding the 0 vector to the $\mathbf{Q}_R$ and expand participants' KS to the desired sample size, ensuring the total test score conforms to normal distribution.

Step 2:

(1) Select the corresponding number of questions from the $\mathbf{Q}_R$ according to the required number of test questions.

(2) Given that all test questions have four answers, select subsets from test question Q vectors as answer Q vectors, ensuring that the subsets meet the cognitive attribute structure selected in Step 1.

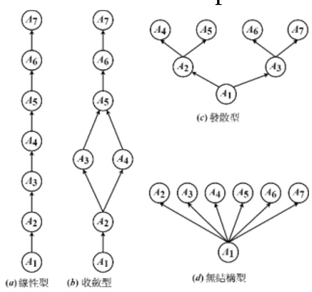

Figure 5 Four-Part Basic Cognitive Attribute Structure

Step 3:

(1) Use the KS in Step 1 nd test question and answer Q-matrices in Step 2–(2) to calculate the ideal responses of the binary sum answers.

(2) Apply the MC-S-DINA model presented by Ozaki (2015) on the KS in Step 1–(4)and test question and answer Q-matrices in Step 2–(2) to calculate the probability that Participant i chooses answer c for test question j($P(X_{ij}(c) = 1, \alpha_i)$) according to the probability of them guessing or mistakenly choosing the incorrect answers.

(3) Generate the simulated binary sum answers based on the probabilities obtained in Step 3–(2).

Step 4:

(1) Use the ideal responses of binary sum answers in Step 3–(1) to calculate the ideal order coefficient matrices.
(2) Employ the 100 simulated binary sum answers in Step 3–(3) to calculate the estimated order coefficient matrices.
(3) Apply the ideal order coefficient matrices in Step 4–(1) as the criterion validity to calculate the root-mean-square error (RMSE) of the order coefficient matrices. This then serves as the judgment criterion to determine the effectiveness of OT in comparison to OCWOT.

Constructing the test question answer Q-matrices. This study used the method presented in Figure2 to construct the adjacency and reachability matrices for the four-part cognitive attribute structure depicted in Figure 5; all seven attributes were measured, producing 127 ($2^7 - 1 = 127$) possible test questions. Subsequently, inefficient test questions (test questions that did not conform to the cognitive attribute structure) were removed to obtain the number of efficient test questions for each attribute structure (Table4) and the number of test questions for the attributes measured. For example, the divergent and linear structure had 25and 7 efficient test questions, respectively, and the numbers of efficient test questions that measured three attributes in the divergent structure and unstructured part were 5 and 15, respectively. The test questions were extracted from efficient test questions and had cognitive attribute-containing answers as their subsets. Therefore, these test questions must conform to the original cognitive attribute structure. If the number of test questions to be extracted for a certain number of measurement attributes was greater than the number of efficient test questions, extractions were repeated. For example, when the number of test questions was 20, only two efficient test questions measuring three attributes were in the convergent structure. However, because six test questions were required, the two efficient test questions were applied three times. By contrast, if the number of test questions was less than the number of efficient test questions, test questions with attributes that were not repeated were randomly extracted from efficient test questions. For example, the divergent structure contained six questions measured four attributes. Because only five questions were required in the simulated test questions, five questions were randomly selected from these six questions.

This simulation study introduced three different test question sets, with the first, second, and third question sets containing10, 20, and 30 test questions, respectively. The number of attributes listed in Table 5 represented the number of attributes measured using the correct answers to each test question. For example, in the set with 30test questions, five test questions measured six attributes; in the set containing 10test questions, one test question measured seven attributes. Table 6details the convergent structure of the Q-matrix with 10 test questions, whereby the first answers to all test questions were the correct answers.

Table7 presents the convergent structure of the Q-matrix with 10 test question answers, with the convergent structure used as the subsequent structure.

Table4

Number of Effective Test Questions in the Four-Part Cognitive Attribute Structure

Number of Attributes measured	Convergent	Divergent	Linear	Unstructured
1	1	1	1	1
2	1	2	1	6
3	2	5	1	15
4	1	6	1	20
5	1	6	1	15
6	1	4	1	6
7	1	1	1	1
Total	8	25	7	64

Table5
Number of Test Questions Required for each Attribute Number in the Simulation Study Tests

Number of attributes		3	4	5	6	7
Number of Test questions	10	3	2	2	2	1
	20	6	5	4	3	2
	30	8	7	6	5	4

Table 6
Convergent Structure Q-Matrix With 10 Test Questions

	Attribute						
Test question	1	2	3	4	5	6	7
1	3	2	0	1	0	0	0
2	3	2	1	0	0	0	0
3	3	2	0	1	0	0	0
4	4	3	2	1	0	0	0
5	4	3	2	1	0	0	0
6	4	4	3	2	1	0	0
7	4	4	3	2	1	0	0
8	4	4	3	4	2	1	0
9	4	4	4	3	2	1	0
10	4	4	4	4	3	2	1

Table7
Convergent Structure Q-Matrix With 10 Test Questions Answers

		Cognitive attributes			Cognitive attributes
Test questions	Answers	1 2 3 4 5 6 7	Test questions	Answers	1 2 3 4 5 6 7
1	1	1 1 0 1 0 0 0	6	1	1 1 1 1 1 0 0
	2	1 1 0 0 0 0 0		2	1 1 1 1 0 0 0
	3	1 0 0 0 0 0 0		3	1 1 1 0 0 0 0
	4	0 0 0 0 0 0 0		4	1 1 0 0 0 0 0
2	1	1 1 1 0 0 0 0	7	1	1 1 1 1 1 0 0
	2	1 1 0 0 0 0 0		2	1 1 1 1 0 0 0
	3	1 0 0 0 0 0 0		3	1 1 1 0 0 0 0
	4	0 0 0 0 0 0 0		4	1 1 0 0 0 0 0
3	1	1 1 0 1 0 0 0	8	1	1 1 1 1 1 1 0
	2	1 1 0 0 0 0 0		2	1 1 1 1 1 0 0
	3	1 0 0 0 0 0 0		3	1 1 1 1 0 0 0
	4	0 0 0 0 0 0 0		4	1 1 0 1 0 0 0
4	1	1 1 1 1 0 0 0	9	1	1 1 1 1 1 1 0
	2	1 1 1 0 0 0 0		2	1 1 1 1 1 0 0
	3	1 1 0 0 0 0 0		3	1 1 1 1 0 0 0
	4	1 0 0 0 0 0 0		4	1 1 1 0 0 0 0
5	1	1 1 1 1 0 0 0	10	1	1 1 1 1 1 1 1
	2	1 1 1 0 0 0 0		2	1 1 1 1 1 1 0
	3	1 1 0 0 0 0 0		3	1 1 1 1 1 0 0
	4	1 0 0 0 0 0 0		4	1 1 1 1 0 0 0

Establishing the KS of the test sample. To expand the KS to the number of samples required for the simulations, this study first obtained the ideal responses of each KS. Next, the numbers of ideal responses recorded were tallied, and the order of the corresponding KS was rearranged according to the number of test questions answered correctly (which were listed in ascending order). The sample sizes were organized to form a normal distribution in each KS. Subsequently, the KS was repeatedly generated according to the size to produce the true values of participants' KS in the simulation study.

Table 8presents the distribution of the 250 test samples in each KS in the convergent structure.

Generating the ideal responses. When participants' KS is identical to the cognitive attributes measured using their answers, or when the cognitive attributes measured using their answers are a subset of their KS, these answers constitute ideal responses. If a test that measures K attributes has n test questions, and each test question has l answers, given that α_i is Participantsi's KS, and that $q_j(c)$ is the Q vector of test question j answer c, Participanti's ideal response to test question j.

Table 8
Frequency Allocation of the 250 samples in the Convergent Structure

Knowledge state	Ideal response	Number of questions answered correctly	Sample size	Percentage
α_1	0000000	00000000	0	1	0.4%
α_2	1000000	10000000	1	13	5.2%
α_3	1100000	11000000	2	41	16.4%
α_4	1110000	11100000	3	35	14.0%
α_5	1101000	11010000	3	35	14.0%
α_6	1111000	11111000	4	70	28.0%
α_7	1111100	11111100	5	41	16.4%
α_8	1111110	11111110	6	13	5.2%
α_9	1111111	11111111	7	1	0.4%

Table9
Ideal Responses in the Convergent Structure

Knowledge state	Test questions1	Test question2	Test question3	Test question4	Test question5	Test question6	Test question7	Test question8	Test question9	Test question10
1	4	4	4	4	4	4	4	4	4	4
2	3	3	3	4	4	4	4	4	4	4
3	2	2	2	3	3	4	4	4	4	4
4	1	2	1	3	3	4	4	4	4	4
5	2	1	2	2	2	3	3	4	4	4
6	1	1	1	1	1	2	2	3	3	4
7	1	1	1	1	1	1	1	2	2	3
8	1	1	1	1	1	1	1	1	1	2
9	1	1	1	1	1	1	1	1	1	1

Is $h_{ij}(c)=\prod_{k=1}^{K}\alpha_{ik}^{q_{kj}(c)}$ ($j=1,2,...,n;c=1,2,...,l$).If $h_{ij}(c)=1$, the participant has all the attributes required to select the correct answers. Conversely, if $h_{nj}(c)=0$, the participant lacks at least one of the attributes required to select the correct answers. Because the number of attributes measured using the test question answers in this simulation study were gradually decreased, when participants' KS allowed them to answer multiple questions correctly, the answers that measured the most attributes were regarded as the participants' ideal responses to test question j(de la Torre, 2009).If the participants' KS did not enable them to answer any question correctly, the participants' ideal responses were answers that measured the least number of attributes. The convergent structure ideal responses are reported in Table9. In this study, the correct answers to the test questions were always the first answers.

Simulating the participants' answers and binary responses. The MC-S-DINA model introduced by Ozaki(2015) calculates the probabilities that participants select certain answers. Thus, this study employed this model to simulate the participants' answers. The percentage of guesses and incorrect answers chosen by mistake was set as10%, 20%,30%, and 40%, which corresponded to the percentage of guesses and incorrect answers chosen by mistake in the MC-S-DINA model (i.e., δ_{j1} ,j=1,2,…,20). The percentage of each answer being a guess or incorrect answers chosen by mistake is calculated as follows:

$$\left[\delta_{j1},\delta_{j2},\delta_{j3},\delta_{j4}\right]=[\delta_{j1},U(\delta_{j1},0.5),U(0.5,1-\delta_{j1}),1-\delta_{j1}] \quad (11)$$

Here, U(a, b)is the uniform distribution from a to b. Let $P_{ij}(C=c)$ be the probability that Participant i chooses c as the answer to test question j, and r be a uniformly distributed random number conforming to U(0,1). When $r \le P_{ij}(C=1)$, the participant's answer is Answer 1; when $P_{ij}(C=1) < r \le P_{ij}(C=1)+P_{ij}(C=2)$, the participant's answer is Answer 2, and so on. Because Answer1 is the correct answer, when chosen, the binary response recorded is one. If a different answer is chosen, the binary response recorded is zero.

Empirical Data

This study used empirical data obtained from the Eight Intelligences: Mathematical Logic Test (Chen, 2011) to perform analyses. Mathematical logic intelligence refers to the ability to effectively use numerical calculations, perform measurements and classifications, and apply logic and reasoning to analyze problems, complete mathematical operations, and explore problems scientifically. The test contained 30 test questions with four possible answers, and the test evaluation content covered four attributes (i.e., number, sequence, figurate numbers, and real life applications). The test questions, which had a reliability coefficient (Cronbach's α) of .930, were administered to104 Grade 6 students studying in the second semester of the 2015 to 2016 academic year.

Study Results

Simulation Study Assessment Criteria

When determining the effectiveness of different OT, researchers analyzed the correlation coefficients between OT-based and ideal response-based test question order structures (Liu, 2013; Liu et al., 2015).The higher the correlation coefficients are, the closer the estimated order structure was to the ideal test question order structure.

OT calculates the order coefficients between two test question responses. When the order coefficient is greater than the preset threshold, the two test questions are regarded as sharing a sequential relationship, on the basis of which the order structure of all test questions can be established. According to their objectives, researchers often use different thresholds to generate different order structures, rendering a comparison across theoretical models applying previous methods unsuitable.

A response set has only one order coefficient matrix, and ideal responses represent participants' responses when answers are neither guessed nor the incorrect answer mistakenly chosen. Therefore, the order coefficient matrix calculated using ideal responses is surmised to be the optimal order coefficient matrix or the "ideal test question order coefficient matrix." Thus, the closer that the estimated test question order coefficient matrices calculated using simulated responses are to the ideal test question order coefficient matrices, the more optimal the OT effectiveness. This simulation study calculated the order coefficient matrices of ideal and simulated responses. OT and OCWOT calculated order coefficient matrices using ideal binary responses and ideal answers, respectively. Next, RMSEs were used to determine OT and OCWOT effectiveness.

$$RMSE = \frac{1}{G}\sum_{g=1}^{G}\frac{1}{n^2}\sqrt{\sum_{j=1}^{n}\sum_{m=1}^{n}\left(\hat{g}_{jm}^{(g)} - g_{jm}\right)^2} \quad (12)$$

Here, G is the number of response groups generated when the relevant parameters are fixed in a simulation study. In this simulation study, $G = 100$, n is the number of test questions in the test set, g_{jm} is the test question order coefficients of test questions j and m obtained through calculating the ideal responses, and $\hat{g}_{jm}^{(g)}$ is the test question order coefficient obtained through calculating the simulated responses of G. Through calculation of the RMSE of the test question order coefficient of G and the ideal test question order coefficient and application of the average value of the two OTs, the effectiveness of OT was compared; the smaller the RMSE, the more accurate the OT estimation result.

Simulation Study Results

Simulation study results. A four-part cognitive attribute structure was established using OT and OCWOT. is the lower-level test question (or prerequisite test question) of Question1.That is, to answer Question 1 correctly, participants must answer Question 5 correctly. The rate at which lower-level test questions are answered correctly is expected to be higher than that of upper-level test questions. Figure6 and 7, the structural diagrams of the two methods are divided into four levels, with a greater number of lower-level test questions generated using OCWOT then using OT in the structure. This is attributable to OT’s use of binary scoring, whereas OCWOT uses polytomous scoring after weighting answers. This enables OCWOT to be more sensitive than OT in determining upper and lower levels after calculating two test question order coefficients, resulting in a greater number of lower-level test questions in structural diagrams created using OCWOT. The overall structure subsequently becomes more complete, and estimations of participants’ mathematical logic-based knowledge structure become more reasonable.

Table10, 11, and 12 present the RMSEs of the four-part cognitive attribute structure with10, 20, and 30test questions, respectively. The following conclusions were derived:

(1) An increased sample size increased the order coefficient estimation accuracy.
(2) When the numbers of test questions differed, participants more frequently guessing and mistakenly choosing incorrect answers increased the convergent, divergent, and linear RMSEs, thereby decreasing the estimation accuracy of both OT and OCWOT. However, the unstructured RMSEs did not exhibit consistent changes (i.e., they either increased or decreased). This indicated that unstructured attributes were independent of each other, that test questions did not have consistent sequential relationships, and that answers provided limited information.
(3) Overall, OCWOT proposed in this study exhibited the optimal effectiveness.

Empirical data from the experimental results. The empirical data involved the analysis of the participants’ answers and binary responses using OCWOT and OT. The test question structure was drawn based on a threshold (ε) of 0.97 to compare the differences between the two theories.

The Eight Intelligences: Mathematical Logic Test evaluated the test question order structural

Question 1 and 2 presented in Figure 8 are both sequential questions. Question 1 can be answered through identifying the patterns of the sequence of prime numbers, whereas Question 2 requires expanding the denominators and then identifying the sequence patterns. In OCWOT, Question 1 is the lower-level test question of Question 2; that is, participants must answer Question 1 correctly before they could successfully answer Question 2. In terms of the test question itself, this thinking order for problem-solving is reasonable, but in OT, these two questions are regarded as test questions of equal status.

diagrams drawn using OT and OCWOT, as depicted in Figure6and Figure7, respectively; the numbers in circles denote the test questions, and the arrows represent the relationships between test questions. For example, as illustrated inFigure7,the arrow of Question5 points to Question1 (i.e., “5→1”), signifying that Question5 is the lower-level test question (or prerequisite test question) of Question1.That is, to answer Question 1 correctly, participants must answer Question 5 correctly. The rate at which lower-level test questions are answered correctly is expected to be higher than that of upper-level test questions. Figure6 and 7, the structural diagrams of the two methods are divided into four levels, with a greater number of lower-level test questions generated using OCWOT then using OT in the structure. This is attributable to OT’s use of binary scoring, whereas OCWOT uses polytomous scoring after weighting answers. This enables OCWOT to be more sensitive than OT in determining upper and lower levels after calculating two test question order coefficients, resulting in a greater number of lower-level test questions in structural diagrams created using OCWOT. The overall structure subsequently becomes more complete, and estimations of participants’ mathematical logic-based knowledge structure become more reasonable.

Question 1 and 2 presented in Figure 8 are both sequential questions. Question 1 can be answered through identifying the patterns of the

sequence of prime numbers, whereas Question 2 requires expanding the denominators and then identifying the sequence patterns. In OCWOT, Question 1 is the lower-level test question of Question 2; that is, participants must answer Question 1 correctly before they could successfully answer Question 2. In terms of the test question itself, this thinking order for problem-solving is reasonable, but in OT, these two questions are regarded as test questions of equal status.

Table10

RMSEs of the Four-Part Cognitive Attribute Structure Calculated Using OT and OCWOT (10Test Questions)

		OT					OCWOT				
Cognitive attributes	Percentage of participants choosing the wrong answer by mistake/making guesses (%)	Sample size									
		50	100	150	200	1000	50	100	150	200	1000
convergent	10	.1060	.0997	.1003	.0994	.0970	.0248	.0243	.0252	.0230	.0218
	20	.1259	.1229	.1202	.1189	.1171	.0321	.0316	.0293	.0284	.0269
	30	.1448	.1370	.1383	.1373	.1345	.0362	.0351	.0338	.0339	.0318
	40	.1532	.1472	.1472	.1480	.1460	.0412	.0387	.0377	.0375	.0354
divergent	10	.1126	.1081	.1058	.1054	.1038	.0240	.0250	.0223	.0213	.0204
	20	.1209	.1110	.1115	.1114	.1095	.0275	.0274	.0232	.0241	.0230
	30	.1217	.1175	.1153	.1141	.1146	.0304	.0307	.0280	.0251	.0242
	40	.1226	.1161	.1141	.1168	.1140	.0328	.0306	.0280	.0287	.0269
linear	10	.1028	.0985	.0986	.0964	.0959	.0261	.0248	.0234	.0231	.0228
	20	.1266	.1205	.1201	.1209	.1165	.0334	.0297	.0293	.0273	.0258
	30	.1448	.1396	.1378	.1359	.1371	.0349	.0340	.0332	.0303	.0309
	40	.1522	.1474	.1473	.1485	.1460	.0379	.0361	.0360	.0352	.0341
unstructured	10	.1256	.1157	.1160	.1136	.1129	.0289	.0238	.0237	.0236	.0219
	20	.1248	.1209	.1172	.1166	.1150	.0306	.0287	.0276	.0260	.0246
	30	.1260	.1183	.1165	.1152	.1138	.0323	.0296	.0270	.0257	.0258
	40	.1288	.1172	.1153	.1155	.1148	.0328	.0297	.0281	.0279	.0269

Table11

RMSEs of the Four-Part Cognitive Attribute Structure Calculated Using OT and OCWOT(20 Test Questions)

		OT					OCWOT				
Cognitive attributes	Percentage of participants choosing the wrong answer by mistake/making guesses (%)	Sample size									
		50	100	150	200	1000	50	100	150	200	1000
convergent	10	.1093	.1062	.1046	.1036	.1007	.0272	.0264	.0250	.0243	.0227
	20	.1336	.1287	.1272	.1259	.1235	.0332	.0335	.0307	.0300	.0287
	30	.1515	.1474	.1438	.1440	.1420	.0391	.0376	.0368	.0361	.0340
	40	.1660	.1587	.1567	.1559	.1552	.0439	.0409	.0406	.0387	.0384
divergent	10	.1119	.1048	.1032	.1029	.1010	.0288	.0254	.0227	.0227	.0222
	20	.1198	.1134	.1112	.1104	.1088	.0342	.0281	.0264	.0244	.0240
	30	.1234	.1192	.1164	.1157	.1148	.0339	.0320	.0289	.0285	.0266
	40	.1288	.1230	.1205	.1203	.1199	.0377	.0331	.0295	.0302	.0292
linear	10	.1090	.1046	.1007	.1000	.0989	.0279	.0264	.0242	.0242	.0236
	20	.1333	.1284	.1275	.1250	.1238	.0335	.0313	.0298	.0288	.0290
	30	.1506	.1449	.1446	.1425	.1428	.0391	.0370	.0350	.0334	.0330
	40	.1627	.1583	.1565	.1562	.1558	.0427	.0401	.0399	.0377	.0370
unstructured	10	.1245	.1181	.1149	.1150	.1139	.0298	.0261	.0249	.0239	.0227
	20	.1286	.1202	.1200	.1166	.1158	.0318	.0293	.0274	.0268	.0254
	30	.1266	.1205	.1176	.1150	.1157	.0330	.0292	.0292	.0275	.0270
	40	.1252	.1210	.1187	.1161	.1165	.0321	.0312	.0305	.0283	.0271

Table12

RMSEs of the Four-Part Cognitive Attribute Structure Calculated Using OT and OCWOT(30Test Questions)

	OT	OCWOT
Cognitive	Sample size	

attributes	Percentage of participants choosing the wrong answer by mistake/making guesses (%)	50	100	150	200	1000	50	100	150	200	1000
convergent	10	.1116	.1093	.1066	.1069	.1049	.0272	.0269	.0252	.0250	.0234
	20	.1363	.1316	.1299	.1274	.1258	.0344	.0326	.0315	.0307	.0289
	30	.1505	.1468	.1448	.1441	.1431	.0386	.0377	.0368	.0352	.0341
	40	.1655	.1575	.1582	.1573	.1553	.0431	.0412	.0402	.0401	.0378
divergent	10	.1141	.1092	.1061	.1057	.1057	.0310	.0278	.0247	.0235	.0242
	20	.1226	.1152	.1119	.1127	.1121	.0342	.0301	.0268	.0264	.0262
	30	.1274	.1186	.1182	.1174	.1168	.0353	.0315	.0301	.0303	.0293
	40	.1310	.1248	.1212	.1209	.1204	.0372	.0347	.0312	.0322	.0307
linear	10	.1086	.1050	.1026	.1027	.1021	.0282	.0267	.0252	.0249	.0247
	20	.1335	.1272	.1257	.1248	.1247	.0349	.0308	.0304	.0288	.0287
	30	.1523	.1448	.1442	.1432	.1434	.0391	.0353	.0352	.0345	.0332
	40	.1607	.1586	.1565	.1558	.1552	.0421	.0409	.0400	.0378	.0377
unstructured	10	.1261	.1203	.1190	.1181	.1176	.0310	.0287	.0269	.0256	.0250
	20	.1312	.1217	.1199	.1198	.1193	.0325	.0297	.0290	.0269	.0270
	30	.1327	.1221	.1204	.1195	.1188	.0354	.0320	.0303	.0301	.0277
	40	.1299	.1228	.1206	.1190	.1188	.0357	.0324	.0328	.0312	.0289

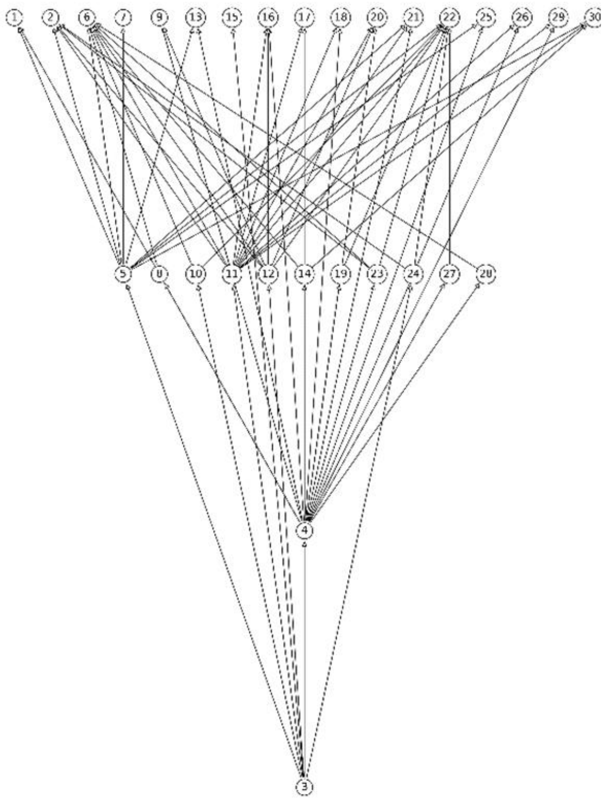

Figure6
Mathematical Logic-Based OT Structure

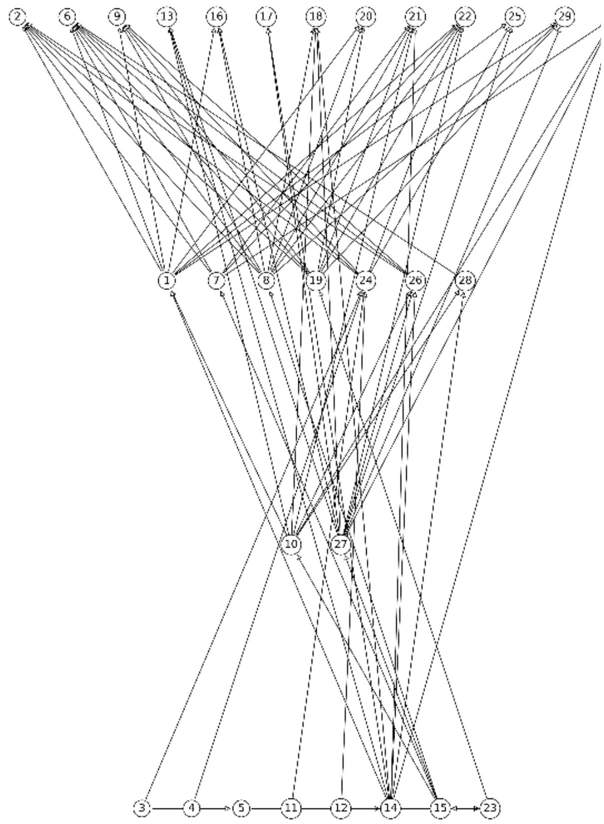

Figure7
Mathematical Logic-Based OCWOT Structure

1. Which number should be in the parentheses based on the pattern of these numbers? 2, 3, 5, 7, ( ), 13, 17,… ,
(A) 9 (B) 10 (C) 11 (D) 12
2. Which number should be in the parentheses based on the pattern of these numbers? 1/2, 1/2, 3/8, ( ), 5/32, 3/32, 7/128, …
(A) 3/4 (B) 3/8 (C) 1/4 (D) 1/8

Figure 8 Mathematical Logic-Based Test Questions 1 and2

Conclusion and Recommendations

Multiple-choice questions as a question format are advantageous because they can be answered quickly and scores can be calculated easily. However, because participants only receive scores if they select the correct answers, these questions do not take the partial knowledge of participants into account, unlike diagnostic tests. The failure to score such knowledge precludes information provided by distractors, resulting in the inaccurate determination of participants' knowledge in diagnostic tests.

As an effective tool for cognitive diagnoses, OT is often used in on-site teaching to analyze students' knowledge structure and diagnose their learning difficulties and deficiencies, gaining information that can be subsequently used as the basis on which teachers modify their courses and provide remedial education. Tests have evolved beyond traditional pen and paper-based tests, transitioning to computerized and online tests using computerized adaptive testing systems. This evolution of test tools allows students to provide answers quickly and reduces the time teachers spend on marking tests and analyzing students' learning difficulties. OT plays a crucial role in computerized adaptive testing systems, which use knowledge structure as their basis. OT is used to establish the knowledge structure of domain experts and references students' responses to construct their knowledge structure and identify their learning difficulties. OT has been proven effective in various practical applications (Chang, Ku, Yu, Wu,& Kuo, 2015; Shih et al., 2012; Wu et al., 2012, 2017). However, it is only applicable to the response data of binary scoring and cannot extract the diagnostic information of distractors in multiple-choice questions. Accordingly, this study used the correlations between answers and correct answers, thus proposing OCWOT that extracts distractor information. Contrary to other methods, the order coefficient matrices of ideal responses were used as the criterion validity and order coefficient matrices obtained from simulated responses were used to calculate RMSEs, which were applied to judge the OT effectiveness of the two test questions. This study investigated OT and OCWOT effectiveness through conducting simulation research and using empirical data, revealing that OCWOT out performed OT in terms of effectiveness and estimation accuracy (i.e., improvement of 50% to 70%). Therefore, when hidden information in answers is taken into account, the estimated effectiveness of test question order coefficients is enhanced. The empirical data system was used to draw test question structural diagrams based on a threshold (ε)of 0.97, which were used to compare the differences between OT and OCWOT. The results demonstrated that, compared with OT,OCWOT's use of polytomous scoring results in greater sensitivity in determining upper and lower levels, thus generating a greater number of lower-level test questions in structural diagrams compared with those generated using OT. This provides a more reasonable method with which teachers can improve their teaching processes and implement remedial education strategies.

Both OT and OCWOT assess group data and do not provide individual diagnosis information. Therefore, the efficacy of nonparametric kernel-smoothing test question OT for individual participants is a topic that warrants study (Liu et al., 2003; Liu, Wu, Sheu, Chung, &Tsai, 2012). Scholars have achieved favorable results in combining MC-DINA(de la Torre, 2009) and computerized adaptive testing systems (Yigit, Sorrel, & de la Torre, 2018).However, the KSAT system, which is based on OT, does not account for the information provided by distractors. By contrast, OCWOT proposed in this study integrates distractor information and thus outperforms OT. OCWOT could be combined with the KSAT system to construct students' cognitive structures to facilitate the development of remedial education.

References

1. Chen, H.J. (2011).The Study of the Multiple Intelligences Computerized Testing. National Taichung University of Education (unpublished master's thesis), Taichung.
2. Liu, H.C., Sheu, T.W., Tsai, H.C., Kniew, K.F., and Guo S.J. (2017). Improving Generalized Polytomous Ordering Theory for any Real Number Scoring. Psychological Testing, 64(4), 295-312.
3. Airasian, P. W., Bart, W. M. (1973). Ordering theory: A new and useful measurement model. Journal of Education Technology, 5, 56-60
4. Appleby, J., Samuels, P., &Treasure-Jones, T. (1997). Diagnosis—a knowledge-based diagnostic test of basic mathematical skills. Computers & Education, 28(2), 113–131. https://doi.org/1.1016/S0360-1315(97)00001-8 M4-
5. Bart, W. M., & Krus, D. J. (1973). An Ordering-Theoretic Method to Determine Hierarchies Among Items. Educational and Psychological Measurement, 33(2), 291–3. https://doi.org/1.1177/001316447303300208 M4
6. Chang, S.-H., Ku, A.-C., Yu, L.-C., Wu, T.-C., & Kuo, B.-C (2015). A science, technology, engineering and mathematics course with computer-assisted remedial learning system support for vocational high school students. Journal of Baltic Science Education, 14(5), 641-654.
7. Chen, L. H. (2011). Enhancement of student learning performance using personalized diagnosis and remedial learning system. Computers & Education, 56(1), 289-299
8. de la Torre, J. (2008). An empirically-based method of Q-matrix validation for the DINA model: Development and applications. Journal of Educational Measurement, 45, 343-362.

9. de la Torre, J. (2009). A Cognitive Diagnosis Model for Cognitively Based Multiple-Choice Options. Applied Psychological Measurement, 33, 163-183. doi:1.1177/0146621608320523
10. de la Torre, J.(2011). The Generalized DINA Model Framework. Psychometrika, 76,179–199. https://doi.org/10.1007/s11336-011-9207-7
11. Gierl, M. J., Leighton, J. P., & Hunka, S. (2000). Exploring the logic of Tatsuoka's rule-space model for test development and analysis. Educational Measurement: Issues and Practice, 19, 34-44. doi:1.1111/j.1745-3992.2.tb00036.x
12. Gierl, M. J., & Lai, H. (2018). Using Automatic Item Generation to Create Solutions and Rationales for Computerized Formative Testing. Applied Psychological Measurement, 42, 42-57. doi:1.1177/0146621617726788
13. Harel, G. (2000). Three principles of learning and teaching mathematics, in: J.-L. Dorier (Ed.), On the teaching of Linear Algebra, Kluwer Academic Publishers, Dordrecht, 2000, pp.177-189.
14. Hillel, J. (2000). Modes of description and the problem of representation in linear algebra, in: J.-L. Dorier (Ed.), On the teaching of Linear Algebra, Kluwer Academic Publishers, Dordrecht, 2000, pp.191-207.
15. Hwang, G. J., Panjaburee, P., Triampo, W., & Shih, B. Y. (2013). A group decision approach to developing concept–effect models for diagnosing student learning problems in mathematics. British Journal of Educational Technology, 44 (3), 453-468.
16. Jonassen, D. H, Beissner, K., and Yacci, M. (1993). Structural knowledge: Techniques for representing, converying, and acquiring structural knowledge. Hillsdale, NJ:Lawrence Erlbaum Associates.
17. 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/1.1177/01466210122032064
18. Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. Educational Psychology Review, 24(2), 313-337
19. Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The Attribute Hierarchy Method for Cognitive Assessment. Journal of Educational Measurement, 41, 205-237. doi:1.1111/j.1745-3984.2004.tb01163.x
20. Liu, H.C., Sheu, T.W., Kou, B.C., Shih, S.C., Shiu, C.L., Item Ordering Theories Based on Nonparametric Item Response Theory, IMPS-2003, International Meeting of the Psychometric Society, Sardinia, Italy,2003.
21. Liu, H.C. (2007). Ordering Theory for Standardized and Normalized Polytomous Response Testing, Journal of Research on Measurement and Statistics, 15,1-11. https://doi.org/1.6773/JRMS.200706.0001
22. Liu, H.C. (2012a). Ideal item ordering matrix and Liu's ordering theory. Asia University.
23. Liu, H.C. (2012b). A novel validity index for evaluating the item ordering structure based on Q-matrix theory, Research Journal of Applied Sciences, Engineering and Technology, 6(1), 2012, 53-56. doi: 1.19026/rjaset.6.4033
24. Liu, H. C., Wu, S. N., Sheu, T. W., Chung, Y. T., & Tsai, H-C. (2012). Polytomous item ordering theory based on ramsay's kernel smoothing nonparametric item response theory. ICIC Express Letters, 6(6), 1537-1542.
25. Liu, H.C. (2013) A Novel Validity Index for Evaluating the Item Ordering Structure Based on Q-matrix Theory. Research Journal of Applied Sciences, Engineering and Technology,6(1), 53-56.
26. Liu, H.C.(2014). Liu's improved polytomous ordering theory based on extensional item ordering consistency theorem. In: International Conference on

Information Management 2014, Taichung, Taiwan.

27. Liu, H. C., Jeng, B. C., Tsai, H. C., Yu, Y. K., & Lu, C. Y. (2013). Liu's Polytomous Item Relational Structure Theory Based on Q-Matrix Theory. Applied Mechanics and Materials, 311, 93–98. https://doi.org/10.4028/www.scientific.net/amm.311.93
28. Liu, H.C., Shia, B.C. and Cheng, D.Y. (2014). An improved Liu's ordering theory based on empirical distributive critical value, in The 27th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems, Kaohsiung, Taiwan.
29. Liu, H.C., Ou, Y.C., Tsai, H.C., Yu, Y.K., Cheng, T.Y., & Su, H.W. (2015). A Novel Ordering Theory For Multiple Choice Question Based on Liu's Polytomous Ordering Theory. Proceedings of International Multi-Conference on Engineering and Technology Innovation 2015 (IMETI2015), Kaohsiung, Taiwan.
30. Junker, B. W., & Sijtsma, K. (2001). Nonparametric Item Response Theory in Action: An Overview of the Special Issue. Applied Psychological Measurement, 25(3), 211–220. https://doi.org/10.1177/0146621 0122032028
31. Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. Measurement: Interdisciplinary Research and Perspectives, 1(1), 3-62. doi: 1.1207/S15366359MEA0101_02
32. Ozaki, K. (2015). DINA Models for Multiple-Choice Items with Few Parameters: Considering Incorrect Answers. Applied Psychological Measurement, 39, 431-447. doi:1.1177/0146621615574693
33. Shih, S.C., Kuo, B.C., & Liu, Y.L. (2012). Adaptively ubiquitous learning in campus math path. Educational Technology & Society, 15(2), 298-308.
34. Tatsuoka, K K. (1983). Rule space: an approach for dealing with misconceptions based on item response theory. Journal of Educational Measurement, 20(4): 345-354.
35. Takeya,M. (1991). New item structure theorem. Tokyo: Waseda University.
36. 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
37. Ting, M.Y., & Kuo, B.C. (2016). A Knowledge-structure-based Adaptive Dynamic Assessment System for Calculus Learning. J. Comp. Assist. Learn.,32(2), 105–119. https://doi.org/1.1111/jcal.12119
38. Wu, H.M., Kuo, B.C., & Yang, J.M. (2012). Evaluating knowledge structure based adaptive testing algorithms and system development. Educational Technology & Society, 15(2), 73-88.
39. Wu, H. M., Kuo, B. C., & Wang, S. C. (2017). Computerized Dynamic Adaptive Tests with Immediately Individualized Feedback for Primary School Mathematics Learning. Educational Technology & Society, 20 (1), 61–72
40. Yang, X. D., & Embretson, S. E. (2007). Construct validity and cognitive diagnostic assessment. In J. P. Leighton. & M. J. Gierl (Eds.), Cognitive diagnostic assessment for education: Theory and applications(pp. 119–145). Cambridge, UK: Cambridge University Press.
41. Yigit, H. D., Sorrel, M. A., & de la Torre, J. (2018). Computerized Adaptive Testing for Cognitively Based Multiple-Choice Data. Applied Psychological Measurement, 43(5), 388-401. doi:1.1177/0146621618798665