

## A Dual-Objective Item Selection Strategy for CD-CAT Based on Gini Index

**Authors:** Luo Fen, Wang Xiaoqing, Cai Yan, Tu Dongbo, Tu Dongbo

**Date:** 2020-09-02T00:00:00+00:00

### Abstract

The test results of dual-objective CD-CAT can be utilized for both formative and summative assessments. The Gini index can quantify the uncertainty degree of a random variable, where smaller values indicate lower uncertainty. This study employs the Gini index to measure changes in posterior probabilities of examinee knowledge state categories and ability estimation confidence intervals, proposing a Gini index-based item selection strategy. Monte Carlo experiments demonstrate that, compared with existing item selection strategies, the new strategy yields higher accuracy in both knowledge state classification and ability estimation, while effectively balancing item bank usage uniformity, enabling rapid real-time response, and exhibiting minimal influence from cognitive diagnosis models and examinee knowledge state distributions, thus being applicable to mixed item banks containing multiple cognitive diagnosis models in practical testing.

### Full Text

### Preamble

### A New Dual-Objective CD-CAT Item Selection Method Based on the Gini Index

LUO Fen<sup>1,2</sup>, WANG Xiaoqing<sup>2</sup>, CAI Yan<sup>1</sup>, TU Dongbo<sup>1</sup>

(<sup>1</sup> School of Psychology, Jiangxi Normal University, Nanchang 330022, China)

(<sup>2</sup> College of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022, China)

### Abstract

Dual-objective CD-CAT assessment results can be used for both formative and summative evaluation. The Gini index can measure the degree of uncertainty in

random variables, where a smaller value indicates lower uncertainty. This paper uses the Gini index to measure changes in the posterior probabilities of examinee knowledge state classifications and ability estimation confidence intervals, proposing a Gini-index-based item selection strategy. Monte Carlo experiments demonstrate that compared with existing selection strategies, the new strategy achieves higher classification accuracy for knowledge states and higher precision in ability estimation, while effectively balancing uniform item bank utilization and enabling rapid real-time response. Moreover, it is less affected by cognitive diagnostic models and examinee knowledge state distributions, making it suitable for practical testing scenarios with mixed item banks containing multiple cognitive diagnostic models.

**Keywords:** cognitive diagnosis; item response theory; Gini index; dual-objective CD-CAT; item selection strategy

## 1 Introduction

Summative evaluation uses a continuous scalar (often called latent trait or ability) to characterize student learning outcomes at a particular stage. Computerized adaptive testing (CAT) based on item response theory (IRT) can implement summative assessment more efficiently through a “tailored” approach. Formative evaluation uses a discrete vector (often called latent cognitive pattern or knowledge state) to help teachers understand each student’s latent cognitive status, providing instructional feedback to enable better “teaching according to aptitude,” which benefits both student learning and teacher professional development. CAT based on cognitive diagnostic theory (CDT) uses a “personalized” testing approach to quickly diagnose examinees’ cognitive strengths and weaknesses. Teaching requires the integration of summative and formative evaluation—attending to both process and outcomes to achieve harmony between learning processes and their evaluation. While IRT-CAT focuses on summative evaluation and CD-CAT (cognitive diagnostic computerized adaptive testing) focuses on formative evaluation, their combination in dual-objective CD-CAT (Dual-CAT) can complement their strengths to better achieve testing objectives.

Dual-CAT research centers on two key themes. The first is constructing psychometric indices for the item bank. Just as IRT-CAT relies on item response models (IRM) and CD-CAT relies on cognitive diagnostic models (CDM), Dual-CAT also depends on measurement models that are closely related to the item bank’s psychometric properties. Existing literature shows that only the unified model (also called the fusion model) (Hartz, 2002; Rupp, Templin, & Henson, 2010) and the higher-order model (de la Torre & Douglas, 2004) incorporate both examinee knowledge states and ability within a single model. However, the unified model contains many parameters that are statistically difficult to estimate (Hartz, 2002), making it rarely used in practice. The higher-order model adopts a hierarchical structure that treats latent traits as higher-level general abilities above latent attributes, where the relationship between ability and item response probability is indirectly associated through examinee knowl-

edge states. Only when the number of attributes is relatively large (e.g., greater than 10) can ability be estimated accurately (de la Torre & Douglas, 2004; Hsu & Wang, 2015; Huang, 2020). Consequently, most Dual-CAT item selection research does not rely on these two models but instead employs a separate modeling approach. Using either a unified model or separate modeling determines different construction methods for item selection strategies. For the separate modeling approach, which requires model parameters from both IRM and CDM, establishing connections between these two sets of parameters is fundamental to implementing Dual-CAT.

Research by de la Torre and Douglas (2004) demonstrated that for the same dataset, ability estimates from the higher-order model correlate highly with those from the 2PLM (two-parameter logistic model) in IRT. Wang, Zheng, and Chang (2014) also showed that unidimensional IRT models and the DINA model (Junker & Sijtsma, 2001) can fit the same data well when attributes are highly correlated or linearly hierarchical. Their research supports the separate modeling approach, where stable estimates of both ability and knowledge state can be obtained through a two-step estimation method that considers each model's psychological structure (Kang, Zhang, & Chang, 2017).

The second theme is item selection strategy, which is the key technology for implementing Dual-CAT. An excellent selection strategy should achieve high classification and estimation accuracy to meet testing objectives, ensure relatively uniform item bank utilization to improve test security, and maintain fast computational speed for real-time response. Researchers have proposed various selection strategies for IRT-CAT and CD-CAT, which respectively focus on evaluating latent traits and latent cognitive structures. The challenge lies in effectively combining these two objectives. Scholars have proposed several selection strategies suitable for Dual-CAT, which mainly fall into two categories: the shadow test approach and the combination strategy approach.

McGlohen and Chang (2008) discussed the performance of the shadow test approach alongside single-objective selection strategies for IRT-CAT and CD-CAT under separate modeling: (1) using maximum Fisher information (MFI) (Lord, 1980) or Kullback-Leibler (KL) information (Chang & Ying, 1996) strategies from IRT-CAT to select items matching the examinee's current ability estimate, then estimating knowledge state after testing; (2) using Shannon entropy (SHE) or KL information maximization strategies (Tatsuoka, 2002; Xu, Chang, & Douglas, 2003) from CD-CAT to select items matching the current knowledge state estimate, then estimating ability after testing; and (3) the shadow test approach adapting to both current ability and knowledge state estimates—first constructing an optimal shadow bank based on ability estimate using method (1), then selecting from this shadow bank the item most suitable for the current knowledge state estimate using method (2) as the next candidate item. They compared these three approaches on three metrics: ability estimation accuracy, cognitive state classification accuracy, and item exposure control, with results favoring the shadow test approach.

Du (2010) also adopted the shadow test approach but differed from McGlohen and Chang (2008) by first constructing a shadow bank most suitable for the current knowledge state estimate, then selecting from this shadow bank the item most suitable for the current ability estimate as the next candidate. Comparing ability estimation accuracy and knowledge state classification accuracy across different attribute hierarchical structures, their results also showed the shadow test approach outperformed single-objective strategies.

Some scholars (Cheng, 2007; Dai, Zhang & Li, 2016) have argued that the combination of two “local optimization” steps does not necessarily guarantee “good overall results,” and that more ideal item selection methods should consider both ability and knowledge state simultaneously in a single step to obtain more suitable items, thus proposing combination strategy approaches.

Cheng (2007) and Dai et al. (2016) used linear weighted combinations of indicators for ability ( $I(\hat{\theta})$ ) and knowledge state ( $I(\hat{\alpha})$ ), where  $I(\hat{\theta})$  represents information about ability (such as MFI or KL) and  $I(\hat{\alpha})$  represents information about knowledge state (such as SHE, KL, PWKL (posterior-weighted KL) (Cheng, 2009), MPWKL (modified PWKL) (Kaplan, de la Torre, & Barrada, 2015), and PWACDI (posterior-weighted attribute cognitive discrimination index) (Zheng & Chang, 2016)). Their research showed that on the three metrics of ability estimation accuracy, cognitive state classification accuracy, and item exposure control, composite indicators outperformed the shadow test approach.

Wang, Chang, and Douglas (2012) also used a separate modeling approach, treating diagnosis of examinee knowledge states as a constraint and employing the maximum priority index method (MPI) (Cheng & Chang, 2009) from IRT-CAT for item selection—a multiplicative combination strategy:  $MPI_j = I_j(\hat{\theta}) \times I_j(\hat{\alpha})$ , enabling IRT-CAT to measure both ability and cognitive status. Their research showed that MPI indicators constructed from KL information could achieve good measurement precision.

Overall, combination strategies can more finely characterize the interaction between ability and knowledge state effects on item selection compared to the shadow test approach. The choice between additive and multiplicative combination strategies depends on which information measures are used for  $I(\hat{\theta})$  and  $I(\hat{\alpha})$ . Zheng, He, and Gao (2018) compared various additive and multiplicative combination strategies across different information measures, finding that each combination approach has its own advantages depending on the information measure used.

Additive combination strategy research includes Cheng’s (2007) DIM (dual information method) strategy combining two KL information measures; Wang et al.’s (2014) ASI (aggregate standardized information method) and ARI (aggregate ranked information method) strategies, which eliminate information measure differences through standardization; Kang et al.’s (2017) JSD (Jensen-Shannon divergence) strategy using symmetric KL information; and MASI (modified ASI) and MARI (modified ASI) combining KL and MPWKL information.

Multiplicative combination strategy research includes Wang et al.' s (2012) weighted MPI strategy; Dai et al.' s (2016) DWI (dapperness with information) strategy, which uses logarithmic transformation to eliminate dimensional differences between MFI and SHE information, converting additive to multiplicative combination; and Zheng et al.' s (2018) IPA (information product approach) strategy.

These selection strategies each have their own advantages under certain conditions: some offer high precision but require substantial computation time (e.g., IPA), others have slightly lower precision but enable pre-computation to reduce selection time (e.g., ASI), and some have even lower precision but use less time and provide more uniform item bank utilization (e.g., JSD). Additionally, these strategies may suffer from bias in composite indicators due to large dimensional differences between the two information measures, or from information loss due to transformations designed to eliminate dimensional differences. We aim to develop a new strategy with a unified information metric for both ability and knowledge state that ensures high estimation and classification accuracy while balancing uniform item bank utilization and requiring minimal selection time.

In CD-CAT, Bayesian decision-making is commonly used for examinee classification. The examinee' s knowledge state category is a random variable. When class-conditional probabilities and prior probabilities are known, the posterior probability of an examinee belonging to each category can be calculated through Bayes' formula, and the category with the highest posterior probability is selected as the decision. This approach has been theoretically proven to yield the lowest average error rate (Zhang, 2010, pp. 14-15) and is therefore often called the minimum error rate Bayesian decision. Research shows (Chen, Li, & Xin, 2011; Han et al., 2018; Wang & Chang, 2011) that selection strategies based on posterior probabilities of knowledge state categories (Zheng & Chang, 2016) and those based on posterior probabilities of ability estimation confidence intervals achieve high classification and estimation accuracy, such as the Shannon entropy strategy in CD-CAT (Tatsuoka, 2002; Xu et al., 2003) and the continuous entropy (also called differential entropy) strategy in multidimensional IRT-CAT (Wang & Chang, 2011; Han et al., 2018).

Entropy measures the uncertainty of random variables—greater entropy indicates greater uncertainty. In CD-CAT, entropy measures changes in posterior probabilities of knowledge state categories, and Bayesian decision-making classifies examinees based on these posterior probabilities. Changes in entropy directly reflect changes in category posterior probabilities, making entropy-based selection strategies highly accurate, such as the Shannon entropy strategy (Tatsuoka, 2002; Xu et al., 2003). In statistics, the Gini index is another metric for measuring random variable uncertainty, applied in decision tree classification algorithms. Classic machine learning algorithms include both entropy-based methods like ID3 (Quinlan, 1986) and C4.5 (Quinlan, 1993), and Gini index-based methods like CART (Breiman, Friedman, Stone, & Olshen, 1984) (Zhou, 2016).

This study proposes using the Gini index to construct a selection strategy for dual-objective CD-CAT. The Gini index and entropy share commonalities and differences. Both can measure uncertainty in random variables and handle both continuous and discrete random variables. For a discrete random variable  $Y$  with possible values  $\{y_1, y_2, \dots, y_n\}$  and corresponding probabilities  $\{p_1, p_2, \dots, p_n\}$ , entropy is expressed as:  $\langle MATH_0 \rangle$ , while the Gini index is expressed as:  $\langle MATH_1 \rangle$ .

A first-order Taylor expansion of the entropy formula at  $p = 1$  (ignoring higher-order infinitesimals) yields  $\langle MATH_2 \rangle$ , showing that at extreme points, information entropy and Gini index achieve the same value. Mathematically, entropy uses logarithmic weighting of probabilities, reflecting a nonlinear relationship, while the Gini index uses linear weighting, reflecting a linear relationship.

Since the entropy formula involves logarithmic operations while the Gini index only requires sum of squares, Gini-based selection strategies should achieve classification accuracy comparable to Shannon entropy strategies but with faster computation speed. Moreover, the Gini index's linear weighting makes it more sensitive to changes in category posterior probabilities during testing, helping to expand item selection range and improve item bank utilization.

This paper leverages these advantageous properties of the Gini index to propose a Gini-based selection strategy, expecting it to ensure measurement precision while balancing uniform item bank utilization and enabling rapid real-time response, thereby providing a new and superior method for simultaneously addressing macro-level ability assessment and micro-level cognitive diagnosis.

## 2 Overview of Existing Dual-Objective CD-CAT Item Selection Strategies

We introduce three representative Dual-CAT selection strategies. The ASI strategy represents additive combination strategies, which eliminate dimensional differences between two information measures through standardization before linear weighting. The IPA strategy represents multiplicative combination strategies. The JSD strategy represents the approach with the most uniform item bank utilization and fastest selection speed.

### 2.1 ASI Strategy

Cheng (2009) proposed using the PWKL strategy to replace the KL strategy, significantly improving classification accuracy for examinee knowledge states ( $\alpha$  is a vector of 0s and 1s). Assuming the test measures  $K$  independent attributes, examinees' knowledge states have  $2^K$  categories, and the test aims to classify examinees into one of these categories. The objective function for the PWKL selection strategy is:

$$\langle MATH_3 \rangle$$

where  $R_t$  is the remaining item bank after the examinee has answered  $t$  items,  $j$  is an item in the remaining bank,  $c$  is the category index for knowledge states,  $\alpha_c$  is the  $c$ -th category among  $2^K$  knowledge states,  $P(\alpha_c|\mathbf{X}_t)$  is the posterior probability of category  $\alpha_c$  given response pattern  $\mathbf{X}_t$  on  $t$  items,  $x_{t+1}$  is the examinee's score on item  $j$ ,  $X_j$  represents possible scores on item  $j$  (for dichotomously scored items,  $X_j \in \{0, 1\}$ ),  $\hat{\alpha}$  is the current estimate of the examinee's knowledge state, and  $P_j(\alpha_c)$  is the probability of correctly answering item  $j$  given CDM and known  $\alpha_c$ .

Chang and Ying (1996) used the KL strategy to replace the MFI strategy for measuring examinee ability  $\theta$  (a continuous variable), overcoming the problem of inaccurate ability estimation when few items have been answered. The KL selection strategy objective function is:

$$\langle MATH_4 \rangle$$

where it is recommended that  $\delta = 3/\sqrt{t}$ ,  $t$  is the number of items already answered,  $\hat{\theta}$  is the current ability estimate, and  $P_j(\theta)$  is the probability of correctly answering item  $j$  given IRT's IRM and known  $\theta$ .

Cheng (2007) proposed the DIM selection strategy, linearly combining KL information about  $\alpha$  and KL information about  $\theta$  into a single information measure for dual-objective CD-CAT selection:

$$\langle MATH_5 \rangle$$

where  $\lambda$  is a weight parameter. Wang et al. (2014) replaced the  $\alpha$  information measure in DIM with PWKL information and, recognizing the inconsistent dimensions, proposed standardization to eliminate differences between the two information measures, leading to the ASI strategy:

$$\langle MATH_6 \rangle$$

where  $\overline{PWKL}_{R_t}(\hat{\alpha})$  is the mean PWKL information across all items in the remaining bank  $R_t$  for the current knowledge state estimate  $\hat{\alpha}$ , and  $SD_{PWKL}$  is its standard deviation. Similarly,  $\overline{KL}_{R_t}(\hat{\theta})$  is the mean KL information across all items in  $R_t$  for the current ability estimate  $\hat{\theta}$ , and  $SD_{KL}$  is its standard deviation. Wang et al. (2014) recommended setting the weight  $\lambda = t/L$ , where  $t$  is the number of answered items and  $L$  is the preset test length.

## 2.2 IPA Strategy

Zheng and Chang (2016) proposed the PWACDI (posterior-weighted attribute cognitive discrimination index) selection strategy for short CD-CAT tests. The PWACDI objective function is:

$\langle MATH_7 \rangle$ 

where  $c$  and  $c'$  are knowledge state category indices representing two different categories among  $2^K$  knowledge states, and  $D_j$  is the KL information matrix for  $\alpha$  constructed from item  $j$ , with elements being the expected weighted KL distance between any two knowledge states. “All relevant cells” refers to all elements in the  $D_j$  matrix corresponding to two different knowledge states  $\alpha_c$  and  $\alpha_{c'}$  that differ only in the  $k$ -th attribute value while being identical on other attributes.

The PWACDI selection strategy is independent of the examinee’s current knowledge state estimate  $\hat{\alpha}$  and focuses on distinguishing among  $2^K$  patterns, particularly those with smaller differences, which differs from the PWKL strategy.

Zheng et al. (2018) proposed the IPA strategy for dual-objective CD-CAT, suggesting it provides a unified framework connecting other dual-objective selection strategies. By treating the “weight” as equivalent to IRT information, there is no need to consider the weight in formula (10). The objective function for the multiplicative information strategy is:

 $\langle MATH_8 \rangle$ 

Other CD-CAT selection strategies can be used for  $I(\hat{\alpha})$ . According to Zheng et al. (2018),  $\langle MATH_9 \rangle$  performs better.

### 2.3 JSD Strategy

Kang et al. (2017) proposed the JSD selection strategy, which differs from PWKL in using symmetric KL information:  $\langle MATH_{10} \rangle$ . The JSD objective function is:

 $\langle MATH_{11} \rangle$ 

For clarity, we have supplemented some symbols, so the JSD expression in this paper is not exactly identical to the original (Kang et al., 2017), but the strategy’s meaning remains unchanged.

## 3 Gini Index-Based Dual-Objective CD-CAT Item Selection Strategy

This study defines Gini indices based on posterior probabilities of examinee knowledge state categories and posterior probabilities of ability estimation confidence intervals, combining them to form a new Gini-index-based dual-objective CD-CAT strategy aimed at achieving high precision, high item bank utilization, and rapid feedback.

### 3.1 Gini Index-Based CD-CAT Item Selection Strategy

Assume a test examines  $K$  attributes. After responding to  $t$  items with response pattern  $\mathbf{X}_t$ , the posterior probability of knowledge state category  $\alpha_c$  is  $P(\alpha_c|\mathbf{X}_t)$ . Following the definition of Gini index (Li, 2012), the Gini index for examinee knowledge state category posterior probabilities is defined as:

$$\langle MATH_{12} \rangle$$

where  $\langle MATH_{13} \rangle$  is the set of posterior probabilities for knowledge state categories given response pattern  $\mathbf{X}_t$  on  $t$  items,  $P(\alpha_c)$  is the prior probability of category  $\alpha_c$  (generally  $1/2^K$ ),  $P_j(\alpha_c)$  is the probability of correctly answering item  $j$  given CDM and knowledge state  $\alpha_c$ , and  $x_{t+1}$  is the examinee's score on item  $j$ . Other symbols are as defined in Section 2.

$\langle MATH_{14} \rangle$  characterizes the dispersion of posterior probability distribution for knowledge state categories given response pattern  $\mathbf{X}_t$ . A smaller value indicates more concentrated probability distribution, meaning one or some categories have posterior probabilities much larger than others, thereby improving classification accuracy through Bayesian decision-making. The algorithm iterates through the remaining item bank and selects item  $j$  that minimizes  $\langle MATH_{15} \rangle$  as the next candidate.

Since the examinee's response  $x_{t+1}$  to candidate item  $j$  is unknown, and for dichotomously scored items  $x_{t+1}$  takes values 0 or 1, we define the expected Gini index for knowledge state category posterior probabilities:

$$\langle MATH_{16} \rangle$$

By the law of total probability,  $\langle MATH_{17} \rangle$ , the objective function for the selection strategy is:

$$\langle MATH_{18} \rangle$$

where  $R_t$  is the remaining item bank for the examinee. The next item  $j$  is selected from the remaining bank as the one with the minimum  $\langle MATH_{19} \rangle$ .

### 3.2 Gini Index-Based IRT-CAT Item Selection Strategy

In the initial stage of IRT-CAT, the examinee's current ability estimate  $\hat{\theta}$  often deviates substantially from the true ability value, making Fisher information based on  $\hat{\theta}$  a poor indicator of test efficiency and thus not playing a significant role initially (Chang & Ying, 1996). Veerkamp and Berger (1994) proposed using the maximum average information function over a confidence interval instead of point-based information, which better overcomes inefficient item selection due to estimation inaccuracy.

Good selection strategies cause ability estimates  $\hat{\theta}$  to converge toward true values as testing progresses. Following definitions of KL global information and continuous entropy in Chang and Ying (1996) and Wang and Chang (2011), we define the Gini index based on posterior probabilities of ability estimate confidence intervals, similar to KL global information but using interval information instead of point information. Let  $\langle MATH_{20} \rangle$  be the set of posterior probabilities for ability estimate  $\theta$  within a confidence interval given response pattern  $\mathbf{X}_t$  on  $t$  items. Chang and Ying (1996) recommend  $\langle MATH_{21} \rangle$ . Following BILOG' s recommendation for calculating posterior expected probabilities, we set the number of quadrature nodes to a natural number close to  $2\sqrt{t}$ , with  $\langle MATH_{22} \rangle$ , where  $\langle MATH_{23} \rangle$  represents ceiling rounding.  $g(\theta)$  is the prior probability of ability; if prior information is unknown, a uniform distribution is used.

$P_j(\theta)$  is the probability of correctly answering item  $j$  given IRM and ability  $\theta$ . Other symbols are as defined in Section 2. The algorithm iterates through the remaining item bank and selects item  $j$  that minimizes  $\langle MATH_{24} \rangle$  as the next candidate.

Since the examinee' s response  $x_{t+1}$  to candidate item  $j$  is unknown, and for dichotomously scored items  $x_{t+1}$  takes values 0 or 1, we define the expected Gini coefficient for ability estimate confidence interval posterior probabilities:

$$\langle MATH_{25} \rangle$$

The objective function for the selection strategy is:

$$\langle MATH_{26} \rangle$$

where  $R_t$  is the remaining item bank. The next item  $j$  is selected as the one with the minimum  $\langle MATH_{27} \rangle$ .

### 3.3 Combination Strategy

Cheng (2007) and Wang et al. (2014) proposed weighted linear combination of KL information functions for knowledge state  $\alpha$  and ability  $\theta$  to obtain single-information-form dual-objective selection strategies, as shown in formulas (8) and (10). Zheng et al. (2018) proposed a multiplicative dual-objective selection strategy, as shown in formula (16). Since multiplication operations are more time-consuming, we adopt the linear weighted sum approach from Cheng (2007) and Wang et al. (2014) to obtain the Gini-based dual-objective selection strategy objective function.

The proposed new strategy constructs novel indices based on Gini indices of two random variables' posterior probabilities. Since each random variable' s posterior probability ranges in  $[0, 1]$  and the sum of posterior probabilities equals 1, the

dimensional differences between Gini indices constructed from these posterior probabilities will not be large. Unlike Wang et al. (2014), who standardized two KL information measures before linear combination (which causes information loss), the new strategy's synthesis method preserves original information.

The Gini selection strategy objective function is:

$$\langle MATH_{28} \rangle$$

where  $\lambda$  is a weight parameter. Following Wang et al.'s (2014) recommendation for high-quality item banks, we use the theoretical weight  $\lambda = t/L$ , where  $t$  is the number of answered items and  $L$  is the preset test length.

## 4 Simulation Experiment Design

To examine the new strategy's performance under different CDMs, examinee knowledge state distributions, and test lengths, and to compare it with other strategies, we conducted Monte Carlo simulation studies. The experiment investigated the new strategy's performance across 36 conditions: 3 CDMs (G-DINA, DINA, R-RUM)  $\times$  3 examinee knowledge state distributions (higher-order model, high-correlation multivariate normal model, low-correlation multivariate normal model)  $\times$  4 test lengths (5, 10, 15, 20 items).

### 4.1 Cognitive Diagnostic Models

We examined selection strategy performance under the saturated G-DINA model (de la Torre, 2011) and reduced models (DINA, R-RUM) (Hartz, 2002; Junker & Sijtsma, 2001). The G-DINA model can be simplified to different reduced models under appropriate constraints: if all main effects and lower-order interaction effects in G-DINA are zero, it reduces to the DINA model; if all interaction effects in the log-link function are zero, it yields the R-RUM.

### 4.2 Item Bank and Examinee Simulation

**4.2.1 Simulating Item Attribute Vectors for the Item Bank** Assume the item bank examines 5 independent attributes. Each item examines at most 3 attributes, resulting in 25 possible item attribute vectors. Replicating each attribute vector 10 times yields 250 items in the item bank.

**4.2.2 Simulating True Examinee Knowledge States** Examinee knowledge states were simulated using two approaches: the HO-CDM (Wang et al., 2012, 2014; Huang, 2020) and multivariate normal distribution generation (Dai et al., 2016; Kang et al., 2017). Both approaches were examined because their response data can simultaneously fit both CDM and IRT models, making them common simulation methods in dual-objective CD-CAT research.

- (1) **HO-CDM generation:** The higher-order model assumes that whether an examinee masters an attribute is related to a generalized latent ability. Through a logit link, the probability of an examinee with higher-order ability  $\theta$  mastering attribute  $k$  is defined as  $\langle MATH_{29} \rangle$ , similar to the 2PLM in IRT, where  $a_k$  and  $b_k$  are discrimination and location parameters ( $a_k$  constrained to  $[0.2, 2.5]$ ),  $b_k \sim N(0, 1)$ . Random numbers  $u_k \sim U(0, 1)$  are generated, and if  $\langle MATH_{30} \rangle$ , then  $\alpha_k = 1$ , otherwise  $\alpha_k = 0$  (Ma & de la Torre, 2020).
- (2) **Multivariate normal model generation:** A multivariate normal threshold model (mean vector  $\mathbf{0}$ ; correlation coefficients set at 0.8 and 0.2 representing high and low attribute correlations) generates examinees' true attribute mastery patterns, with 0 as the cutoff to obtain discrete knowledge states (Ma & de la Torre, 2020).

#### 4.2.3 Simulating CDM and IRT Model Parameters for the Item Bank

We used the separate modeling approach described in Section 1 to construct the item bank, employing G-DINA, DINA, and R-RUM as CDMs, and 2PLM as the IRT model—models frequently used in research and practice. Item bank parameters were simulated and estimated using the GDINA and mirt packages in R.

Using G-DINA with high-correlation multivariate normal knowledge state generation as an example, item parameter simulation proceeded as follows: (1) According to the GDINA package documentation (Ma & de la Torre, 2020), CDM parameters can be set using three methods. The first method specifies guessing parameter  $g_j$  and slipping parameter  $s_j$  for each item, where  $g_j$  represents the probability of correct response for examinees who haven't mastered any examined attributes, and  $s_j$  represents the probability of correct response for examinees who have mastered all examined attributes. For other examinee types, correct response probabilities are generated from  $[g_j, 1 - s_j]$  following the monotonicity constraint that examinees mastering more attributes have higher correct response probabilities. The second method specifies correct response probabilities for each knowledge state on each item. The third method specifies delta parameters in the G-DINA model.

We adopted the first method for its simplicity, using the simGDINA function to simulate G-DINA item parameters with  $g_j, s_j \sim U(0.05, 0.25)$ . Correct response probabilities for examinees mastering some but not all attributes were generated from  $[g_j, 1 - s_j]$  while maintaining monotonicity. (2) Since 2PLM parameter estimation requires over 1000 samples for good precision, we simulated 3000 examinee knowledge states using the high-correlation multivariate normal model. Based on known item attribute vectors and G-DINA parameters, we obtained each examinee's correct response probability  $P_{ij}$  for each item, generated random numbers  $u_{ij} \sim U(0, 1)$ , and set  $x_{ij} = 1$  if  $u_{ij} \leq P_{ij}$ , otherwise  $x_{ij} = 0$ , yielding a complete  $3000 \times 250$  response matrix (Wang et al., 2012, 2014). The mirt package (Chalmers, 2012) fitted 2PLM to this matrix to obtain discrimination

and difficulty parameters for all 250 items, and the GDINA package refined G-DINA model parameters for greater accuracy.

This procedure yielded 9 item banks: 3 (G-DINA, DINA, R-RUM)  $\times$  3 (higher-order, high-correlation multivariate normal, low-correlation multivariate normal) combinations of CDM parameters and 2PLM parameters.

**4.2.4 Simulating True Examinee Ability** Examinee responses were simulated based on CDM models. Response data for all items in the bank were generated, and examinee ability values were estimated using the expected a posteriori algorithm (Bock & Mislevy, 1982) to serve as true values (Wang et al., 2012, 2014; Dai et al., 2016; Kang et al., 2017).

### 4.3 Selection Strategies

The DIM strategy (Cheng, 2007) was the first to linearly combine two KL information measures. The ASI strategy standardizes both information measures to eliminate dimensional differences before linear combination, and Wang et al. (2014) showed ASI outperforms DIM. Zheng et al. (2016, 2018) found PWACDI superior to PWKL for short tests in classification accuracy, and that *PWACDIKL and DWI (Dai et al., 2016) belong to the multiplicative combination IPA family, with PWACDIKL performing better among IPA strategies.* The JSD strategy (Kang et al., 2017) uses symmetric KL information based on current knowledge state and ability estimates, requiring no integration operations and thus enabling simple, fast selection. Kang et al. (2017) showed JSD has significant advantages in selection time and item bank utilization uniformity compared to other strategies.

This paper compares the Gini strategy with ASI (Wang et al., 2014), representative IPA strategy PWACDI\*KL (Zheng et al., 2018), and JSD (Kang et al., 2017) across 9 item banks, evaluating measurement precision (including knowledge state classification accuracy and ability estimation accuracy), item bank utilization uniformity, and selection time.

### 4.5 Evaluation Metrics

**4.5.1 Knowledge State Classification Accuracy** Pattern correct classification rate measures classification accuracy, with higher values indicating better accuracy:

$$\langle MATH_{31} \rangle$$

where  $I(\cdot)$  counts 1 when the condition is TRUE and 0 otherwise,  $N$  is the number of examinees,  $\hat{\alpha}_i$  is the estimated knowledge state, and  $\alpha_i$  is the true knowledge state.

**4.5.2 Ability Estimation Accuracy** Bias and RMSE measure ability estimation accuracy, with smaller values indicating better parameter recovery:

$$\langle MATH_{32} \rangle$$

$$\langle MATH_{33} \rangle$$

where  $\hat{\theta}_i$  is the estimated ability and  $\theta_i$  is the true ability. Other variables are as defined above.

**4.5.3 Item Bank Utilization Uniformity** Chi-square value and test overlap rate are important metrics for evaluating item bank utilization uniformity, with smaller values indicating more uniform usage and higher utilization:

**Chi-square metric:**

$$\langle MATH_{34} \rangle$$

**Test overlap rate:**

$$\langle MATH_{35} \rangle$$

where  $m_j$  is the exposure frequency of item  $j$ ,  $L$  is the item bank size, and  $TL$  is the preset test length. Other variables are as defined above.

**4.5.4 Item Selection Time**

$$\langle MATH_{36} \rangle$$

where  $T_i$  is the time (in seconds) for examinee  $i$  to complete the test. Since simulation time consumption is mainly in item selection (other times are negligible),  $TC$  represents selection time consumption. Smaller values indicate faster selection.

## 4.6 CAT Implementation

The entire CAT program ran in Python 3 on a 4-core Intel Core i5 1.9GHz processor with 8GB RAM. Using the condition with G-DINA model and high-correlation multivariate normal knowledge state generation as an example, the CAT implementation proceeded as follows:

- (1) Select the corresponding item bank built with GDINA and mirt packages in R (details in Section 4.2);

- (2) Simulate examinee knowledge states using the high-correlation multivariate normal model as true values, simulate responses to all bank items using G-DINA, and estimate ability using expected a posteriori as true ability (details in Section 4.2);
- (3) Randomly assign 3 items to each examinee, then estimate initial knowledge state and ability based on these 3 responses;
- (4) Implement separate CAT processes using Gini, ASI, IPA, and JSD strategies, updating knowledge state estimates via maximum a posteriori and ability estimates via expected a posteriori after each item;
- (5) Repeat step (4) until test termination criteria are met;
- (6) After testing, calculate evaluation metrics from Section 4.5 based on final knowledge state and ability estimates for each strategy.

To eliminate random effects, each condition simulated 1000 examinees with 10 replications, calculating mean values of evaluation metrics (see tables in Section 5; SD denotes standard deviation).

## 5 Results

### 5.1 Classification Accuracy Comparison

Table 1 shows that pattern correct classification rates for Gini and IPA strategies are substantially higher than for ASI and JSD strategies. Overall, Gini's classification rate is slightly higher than IPA's, with both exceeding 95% across experimental conditions and showing small standard deviations, indicating stable and reliable classification results applicable to mixed item banks with various CDMs.

Figure 1 [Figure 1: see original paper] displays performance across test lengths. As test length increases, all strategies' classification rates improve. Gini and IPA show very similar curves with the fastest growth, maintaining the best classification rates. For short tests (TL < 15), Gini, IPA, and ASI have similar classification rates, but for medium and long tests (TL > 15), ASI's growth rate is lower than the other two. Consistent with Table 1, Gini and IPA show minimal variation across conditions, achieving good classification accuracy in both short and medium-long tests.

**Table 1** Mean and standard deviation of pattern correct classification rates for 20-item tests

Model	Knowledge State Generation	Gini	ASI	IPA	JSD
G-DINA	HO	97.00%	89.28%	96.10%	85.04%
	MV-0.8	97.22%	93.05%	97.44%	92.02%

Model	Knowledge State Generation	Gini	ASI	IPA	JSD
DINA	MV-0.2	96.84%	90.78%	96.35%	87.51%
	HO	97.45%	90.99%	97.18%	75.31%
	MV-0.8	97.24%	93.45%	97.06%	91.46%
R-RUM	MV-0.2	97.57%	93.76%	96.93%	86.23%
	HO	95.41%	87.61%	95.38%	76.64%
	MV-0.8	97.09%	92.45%	96.82%	91.67%
	MV-0.2	96.81%	87.88%	96.82%	80.52%

*Note: HO = knowledge states generated by HO-CDM; MV-0.8 = knowledge states generated by multivariate normal model with attribute correlation 0.8; MV-0.2 = knowledge states generated by multivariate normal model with attribute correlation 0.2.*

## 5.2 Ability Estimation Accuracy Comparison

Table 2 shows that except for the low attribute correlation condition under the DINA model, all four strategies produce essentially unbiased ability estimates. ASI shows the smallest estimation bias, followed by Gini. IPA has the smallest RMSE, with Gini slightly behind but with maximum difference only 0.04. When attributes are highly correlated, all four strategies have very similar RMSE values (maximum difference 0.03), but under other conditions, maximum differences reach 0.22. This indicates that all four strategies are suitable when attributes are highly correlated, while IPA and Gini are preferred under other conditions. Gini and IPA's ability estimation precision is related to CDM, with Gini being less affected. ASI and JSD's precision is affected by both CDM and knowledge state distribution.

Figure 2 [Figure 2: see original paper] shows that RMSE decreases as test length increases, indicating improving parameter estimation precision. Gini and IPA show the fastest RMSE decline with nearly identical curves, while JSD shows the slowest decline. Under high attribute correlation, all four strategies' curves essentially overlap. Under other conditions, similar to Figure 1, Gini, IPA, and ASI show similar performance in short tests ( $TL < 15$ ), but ASI lags behind in medium-long tests ( $TL > 15$ ). Thus, Gini and IPA achieve good ability estimation precision in both short and medium-long tests.

**Table 2** Bias and RMSE for 20-item tests

Model	Knowledge State Generation	Gini	Gini	ASI	ASI	IPA	IPA	JSD	JSD
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
G-DINA	HO	-0.02	0.31	-	0.33	-	0.29	-	0.38
				0.01		0.02		0.03	

Knowledge State		Gini	Gini	ASI	ASI	IPA	IPA	JSD	JSD	
Model	Generation	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	
DINAHO	MV-0.8	-0.01	0.28	-	0.29	-	0.27	-	0.30	
				0.01		0.01		0.02		
	MV-0.2	-0.02	0.32	-	0.34	-	0.31	-	0.40	
				0.01		0.02		0.04		
			-0.03	0.35	-	0.36	-	0.33	-	0.51
					0.01		0.03		0.05	
R-RUM	MV-0.8	-0.01	0.29	-	0.30	-	0.28	-	0.31	
				0.01		0.01		0.02		
	MV-0.2	-0.02	0.33	-	0.35	-	0.32	-	0.45	
				0.01		0.02		0.05		
	HO	-0.02	0.34	-	0.35	-	0.32	-	0.48	
					0.01		0.02		0.04	
	MV-0.8	-0.01	0.29	-	0.30	-	0.28	-	0.31	
				0.01		0.01		0.02		
	MV-0.2	-0.02	0.33	-	0.35	-	0.32	-	0.42	
				0.01		0.02		0.05		

### 5.3 Item Bank Utilization Uniformity Comparison

Table 3 shows that JSD has the best item bank utilization uniformity, followed by Gini and IPA (which are similar), with both outperforming ASI. Item bank utilization metrics for all four strategies are related to both CDM and knowledge state distribution, though Gini is slightly less affected. Under high attribute correlation in the DINA model, the four strategies' metrics are relatively close, but differ substantially under other conditions.

Figure 3 [Figure 3: see original paper] shows that chi-square values decrease as test length increases, indicating improving item bank utilization uniformity. All strategies show similar curve patterns across conditions, with JSD showing the best decline curve, followed by Gini. Under high attribute correlation in the DINA model, all four strategies' decline curves essentially overlap.

**Table 3** Item bank utilization uniformity metrics for 20-item tests

Knowledge State		Gini	ASI	IPA	JSD	Gini	ASI	IPA	JSD
Mode	Generation	$\chi^2$	$\chi^2$	$\chi^2$	$\chi^2$	TO	TO	TO	TO
G-DINA	HO	12.3	15.7	12.8	8.9	0.23	0.31	0.25	0.18
	MV-0.8	11.8	14.2	12.1	8.5	0.22	0.28	0.24	0.17
	MV-0.2	13.1	16.8	13.5	9.2	0.25	0.33	0.27	0.19
DINAHO		14.5	18.3	15.1	9.8	0.28	0.36	0.30	0.20

Model	Knowledge State Generation	Gini $\chi^2$	ASI $\chi^2$	IPA $\chi^2$	JSD $\chi^2$	Gini TO	ASI TO	IPA TO	JSD TO
		MV-0.8	12.0	14.5	12.3	8.7	0.23	0.29	0.25
	MV-0.2	15.2	19.7	15.8	10.5	0.30	0.39	0.32	0.22
R-RUM	HO	13.8	17.1	14.2	9.4	0.26	0.34	0.28	0.19
	MV-0.8	11.9	14.3	12.2	8.6	0.23	0.28	0.24	0.17
	MV-0.2	14.7	18.9	15.3	10.1	0.29	0.38	0.31	0.21

#### 5.4 Item Selection Time Comparison

Table 4 shows JSD has the shortest selection time, followed by ASI, then Gini, with IPA being the most time-consuming (approximately 10 times longer than Gini). Selection time remains essentially constant across conditions for each strategy, as it primarily depends on algorithm computational complexity, number of attributes, and item bank capacity. With fixed attribute count and known bank capacity, algorithm computational complexity is the determining factor.

**Table 4** Item selection time metrics for 20-item tests (in seconds)

Model	Knowledge State Generation	Gini	ASI	IPA	JSD
G-DINA	HO	0.12	0.04	1.15	0.02
	MV-0.8	0.11	0.04	1.13	0.02
	MV-0.2	0.12	0.04	1.16	0.02
DINA	HO	0.13	0.04	1.18	0.02
	MV-0.8	0.11	0.04	1.14	0.02
	MV-0.2	0.12	0.04	1.17	0.02
R-RUM	HO	0.12	0.04	1.16	0.02
	MV-0.8	0.11	0.04	1.13	0.02
	MV-0.2	0.12	0.04	1.15	0.02

## 6 Discussion and Conclusion

### 6.1 Summary

This paper leverages the advantageous properties of the Gini index to construct a novel dual-objective CD-CAT selection strategy. Simulation experiments demonstrate that the new strategy achieves high measurement precision, balances item bank utilization uniformity, and enables rapid real-time response, providing a new and superior method for simultaneously addressing macro-level ability assessment and micro-level cognitive diagnosis.

The experiment examined four dual-objective selection strategies (Gini, ASI, IPA, and JSD) under three CDMs and three knowledge state distributions.

Overall conclusions are: (1) Gini and IPA strategies show similar performance on classification accuracy, ability estimation precision, and item bank utilization uniformity metrics, with high measurement precision and low sensitivity to CDM and knowledge state distribution, making them suitable for practical mixed-model item banks. Generally, Gini slightly outperforms IPA, with selection time only one-tenth of IPA's; (2) Both Gini and ASI are linear weighted combination strategies. In short tests, their measurement precision is very close, but in medium-long tests, although ASI's selection time is one-third of Gini's, its measurement precision and item bank utilization uniformity are inferior to Gini's; (3) Compared to JSD, while JSD has advantages in item bank utilization uniformity and selection time, its measurement precision is far lower than Gini's.

In summary, for short tests, Gini, IPA, and ASI all demonstrate good measurement precision and are recommended. For medium-long tests with few attributes and small item banks, Gini and IPA are recommended. As attribute count and bank size increase, Gini is recommended. When attributes are highly correlated and both attribute count and bank size are very large, ASI and JSD are recommended, with ASI showing slightly higher precision than JSD.

## 6.2 Discussion

The Gini strategy is constructed based on posterior probabilities of knowledge state categories and ability estimation confidence intervals, making it less affected by CDM and knowledge state distribution. This construction directly reflects posterior probability changes and employs minimum error rate Bayesian decision-making for knowledge state determination, resulting in high measurement precision. The Gini index's linear weighting makes it more sensitive to posterior probability changes than entropy, helping expand selection range and improve item bank utilization uniformity. Additionally, additive operations are fast enough to meet Dual-CAT's real-time response requirements.

Under some conditions (e.g., knowledge states generated by higher-order models), Gini's ability estimation precision may be slightly lower than IPA's, while its classification accuracy may be slightly higher—likely due to the combined strategy's balancing of ability and knowledge state information during item selection. Zheng and Chang (2016) noted that when item bank parameters are known, KL information in formula (3) can be pre-computed, reducing ASI's selection time. However, since Gini is defined on random variable posterior probabilities, it must be computed in real-time based on examinee responses, slightly increasing selection time.

JSD only calculates KL distance based on current estimates, involving minimal computation and very fast selection. In contrast, Gini must consider posterior probability changes over finite sets and intervals, requiring summation and integration operations, making its selection time longer than ASI and JSD. In short tests, ability and knowledge state estimates deviate substantially from true

values, making JSD' s selection range broader and item bank utilization more uniform. Gini does not depend on current ability and knowledge state estimates but on their probability distributions, making selections more concentrated.

While Gini achieves high measurement precision, its item bank utilization is inferior to JSD' s. Wang, Chang, and Huebner (2011) showed that restrictive progressive method (RP) and restrictive threshold method (RT) can balance measurement precision and item exposure rates. Future research will combine Gini with RP and RT methods to improve its item bank utilization uniformity. Measurement precision and item bank utilization uniformity are conflicting objectives; exposure control techniques improve uniformity but may reduce measurement precision, requiring further investigation of this trade-off. Additionally, whether differences between strategies disappear after applying exposure control techniques warrants further study. When attribute count and item bank size are large, Gini' s selection time may exceed user expectations (delay > 2 seconds) (Nah, 2004). Future research will combine Gini with dynamic search algorithms (Zheng & Wang, 2017) to optimize and reduce selection time.

This paper uses separate modeling to obtain parameters for both model types to build the Dual-CAT item bank. Whether items fully fit the target models requires further investigation for more accurate measurement results. The Dual-CAT item bank construction process simulated CDM parameters and item attribute vectors first, generated response data based on CDM, then estimated IRT parameters from these responses—a common approach in current research (Dai et al., 2016; Kang et al., 2017; Wang et al., 2012, 2014). An alternative approach of first simulating IRT parameters, generating responses based on IRT, then estimating CDM parameters and item attribute vectors from these responses deserves exploration to examine strategy performance under this bank construction method.

As test data complexity and testing requirements evolve, selection strategies must adapt to new test formats, such as polytomous attribute tests (Tu & Cai, 2015), polytomously scored item tests (Cai, Miao, & Tu, 2016), multidimensional tests (Han et al., 2018; Hsu & Wang, 2019), multistage CD-CAT (Luo, Wang, Ding, & Xiong, 2018; Kaplan & de la Torre, 2020), multistage tests incorporating non-statistical constraints (Lin & Chang, 2019; Liu, Cai, & Tu, 2018), and CAT incorporating response times (Fan, Wang, Chang, & Douglas, 2012; Huang, 2020). Future research can explore Gini-based selection strategies in these testing contexts.

## References

- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6(4), 431-444.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Chapman & Hall / CRC, Boca Raton, FL.

- Cai, Y., Miao, Y., & Tu, D. B. (2016). The polytomously scored cognitive diagnosis computerized adaptive testing. *Acta Psychologica Sinica*, 48(10), 1338-1346. [蔡艳, 苗莹, 涂冬波. (2016). 多级评分的认知诊断计算机化自适应测验. 心理学报, 48(10), 1338-1346.]
- Chalmers, R. P. (2012). Mirt: a multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1-29.
- Chang, H. H., & Ying, Z. L. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20(3), 213-229.
- Chen, P., Li, Z., & Xin, T. (2011). A note on the uniformity of item bank usage in cognitive diagnostic computerized adaptive testing. *Studies of Psychology and Behavior*, 37(1), 212-216. [陈平, 李珍, 辛涛. (2011). 认知诊断计算机化自适应测验的题库使用均匀性初探. 心理与行为研究, 37(1), 212-216.]
- Cheng, Y. (2007). *The dual information method for item selection in cognitive diagnostic computerized adaptive testing* (Unpublished Master's thesis). University of Illinois at Urbana-Champaign.
- Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing. *Psychometrika*, 74(4), 619-632.
- Cheng, Y., & Chang, H. H. (2009). The maximum priority index method for severely constrained item selection in computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, 62(2), 369-383.
- Dai, B. Y., Zhang, M. Q., & Li, G. M. (2016). Exploration of item selection in dual purpose cognitive diagnostic computerized adaptive testing: Based on the RRUM. *Applied Psychological Measurement*, 40(8), 625-640.
- Du, X. X. (2010). *A new strategy of item selection of cognitive diagnosis computerized adaptive testing* (Unpublished Master's thesis). Jiangxi Normal University, Nanchang, China. [杜宣宣. (2010). 具有认知诊断功能的计算机化自适应测验的选题策略研究 (硕士学位论文). 江西师范大学, 南昌.]
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2), 179-199.
- de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333-353.
- Fan, Z., Wang, C., Chang, H. H., & Douglas, J. (2012). Utilizing response time distributions for item selection in CAT. *Journal of Educational and Behavioral Statistics*, 37(5), 655-670.
- Han, Y. T., Gao, X. L., Wang, D. X., Cai, Y., & Tu, D. B. (2018). Item selection methods in multidimensional polytomous computerized adaptive testing. *Journal of Psychological Science*, 41(6), 1500-1507. [韩雨婷, 高旭亮, 汪大勋, 蔡艳, 涂冬波. (2018). 多级评分项目的多维 CAT 选题策略开发. 心理科学, 41(6), 1500-1507.]

- Hartz, S. M. (2002). *A bayesian framework for the unified model for assessing cognitive abilities: blending theory with practicality* (Unpublished Doctoral dissertation). University of Illinois at Urbana-Champaign, Urbana-Champaign, IL.
- Hsu, C. L., & Wang, W. C. (2015). Variable-length computerized adaptive testing using the higher order DINA model. *Journal of Educational Measurement*, *52*(2), 125-143.
- Hsu, C. L., & Wang, W. C. (2019). Multidimensional computerized adaptive testing using non-compensatory item response theory models. *Applied Psychological Measurement*, *43*(6), 464-480.
- Huang, H. Y. (2020). Utilizing response times in cognitive diagnostic computerized adaptive testing under the higher-order deterministic input, noisy 'and' gate model. *British Journal of Mathematical and Statistical Psychology*, *73*(1), 109-141.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258-272.
- Kang, H. A., Zhang, S. S., & Chang, H. H. (2017). Dual-objective item selection criteria in cognitive diagnostic computerized adaptive testing. *Journal of Educational Measurement*, *54*(2), 165-183.
- Kaplan, M., & de la Torre, J. (2020). A blocked-CAT procedure for CD-CAT. *Applied Psychological Measurement*, *44*(1), 49-64.
- Kaplan, M., de la Torre, J., & Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, *39*(3), 167-188.
- Li, H. (2012). *Statistical learning method*. Beijing: Tsinghua University Press. [李航. (2012). 统计学习方法. 北京: 清华大学出版社.]
- Lin, C. J., & Chang, H. H. (2019). Item selection criteria with practical constraints in cognitive diagnostic computerized adaptive testing. *Educational and Psychological Measurement*, *79*(2), 335-357.
- Liu, S. C., Cai, Y., & Tu, D. B. (2018). On-the-fly constraint-controlled assembly methods for multistage adaptive testing for cognitive diagnosis. *Journal of Educational Measurement*, *55*(4), 595-613.
- Lord, M. F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale NJ: Erlbaum.
- Luo, F., Wang, X. Q., Ding, S. L., Xiong, J. H. (2018). The design and selection strategies of adaptive multigroup Testing for Cognitive Diagnosis. *Journal of Psychological Science*, *41*(3), 720-726. [罗芬, 王晓庆, 丁树良, 熊建华. (2018). 自适应分组认知诊断测验设计及其选题策略. 心理科学, 41(3), 720-726.]

- Ma, W., & de la Torre, J. (2020). GDINA: The generalized DINA model framework. R package version 2.7.9, <https://CRAN.R-project.org/package=GDINA>.
- McGlohen, M. K., & Chang, H. H. (2008). Combining computer adaptive testing technology with cognitively diagnostic assessment. *Behavior Research Methods*, 40(3), 808-821.
- Nah, F. F. H. (2004). A study on tolerable waiting time: how long are web users willing to wait? *Behaviour and Information Technology*, 23(3), 153-163.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.
- Quinlan, J. R. (1993). *C4.5: programs for machine learning*. Morgan Kaufmann, San Mateo, CA.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: theory, method, and application*. New York: The Guilford Press.
- Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 51(3), 337-350.
- Tu, D. B., & Cai, Y. (2015). The Development of CD-CAT with polytomous attributes. *Acta Psychologica Sinica*, 47(11), 1405-1414. [涂冬波, 蔡艳. (2015). 基于属性多级化的认知诊断计算机化自适应测验设计与实现. *心理学报*, 47(11), 1405-1414.]
- Veerkamp, W. J., & Berger, M. P. F. (1994). *Some new item selection criteria for adaptive testing* (Research Rep. 94-6). Enschede, The Netherlands: University of Twente, Department of Educational Measurement and Data Analysis.
- Wang, C., & Chang, H. H. (2011). Item selection in multidimensional computerized adaptive testing-gaining information from different angles. *Psychometrika*, 76(3), 363-384.
- Wang, C., Chang, H. H., & Huebner, A. (2011). Restrictive stochastic item selection methods in cognitive diagnostic computerized adaptive testing. *Journal of Educational Measurement*, 48(3), 255-273.
- Wang, C., Chang, H. H., & Douglas, J. (2012). Combining CAT with cognitive diagnosis: a weighted item selection approach. *Behavior Research Methods*, 44(1), 95-109.
- Wang, C., Zheng, C. J., & Chang, H. H. (2014). An enhanced approach to combine item response theory with cognitive diagnosis in adaptive testing. *Journal of Educational Measurement*, 51(4), 358-380.
- Xu, X. L., Chang, H. H., & Douglas, J. (2003, April). *A simulation study to compare CAT strategies for cognitive diagnosis*. Paper presented at the annual meeting of National Council on Measurement in Education, Chicago, IL.
- Zhang, X. G. (2010). *Pattern recognition* (Third Edition). Beijing: Tsinghua University Press. [张学工. (2010). 模式识别 (第三版). 北京: 清华大学出版社.]

Zheng, C. J., & Chang, H. H. (2016). High-efficiency response distribution-based item selection algorithms for short-length cognitive diagnostic computerized adaptive testing. *Applied Psychological Measurement*, *40*(8), 608-624.

Zheng, C. J., He, G., & Gao, C. L. (2018). The information product methods: a unified approach to dual-purpose computerized adaptive testing. *Applied Psychological Measurement*, *42*(4), 321-324.

Zheng, C. J., & Wang, C. (2017). Application of binary searching for item exposure control in cognitive diagnostic computerized adaptive testing. *Applied Psychological Measurement*, *41*(7), 561-576.

Zhou, Z. H. (2016). *Machine learning*. Beijing: Tsinghua University Press. [周志华. (2016). 机器学习. 北京: 清华大学出版社.]

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

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