# High-Efficiency Response Distribution–Based Item Selection Algorithms for Short-Length Cognitive Diagnostic Computerized Adaptive Testing

**Chanjin Zheng[1,2] and Hua-Hua Chang[2]**

## Abstract

Cognitive diagnostic computerized adaptive testing (CD-CAT) purports to obtain useful diagnostic information with great efficiency brought by CAT technology. Most of the existing CD-CAT item selection algorithms are evaluated when test length is fixed and relatively long, but some applications of CD-CAT, such as in interim assessment, require to obtain the cognitive pattern with a short test. The mutual information (MI) algorithm proposed by Wang is the first endeavor to accommodate this need. To reduce the computational burden, Wang provided a simplified scheme, but at the price of scale/sign change in the original index. As a result, it is very difficult to combine it with some popular constraint management methods. The current study proposes two high-efficiency algorithms, posterior-weighted cognitive diagnostic model (CDM) discrimination index (PWCDI) and posterior-weighted attribute-level CDM discrimination index (PWACDI), by modifying the CDM discrimination index. They can be considered as an extension of the Kullback–Leibler (KL) and posterior-weighted KL (PWKL) methods. A pre-calculation strategy has also been developed to address the computational issue. Simulation studies indicate that the newly developed methods can produce results comparable with or better than the MI and PWKL in both short and long tests. The other major advantage is that the computational issue has been addressed more elegantly than MI. PWCDI and PWACDI can run as fast as PWKL. More importantly, they do not suffer from the problem of scale/sign change as MI and, thus, can be used with constraint management methods together in a straightforward manner.

[1]Jiangxi Normal University, China
[2]University of Illinois at Urbana-Champaign, USA

**Corresponding Author:**
Chanjin Zheng, School of Psychology, Jiangxi Normal University, 99 Ziyang Ave., Nanchang, Jiangxi Province, China.
Email: russelzheng@gmail.com

Cognitive diagnosis (CD) can be used to determine the presence or absence of specific skills measured by the items in educational assessment. In the past two decades, various cognitive diagnostic models (CDMs) have been proposed to facilitate obtaining diagnostic information, which specifies whether each of the required skills has been mastered (Hartz, 2002; Junker & Sijtsma, 2001; Mislevy, Almond, Yan, & Steinberg, 2000; Rupp, Templin, & Henson, 2010; K. K. Tatsuoka, 1983).

One application of CDM that has received much attention is combining CD with computerized adaptive testing (CAT), denoted as cognitive diagnostic CAT (CD-CAT; Cheng, 2009; Huebner, 2010). CAT is an approach to individual difference assessment that is administered and scored by computers (Lord, 1968). The major advantage of CAT is that items are selected sequentially based on an examinee's performance on previous items, and thus, the test is tailored to his or her latent trait level. In this regard, CAT can potentially provide a more efficient estimate of the latent trait of interest (Weiss, 1982). It is desired that CD-CAT can be developed to reap the same benefit of measurement efficiency as item response theory (IRT)–based CAT. In particular, some applications of CD-CAT, such as in interim assessment, require to be able to accurately recover examinees' cognitive pattern with a short test length.

One of the crucial elements of CD-CAT is the item selection algorithm, and measurement efficiency is the one of the major goals in the algorithm development. Most item selection algorithms in CD-CAT are developed directly from the information indices in information theory. Few algorithms may not have directly evolved from an information index, but some of them are still connected to an information index–based algorithm, such as the rate function approach (Liu, Ying, & Zhang, 2015), which can be alternatively interpreted as the minimum Kullback–Leibler (KL) information between the log-likelihood ratio distribution and the zero-mean distribution within its exponential family. Therefore, this article will focus on information indices–based algorithms in CD-CAT.

Two general approaches can be identified among the basic algorithms according to the distributions involved in the calculation: the response distribution–based approach and the cognitive pattern posterior-based approach (or simply posterior-based approach). Research on item selection in CD-CAT originated from the SHE algorithm for the sequential classification experiment for CD-CAT by C. Tatsuoka (2002) and C. Tatsuoka and Ferguson (2003) and the KL algorithm by Xu, Chang, and Douglas (2003). The key distinction lies in the distributions involved: The KL algorithm is measured by the expected distance between the distribution of the response conditional on the estimated cognitive pattern and all possible cognitive patterns and, thus, the KL algorithm falls into the response distribution approach, whereas, the SHE algorithm, a member of the posterior-based approach, is the expected Shannon entropy of the new posterior of the cognitive patterns.

Since then, major developments in the CD-CAT item selection algorithm research have evolved around these two original methods. Cheng (2009) proposed two new methods based on the KL algorithm: posterior-weighted KL (PWKL) and hybrid KL (HKL), which can match the SHE in terms of measurement efficiency. Note that the posterior in PWKL is used as the weighting factor and not involved in the calculation of KL distance, so PWKL is still a member of the response distribution–based approach, and the same applies to HKL and other extensions of PWKL. The modified PWKL (MPWKL) is the most recent development for the KL approach (Kaplan, de la Torre, & Barrada, 2015). Kaplan et al. (2015) pointed out that the computational time for MPWKL is much longer than PWKL and the generalized deterministic inputs, noisy "and" gate (G-DINA) model discrimination index (GDI), the other method proposed in that article, although the measurement efficiency has been much improved over PWKL. The heavy computational burden in MPWKL is due to the repetitive computation of the expected KL distance between the conditional response distributions given any pair of two distinct cognitive

patterns, which is defined as the D matrix (Henson & Douglas, 2005). The D matrix can be pre-calculated even before CAT administration actually takes place. During CAT administration, it can be retrieved and reused to calculate MPWKL and, thus, substantial computation can be saved. These apply to all other algorithms of this category including KL, PWKL. This pre-calculation strategy might not offer too much added value for KL or PWKL because their computation is not demanding, but it opens up a new possibility for the complicated methods such as MPWKL.

Recently, Wang (2013) proposed mutual information (MI) algorithm that exploits the expected MI between two posteriors. Simulation studies indicate that MI is more efficient than PWKL and SHE in most conditions especially in short-length tests, but it suffers from the huge computational burden. Wang devised a revised MI, but the mathematical manipulation for the simplification is not straightforward. More importantly, the value and even the sign do not remain the same as the original MI, which makes it inconvenient to use together with other methods such as non-statistical constraint management methods because the most commonly used method to incorporate constraint is to use them as weighting factors for an information index. For example, the restrictive progressive (RP) method (Wang, Chang, & Huebner, 2011) is built upon PWKL, and some new variants of RP can be easily developed by replacing the PWKL with other information indices, such as SHE. MI, however, is not feasible due to its computational burden, and neither is the revised MI for the scale/sign issue mentioned above.

The primary goal of this study is to develop two high-efficiency algorithms in this category for short-length CD-CAT, by modifying the two item discrimination indices for paper-and-pencil test construction for CD, namely, CDM discrimination index (CDI) and attribute-level CDI (ACDI), proposed by Henson and Douglas (2005); Henson, Roussos, Douglas, and He (2008); and Rupp et al. (2010). Furthermore, the second goal is to propose a pre-calculation strategy to address the computational issue for complicated algorithms in the response distribution–based approach. Henson and Douglas (2005) discussed the concept of reliability, or discrimination, for CDMs to describe the ability of a test to distinguish among examinees' cognitive patterns of mastery. CDI is essentially weighted sum of expected KL distance between any pair of cognitive patterns and serves as a quantitative measure of how informative an item is for the classification of examinees in CDM. The two newly proposed item selection algorithms, as the counterpart of MI in response distribution–based item selection methods, are expected to achieve similar or better measurement accuracy for short-length CD-CAT, but with a more elegant method to address the computational issue.

The remaining sections of the article are laid out as follows. The next section will present the CDM used in this study. Section ''Item Selection Methods'' is a brief review of major item selection algorithms in CD-CAT. Section ''CDI and ACDI'' is a summary of CDI and ACDI. A comparison between the KL algorithm and the CDI will be made. In Section ''PWCDI and PWACDI,'' modifications for CDI and ACDI to be an efficient item selection method will be presented. In Section ''Simulation Studies,'' two simulation studies will be conducted to evaluate the efficiency of the two new item selection methods. Finally, Section ''Discussion'' will conclude with a discussion of the findings of this work and directions for future research.

## The CDM

In CD, one attempts to identify the cognitive skills involved in responding to items in an assessment. Each skill is generally referred to as an *attribute*, and typically, content experts determine the attributes required for an item. The general purpose of CD is to identify which attributes each examinee has mastered or not based on the responses.

A Q-matrix is an essential element of most of the CDMs. For an item bank consisting of $J$ items, the Q-matrix is a $J \times K$ matrix of 1s and 0s that specifies the association between items and $K$ attributes (K. K. Tatsuoka, 1983). The entry corresponding to the $k$th attributes for the $j$th item, $q_{jk}$, is equal to 1 if item $j$ requires the mastery of attribute $k$, and $q_{jk} = 0$ otherwise.

The Deterministic Input, Noisy And Gate (DINA) model (Haertel, 1989; Junker & Sijtsma, 2001) is used in this study for its simplicity and widespread use in CDM research (DeCarlo, 2010; de la Torre, 2011). The DINA model assumes that, in principle, an examinee must have mastered each attribute associated with a particular item to respond correctly to that item (''And Gate'') while recognizing that examinees might respond contrary to predictions (''Noisy''). Certain examinees will answer an item incorrectly even though they have mastered all of the required attributes, whereas other examinees will answer an item correctly even though they have not mastered at least one of the required attributes. Given these properties, the DINA model–predicted probability that examinee $i$ will respond correctly to item $j$ is defined by

$$P\left(Y_{ij} = 1 | \boldsymbol{\alpha}_i\right) = \left(1 - s_j\right)^{\eta_{ij}} g_j^{1 - \eta_{ij}},$$

where $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK})$—$\alpha_{ik}$ is 1 if examinee $i$ has mastered attribute $k$, and 0 otherwise—denotes the examinee's mastery profile; $s_j$ is the probability that an examinee with all of the required attributes will ''slip'' and answer item $j$ incorrectly; and $g_j$ is the probability that an examinee with at least one missing attribute will successfully ''guess'' the correct answer. And $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$ is the latent response to item $i$ by examinee $j$. $\eta_{ij} = 1$ if examinee $i$ has mastered all the attributes measured by item $j$, otherwise $\eta_{ij} = 0$.

## Item Selection Methods

This section is a brief review of item selection algorithms concerning measurement accuracy in CD-CAT. A taxonomy of two approaches is adopted: the response distribution–based and posterior-based methods including the original KL and SHE algorithms and their important developments. As regards the distributions involved, the response distribution–based approach attempts to develop a global summative measure for the difference between the conditional distributions of the response to the candidate item given all of the possible true and estimated cognitive patterns. By contrast, the posterior-based approach involves the conditional posterior distribution(s) of cognitive patterns given all of the previous responses and the possible response to the candidate item.

This major difference carries some important implications for measurement efficiency and calculation burden. In light of measurement efficiency, the SHE algorithm is much higher than the KL algorithms as it is a direct measure of possible change in posteriors, whereas the KL algorithm assesses this change through that of the response distributions. Therefore, it is the major task for response distribution–based approach to tap the measurement potential in the response distributions as MPWKL did. As far as computation is concerned, the posterior-based approach needs updating the posterior and, thus, not much work can be done before CAT administration while the pre-calculation of the D matrix offers a new route for the response distribution–based approach. For example, alternatively, the KL indices calculation reduces to pick out and sum up an appropriate column from the D matrix. The comparison of measurement efficiency and computational reduction between these two approaches is the recurring theme of the literature review below.

## The Response Distribution–Based Approach

The KL algorithm is the sum of KL distance between the distribution of the response conditional on the estimated cognitive pattern, $P(Y_{ij}|\widehat{\boldsymbol{\alpha}}_i)$, and the distributions that are conditional on all possible cognitive patterns, $P(Y_{ij}|\boldsymbol{\alpha}_c)$, $c = 1, 2, \ldots, 2^K$ (Xu et al., 2003). This index is formulated as:

$$\text{KL}_j\left(\widehat{\boldsymbol{\alpha}}_i\right) = \sum_{c=1}^{2^K} \left[ \sum_{y=0}^{1} \log\left(\frac{P\left(Y_{ij} = y|\widehat{\boldsymbol{\alpha}}_i\right)}{P\left(Y_{ij} = y|\boldsymbol{\alpha}_c\right)}\right) P\left(Y_{ij} = y|\widehat{\boldsymbol{\alpha}}_i\right) \right].$$

It can be considered as a measure of global discrimination power of item $j$ between $P(Y_{ij}|\widehat{\boldsymbol{\alpha}}_i)$ and all the possible response conditional distributions $P(Y_{ij}|\boldsymbol{\alpha}_c)$. The item with the maximum value for KL, given the cognitive pattern $\widehat{\boldsymbol{\alpha}}_i$ for examinee $i$, is the most discriminating one and thus will be administered.

The original KL method suffers from a low measurement efficiency problem. The simulation studies indicated that the KL index cannot achieve a pattern recovery rate similar to that of the SHE algorithm for the DINA (Cheng, 2009) and the fusion model with a fixed-length CD-CAT (Xu et al., 2003). But the KL algorithm demonstrated great computational advantage for CAT administration. From the equation, it can be seen that all the possible KL indices for each item in the entire bank can be calculated without any information from the CAT administration because only the random variable $Y_{ij}$ with value of either 0 or 1 and $2^K$ possible cognitive patterns are involved. As a result, substantial computation can be done offline, and item selection reduces to picking out the maximum from the pre-calculated and stored D matrix.

The low efficiency issue was remedied by a Bayesian KL algorithm, namely, the PWKL algorithm (Cheng, 2009). To reflect the varying importance of different patterns, the addend in the KL algorithm is weighted by the corresponding posterior probability, and this modification leads to the PWKL (Cheng, 2009):

$$\text{PWKL}_j\left(\widehat{\boldsymbol{\alpha}}_i\right) = \sum_{c=1}^{2^K} \left\{ \sum_{y=0}^{1} \left[ \log\left(\frac{P\left(Y_{ij} = y|\widehat{\boldsymbol{\alpha}}_i\right)}{P\left(Y_{ij} = y|\boldsymbol{\alpha}_c\right)}\right) P\left(Y_{ij} = y|\widehat{\boldsymbol{\alpha}}_i\right) \right] \pi(\boldsymbol{\alpha}_c|\mathbf{y}_{t-1}) \right\},$$

where $\pi(\boldsymbol{\alpha}_c|\mathbf{y}_{t-1}) \propto p(\boldsymbol{\alpha}_c) \prod_{j=1}^{t-1} P(Y_{ij} = 1|\boldsymbol{\alpha}_c)^{y_{ij}} \left[1 - P(Y_{ij} = 1|\boldsymbol{\alpha}_c)\right]^{1-y_{ij}}$, $p(\boldsymbol{\alpha}_c)$ is the prior probability of the cognitive patterns, and $\mathbf{y}_{t-1}$ is the vector of responses on $t-1$ items for examinee $i$. Inspired by Henson and Douglas's (2005) discussion on the relationship between the item discrimination power and the cognitive pattern distance, Cheng (2009) further assigned additional weights to the cognitive patterns that are closer to the current cognitive pattern estimate and defined the HKL whose formulation is presented below:

$$\text{HKL} = \sum_{c=1}^{2^K} \left\{ \sum_{y=0}^{1} \left[ \log\left(\frac{P\left(Y_{ij} = y|\widehat{\boldsymbol{\alpha}}_i\right)}{P\left(Y_{ij} = y|\boldsymbol{\alpha}_c\right)}\right) P\left(Y_{ij} = y|\widehat{\boldsymbol{\alpha}}\right) \right] \pi(\boldsymbol{\alpha}_c|y_{t-1}) \frac{1}{d\left(\boldsymbol{\alpha}_c, \widehat{\boldsymbol{\alpha}}_i\right)} \right\},$$

where $d(\boldsymbol{\alpha}_c, \widehat{\boldsymbol{\alpha}}_i)$ is the Euclidean distance $\sqrt{\sum_{k=1}^{K} (\boldsymbol{\alpha}_{ck} - \widehat{\boldsymbol{\alpha}}_{ik})^2}$ for the possible cognitive pattern $\boldsymbol{\alpha}_c$ and the current cognitive pattern estimate $\widehat{\boldsymbol{\alpha}}_i$.

The MPWKL is the most recent development for the KL approach (Kaplan et al., 2015). MPWKL exploits the fact that the point estimate of cognitive pattern $\widehat{\boldsymbol{\alpha}}_i$ might not be accurate particularly when the test is relatively short, and it might be more desirable to consider the weighted sum of all the PWKL with respect to all the possible cognitive patterns $\boldsymbol{\alpha}_d, d = 1, 2, \ldots, 2^K$. More specifically, the MPWKL is calculated as follows:

$$\text{MPWKL}_j = \sum_{d=1}^{2^K} \left\{ \sum_{c=1}^{2^K} \sum_{y=0}^{1} \left[ \log \left( \frac{P(Y_j = y | \boldsymbol{\alpha}_d)}{P(Y_j = y | \boldsymbol{\alpha}_c)} \right) P(Y_j = y | \boldsymbol{\alpha}_d) \pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}) \right] \pi(\boldsymbol{\alpha}_d | \mathbf{y}_{t-1}) \right\}.$$

If a probability of 1 is assigned to the point estimate $\widehat{\boldsymbol{\alpha}}_i$, then MPWKL reduces to PWKL. MPWKL is more informative than the PWKL, and thus, it is expected to outperform PWKL, which is confirmed by simulation studies. Kaplan et al. (2015) did not assess the performance of MPWKL against MI or in a short-test scenario, or address the computational issue of MPWKL.

PWKL is an important development for the response distribution–based approach. Due to the additional information on the posterior of the cognitive patterns, the PWKL and HKL are more efficient than the original KL algorithm and similar to the SHE algorithm. Furthermore, they also enjoy the advantage of having KL distance part and distance weights (in HKL) calculated beforehand, and only the posterior weights need to be updated. So the running time for PWKL and HKL can be decomposed into two parts: the offline calculation of the KL distance and the online updating of the posterior weighting. Running time for PWKL and HKL is not a concern, but this perspective carries important implications for further development along the response distribution–based approach such as MPWKL and the two new methods proposed in the current study. The calculation of KL distance can be done offline before the administration of the CD-CAT and repeatedly retrieved when necessary, so the computational demand during the test administration is in fact not high at all. But in terms of measurement efficiency, PWKL is not designed for a short test and, therefore, not efficient enough for this scenario (Wang, 2013). The current study is motivated by this fact and attempts to fill this gap in the literature.

## The Posterior-Based Approach

C. Tatsuoka (2002) and C. Tatsuoka and Ferguson (2003) proposed the SHE item selection algorithm. Shannon entropy quantifies the uncertainty inherent in a distribution. Shannon entropy is maximized if the distribution is uniform and is minimized if the probability mass concentrates on a single point. In CD-CAT, an ideal item would be one that minimizes the expected Shannon entropy of the posterior distribution of $\widehat{\boldsymbol{\alpha}}_i$ conditional on previous responses. Thus, the SHE algorithm is defined as follows:

$$\begin{aligned}
\text{SHE} &= \sum_{y=0}^{1} \left[ \sum_{c=1}^{2^K} \pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y) \log \left( \frac{1}{\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y)} \right) \right] \left[ \sum_{c=1}^{2^K} P(Y_t = y | \boldsymbol{\alpha}_c) \pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}) \right] \\
&= \sum_{y=0}^{1} P(Y_t = y | \mathbf{y}_{t-1}) \left[ \sum_{c=1}^{2^K} \pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y) \log \left( \frac{1}{\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y)} \right) \right],
\end{aligned}$$

where $\mathbf{y}_{t-1}$ denotes the response vector of $t - 1$ items for examinee $i$, and $P(Y_t = y | \mathbf{y}_{t-1})$ is the conditional response distribution for the current item given $\mathbf{y}_{t-1}$.

It is easy to observe from the equation that the value of the posterior of the cognitive patterns needs online updating during CAT administration. Thus, no calculation can be made in advance as the KL distance part in the response distribution–based approach. Fortunately, this does not appear to be an issue in the SHE. As mentioned above, in terms of measurement efficiency, it is superior to the KL index and is comparable with the PWKL and HKL.

A recent development for the SHE approach is the MI for CD-CAT (Wang, 2013). The expected MI is calculated as follows:

$$\mathrm{MI} = \sum_{y=0}^{1} P(Y_t = y | \mathbf{y}_{t-1}) \left[ \sum_{c=1}^{2^K} \pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y) \log \left( \frac{\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y)}{\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1})} \right) \right].$$

An ideal item would be one that maximized the expected MI between $\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y)$ and $\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1})$. To analytically show the connection between SHE and MI, MI can be rewritten as the difference between two entropy measures,

$$\mathrm{MI} = \sum_{y=0}^{1} P(Y_t = y | \mathbf{y}_{t-1})$$

$$\left[ \sum_{c=1}^{2^K} \pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y) \log \left( \frac{1}{\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1})} \right) - \sum_{c=1}^{2^K} \pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y) \log \left( \frac{1}{\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y)} \right) \right]$$

$$= E_{Y_t} \big( \mathrm{SHE}(\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1})) - \mathrm{SHE}(\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y)) \big),$$

in which the second term is exactly the SHE algorithm and the first term is the expected SHE of the current posterior $\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1})$ with respect to $Y_t$. SHE, however, intends to minimize the second terms in MI, or, equivalently, maximize $-\mathrm{SHE}(\pi(\boldsymbol{\alpha}_c | \mathbf{y}_{t-1}, Y_t = y))$.

Simulation studies indicated that MI is more efficient than other competing item selection methods, particularly for short tests (Wang, 2013). However, the computational efficiency issue of the posterior-based approach poses a serious practical challenge in the case of MI because the online updating of the posteriors and a triple summation are involved in MI. By some algebraic manipulations, Wang (2013) presented a simplified version of MI:

$$\mathrm{MI} = \frac{1}{h_1} \sum_{y=0}^{1} \left[ \sum_{c=1}^{2^K} P(\mathbf{y}_{t-1}, Y_t = y | \boldsymbol{\alpha}_c) P(\boldsymbol{\alpha}_c) \log \frac{P(Y_t = y | \boldsymbol{\alpha}_c)}{h_2} \right],$$

where $h_1 = \sum_{c=1}^{2^K} P(\mathbf{y}_{t-1} | \boldsymbol{\alpha}_c) P(\boldsymbol{\alpha}_c)$ and $h_2 = \sum_{c=1}^{2^K} P(\mathbf{y}_{t-1}, Y_t = y | \boldsymbol{\alpha}_c) P(\boldsymbol{\alpha}_c)$. The calculation burden can be reduced by dropping some terms only related to $h_1$ because it is a constant term over different items. Wang pointed out that one problem with such a simplification scheme is that it only preserves the rank of the original index and that there is a change in scale and/or sign. Therefore, if weighted by an item exposure control and constraint management index via multiplication, a simplified MI would produce an incorrect ordering of items.

In summary, both approaches have certain strengths and weaknesses. Compared with the indirect response distribution–based approach, the posterior-based approach is a more direct and effective measure in the context of CD-CAT in which an accurate estimate of the cognitive pattern obtained via the updated posterior distribution is the ultimate goal. The response distribution–based approach, however, has an edge over the posterior-based approach with respect to computational efficiency. MI is the most efficient method among all the existing methods but only with a partial solution to the computational issue. No empirical study comparing MI and MPWKL has been found in the existing literature. It is desirable to develop algorithms of both high measurement efficiency and low computational demand. With this goal in mind, the current study attempts to develop two response distribution–based high efficient methods from a CDI which can be considered as a generalized KL algorithm.

**Table 1.** A Simulated D Matrix for a Two-Attribute Test.

|       | (0,0) | (1,0) | (0,1) | (1,1) |
|-------|-------|-------|-------|-------|
| (0,0) | 0.00  | 0.90  | 1.15  | 2.05  |
| (1,0) | 0.80  | 0.00  | 1.94  | 1.15  |
| (0,1) | 1.36  | 2.26  | 0.00  | 0.90  |
| (1,1) | 2.16  | 1.36  | 0.80  | 0.00  |

*Note.* (0,0), (1,0), (0,1), and (1,1) are the four possible cognitive patterns for a two-attribute test; the Kullback–Leibler (KL) values in the table are simulated only for illustration purpose.

## CDI and ACDI

### CDI

CDI is proposed for facilitating paper-and-pencil test construction for CD purposes. The building block for CDI is the D matrix for the $j$th item, $D_j$, and it is a $2^K \times 2^K$ matrix whose entries are the expected KL distance between the response distributions for any two distinct cognitive patterns. Each $u, v$ element in $D_j$, namely, the expected KL distance between the response distributions conditional on the cognitive patterns $\alpha_u$ and $\alpha_v$ (where, $u = 1, 2, \ldots, 2^K$ and $v = 1, 2, \ldots, 2^K$), is

$$D_{juv} = E_{\alpha_u}\left[\log\left(\frac{P_{\alpha_u}(X_j)}{P_{\alpha_v}(X_j)}\right)\right].$$

A simulated example for the D matrix of a two-attribute test is presented in Table 1 for illustration purpose; (0,0), (1,0), (0,1), and (1,1) are the four possible cognitive patterns for a two-attribute test, and all the KL values in the table are simulated only for illustration purpose.

A possible summative measure for the discriminatory power of the $j$th item is the mean of all the elements of $D_j$. But the difficulty of differentiating pairs of cognitive patterns is different. Specifically, an examinee who has not mastered any of the attributes measured by a test is easily discriminated from an examinee who has mastered all attribute patterns. On the contrary, cognitive patterns that differ by only one component are usually the most difficult to discriminate; therefore, $D_{juv}$s for those comparisons require more attention. If a test discriminates well between similar cognitive patterns, it will discriminate well between those dissimilar cognitive patterns. Therefore, a weighted average should be used such that each element is first weighted by the similarity, or inverse ''distance'' between the cognitive patterns. Thus, a larger emphasis is placed on those comparisons of cognitive patterns that are more similar.

Thus, a KL distance–based quantitative measure of how informative an item is for the classification of examinees in CDM, namely, CDI, can be constructed as a weighted mean of all off-diagonal elements of $D_j$.

$$\text{CDI}_j = \frac{1}{\sum_{u \neq v} h(\alpha_u, \alpha_v)^{-1}} \sum_{u \neq v} h(\alpha_u, \alpha_v)^{-1} D_{juv},$$

where $h(\alpha_u, \alpha_v) = \sum_{k=1}^{K} |\alpha_{uk} - \alpha_{vk}|$ is the Hamming distance between two cognitive patterns $\alpha_u$ and $\alpha_v$. Henson et al. (2008) and Rupp et al. (2010) further developed the ACDI for attribute $k$. This is defined as follows:

$$\mathrm{ACDI}_{jk} = \frac{1}{2^K} \sum_{\text{all relevant cells}} D_{juv}.$$

All of the relevant cells are defined as the entries in the D matrix where only the $k$th attribute is different for cognitive patterns $\alpha_u$ and $\alpha_v$. The ACDI for item $j$, $\mathrm{ACDI}_j$, is simply the sum of $\mathrm{ACDI}_{jk}$ over $k$ from 1 to $K$:

$$\mathrm{ACDI}_j = \sum_{k=1}^{K} \mathrm{ACDI}_{jk} = \sum_{k=1}^{K} \frac{1}{2^K} \sum_{\text{all relevant cells}} D_{juv}.$$

To simply the notation,

$$\mathrm{ACDI}_j = \frac{1}{2^K} \sum_{\text{all relevant cells}} D_{juv},$$

where ''all of the relevant cells'' refers to all of the entries for any pair of cognitive patterns with the Hamming distance of 1. $\mathrm{ACDI}_j$ for item $j$ is a partial sum of the matrix D in which only the entries for two cognitive patterns with the Hamming distance of 1 are included. The cognitive patterns with a Hamming distance of 1 are the pair that differ by only one component and usually the most difficult to discriminate, so ACDI is the summation of all the most important elements in the matrix D in terms of discriminating examinees.

### Why CDI Is Better Than the KL Algorithm?

It is easy to note that the D matrix connects the computation of CDI/ACDI and the KL algorithm. If the KL is obtained via the pre-calculation strategy described above, one, in fact, has to construct the D matrix first, although this concept was not proposed in the original article. All the possible KL values for an item can be obtained by summing up each columns corresponding to the interim cognitive pattern estimate, namely, the marginal sum of columns in Table 1. Without the weights of cognitive pattern similarity, CDI coincides with the sum of those possible KL values. In this sense, CDI can be interpreted as the sum of all the possible KL values weighted by cognitive pattern similarity. Viewed in this new perspective, CDI as an item selection algorithm is expected to be more informative than the KL algorithm, particularly during the early stage of CD-CAT where the cognitive pattern estimate might not be accurate because, unlike the KL algorithm, the calculation does not rely on the interim estimate of the cognitive pattern.

ACDI is a partial sum of the D matrix. ACDI selects the most important elements in $D_j$. It would be interesting to assess its performance against the KL algorithm that chooses one particular column in $D_j$.

To further tap the potential, the same line of thinking of using posterior as weighting in developing PWKL from KL can be applied to the CDI and ACDI, and the resultant new methods, posterior-weighted CDI (PWCDI) and posterior-weighted ACDI (PWACDI), are expected to outperform or, at least, match PWKL.

## PWCDI and PWACDI

### The PWCDI and PWACDI

The key change is to incorporate the posterior distribution of cognitive patterns into the static D matrix. It is natural to follow the same reasoning as does PWKL and take the varying

importance of different cognitive patterns into account. One complication, however, is that the entries in $D_j$ are the expected KL distance between the conditional response distributions for any pair of two distinct cognitive patterns and the posterior should be considered for both the rows and columns, whereas in the KL algorithm, the weights are only applied to the columns. Then the posterior-weighted **D** (PWD) matrix for item $j$ can be defined as follows:

$$\text{PWD}_{juv} = E_{\boldsymbol{\alpha}_u}\left[\boldsymbol{\pi}(\boldsymbol{\alpha}_u)\times\boldsymbol{\pi}(\boldsymbol{\alpha}_v)\times\log\left(\frac{P(X_j|\boldsymbol{\alpha}_u)}{P(X_j|\boldsymbol{\alpha}_v)}\right)\right],$$

where $\alpha_u$ and $\alpha_v$ are the updated cognitive pattern posteriors (where, $u = 1, 2, \ldots, 2^K$ and $v = 1, 2, \ldots, 2^K$). The PWCDI and PWACDI can then be easily defined in the same manner as the original CDI and ACDI:

$$\text{PWCDI}_j = \frac{1}{\sum_{u\neq v}h(\boldsymbol{\alpha}_u, \boldsymbol{\alpha}_v)^{-1}}\sum_{u\neq v}h(\boldsymbol{\alpha}_u, \boldsymbol{\alpha}_v)^{-1}\text{PWD}_{juv} \text{ and}$$

$$\text{PWACDI}_j = \frac{1}{2^K}\sum_{\text{all relevant cells}}\text{PWD}_{juv}, \text{ respectively.}$$

It is easy to see that both posterior and distance weights are used in PWCDI and PWACDI.

## Connection to Existing Algorithms

PWCDI is closely related to some existing algorithms that can be clearly revealed through the application of the D matrix in various algorithms. Just as explained above, the calculation of the KL algorithm amounts to picking out the proper column of the D matrix corresponding to the estimated cognitive pattern and then obtaining the column sum. With pre-calculation of the D matrix, one may store the D matrix first and, during the administration, finish the computation according to the cognitive pattern estimate. Similarly, HKL is the column sum weighted by cognitive pattern distance, PWKL the posterior-weighted column sum, and MPWKL the sum of the D matrix weighted by row and column posteriors, and cognitive pattern distance. PWCDI, however, is essentially the sum of the D matrix weighted by row and column posteriors, and cognitive pattern distance. In this sense, PWCDI is a counterpart to HKL as the MPWKL to PWKL and can be alternatively labeled as the modified HKL (MHKL). PWCDI can be interpreted as a posterior-weighted sum of all HKL values. The only difference between PWCDI and MPWKL is the distance weighting, which is the same as that between PWKL and HKL. It can be expected that PWACDI, on the contrary, is not directly related to any existing algorithms.

## Computational Simplification

The dynamic nature of the PWD matrix poses some computational challenge, particularly for CAT administration where real-time delivery is the key. Just like MI, PWCDI and PWACDI require a triple summation over $2^K$ possible cognitive patterns. This problem can be solved easily using the same reasoning for the construction of the PWD matrix. The PWD matrix may be partitioned into the ''dynamic'' posterior weighting and the ''static'' D matrix. The ''dynamic'' posterior weighting requires updating using the cognitive pattern estimate in each iteration of CAT administration whereas the ''static'' D matrix remains constant over different iterations of CAT and examinees. The computational demands for these two parts are drastically different.

Only one multiplication is needed for the calculation of weighting, whereas that for the static part is much more complicated. Translating this into mathematical language, the PWD can be reformulated as follows:

$$\mathrm{PWD}_{juv} = \pi(\boldsymbol{\alpha}_u) \times \pi(\boldsymbol{\alpha}_v) \times \sum_{x=0}^{1} P(X_j|\boldsymbol{\alpha}_u) \times \log\left[\frac{P_{\boldsymbol{\alpha}_u}(X_j|\boldsymbol{\alpha}_u)}{P_{\boldsymbol{\alpha}_v}(X_j|\boldsymbol{\alpha}_v)}\right],$$

and the matrix form is

$$\mathrm{PWD} = \boldsymbol{\pi} \times \boldsymbol{\pi}^T \bullet \mathrm{D},$$

where $\boldsymbol{\pi}$ is the vector for the posterior probability of cognitive patterns and $\pi^T$ is its transpose. The symbol "$\bullet$" indicates the element-wise matrix multiplication. In practice, the D matrix can be calculated beforehand and stored for the repetitive use in CD-CAT administration. In a matrix-oriented programming language such as MATLAB, this simplification can improve calculation speed significantly. Compared with the computational simplification made for MI, algebraic manipulation is much easier in PWD and the issues of negativity and scale change are also conveniently avoided. Therefore, for example, the RP method based on PWCDI and PWACDI can be easily constructed by replacing PWKL with them. In summary, PWCDI and PWACDI are a superior computational alternative to MI.

## Simulation Studies

### Study 1: The Fixed-Length Test

*Design.*  A fixed-length CD-CAT simulation study was carried out to evaluate the efficiency of the new algorithms. Three factors were manipulated in the simulation study: test length (short vs. long), item bank quality (high vs. low), and item selection algorithms. The details were as follows:

*Examinees generation.*  Three thousand examinees were generated assuming that every examinee has a 50% chance of mastering each attribute. In a five-attribute test, there were 32 distinct types of cognitive patterns that were assumed to be equally likely in the population.

*Item bank generation.*  The item bank consisting of 500 items for a five-attribute DINA model is generated in the same manner as in Cheng (2009). The Q-matrix used in this study is generated item by item and attribute by attribute. Each item has a 30% chance of measuring each attribute. This mechanism was employed to ensure that every attribute is adequately and equally represented in the item pool. The item parameters $s_j$ and $g_j$ were both generated from $U(0.05, 0.25)$ for the high-quality item bank and from $U(0.10, 0.30)$ for the low-quality bank.

*Test length.*  The length of the short test was set as 5 items, and the length of the long test was set to be 10 items.

*Item selection algorithms.* Eight selection algorithms were compared in this study: KL, PWKL, ACDI, CDI, MI, MPWKL, PWACDI, and PWCDI. The comparisons of KL, ACDI, and CDI can reveal the efficiency of ACDI and CDI if they were used as item selection algorithms against KL. The original ACDI and CDI were also compared with PWACDI and PWCDI to demonstrate the effectiveness of the static-to-dynamic change of D matrix. The performance of PWACDI and PWCDI against PWKL and MI is of the greatest interest for the current study. The performance of MPWKL against MI is also interesting as no empirical studies have been done.

*Evaluation criteria.* The efficiency of the algorithms can be demonstrated using the high attribute correct classification rate (ACCR) and mastery pattern correct classification rate (PCCR). ACCR is defined as

$$\text{ACCR}_k = \sum_{i=1}^{3,000} I\left(\alpha_{ik} = \widehat{\boldsymbol{\alpha}}_{ik}\right) / 3,000,$$

where $I$ is the indicator function; PCCR is defined as

$$\text{PCCR} = \sum_{i=1}^{3,000} I\left(\boldsymbol{\alpha}_i = \widehat{\boldsymbol{\alpha}}_i\right) / 3,000.$$

*Results.* The ACCR and PCCR for the eight algorithms in various item banks and of various test lengths are presented in Table 2. In the short test under the high-quality item bank, the PCCRs for ACDI and CDI, 0.187 and 0.245 respectively, are higher than those of KL, even though ACDI and CDI are not proposed as an item selection algorithm for CD-CAT. The table shows that PWACDI and PWCDI have a 0.773 PCCR, and outperform ACDI and CDI, which indicates that the modification proposed in this study is quite effective. More interesting results concern the PCCRs for PWKL, MI, PWACDI, and PWCDI. The performances of PWACDI and PWCDI are indistinguishable from those of MI. PWACDI in particular achieves the same measurement precision as PWCDI. The lost information on other entries in the matrix D does not exert a negative effect on item selection. PWCDI and MPWKL work equally well. As expected, there is a substantial difference of 0.154 between the PCCRs for these three algorithms and PWKL. Similar observations can be made easily for the low-quality item bank.

In the long test, regardless of item bank quality, the PCCRs for ACDI and CDI are higher than is the case for KL. The difference between the ACDI/CDI and PWKL is still noticeable. The difference between MI and PWKL almost disappears, and the difference between the PWACDI/PWCDI and the PWKL shrinks to about 0.01 to 0.03. PWACDI, PWCDI, and MPWKL are almost identical in the long test.

To demonstrate the advantage of the pre-calculation of the matrix D, the running time for several algorithms in two strategies for one simulation study condition, namely, without pre-calculation (old) and with pre-calculation (new), is recorded and summarized in Table 3. The running time for SHE and simplified MI is also given as the baseline. The pre-calculation strategy does not make a big difference for PWKL, but for MPWKL and PWCDI, the running time reduces by a factor of about 6.4 and similar to that of PWKL. The running time of MPWKL was reported as 3 times more than that of PWKL (Kaplan et al., 2015). SHE is a little longer than PWKL, and MI is twice as long as SHE. The proposed method is very effective in reducing the calculation time for the response distribution–based algorithm.

## Study 2: The Variable-Length Test

*Design.* Study 2 seeks to investigate the efficiency of the two proposed algorithms against PWKL in a variable-length test. A more efficient algorithm can terminate the test with fewer items than a less efficient algorithm in a variable-length test.

Three factors were manipulated in the simulation study: item bank quality (high vs. low), the termination rule, and three item selection algorithms (PWKL, PWACDI, and PWCDI). Examinees and item banks were simulated in the same manner as in Study 1. The termination rule for the variable-length test was proposed by C. Tatsuoka and Ferguson (2003) and stops

**Table 2.** The ACCR and PCCR for Eight Algorithms in Various Item Banks and Test Lengths.

| Test length | Item quality | Selection algorithms | ACCR | | | | | PCCR | Difference |
|---|---|---|---|---|---|---|---|---|---|
| | | | A1 | A2 | A3 | A4 | A5 | | |
| 5 | High | KL | 0.581 | 0.964 | 0.548 | 0.897 | 0.562 | 0.144 | |
| | | ACDI | 0.720 | 0.989 | 0.492 | 0.941 | 0.492 | 0.187 | |
| | | CDI | 0.503 | 0.989 | 0.939 | 0.939 | 0.491 | 0.245 | |
| | | PWKL | 0.805 | 0.942 | 0.881 | 0.923 | 0.900 | 0.619 | |
| | | MI | 0.942 | 0.945 | 0.925 | 0.948 | 0.926 | 0.774 | 0.155 |
| | | MPWKL | 0.921 | 0.943 | 0.916 | 0.923 | 0.904 | 0.772 | 0.153 |
| | | PWACDI | 0.911 | 0.943 | 0.915 | 0.923 | 0.889 | 0.773 | 0.154 |
| | | PWCDI | 0.923 | 0.943 | 0.917 | 0.923 | 0.906 | 0.773 | 0.154 |
| | Low | KL | 0.571 | 0.943 | 0.587 | 0.845 | 0.539 | 0.158 | |
| | | ACDI | 0.690 | 0.967 | 0.492 | 0.887 | 0.491 | 0.193 | |
| | | CDI | 0.504 | 0.966 | 0.888 | 0.889 | 0.491 | 0.238 | |
| | | PWKL | 0.762 | 0.875 | 0.835 | 0.867 | 0.835 | 0.512 | |
| | | MI | 0.887 | 0.888 | 0.846 | 0.894 | 0.866 | 0.627 | 0.115 |
| | | MPWKL | 0.852 | 0.891 | 0.859 | 0.857 | 0.824 | 0.621 | 0.109 |
| | | PWACDI | 0.838 | 0.892 | 0.856 | 0.856 | 0.811 | 0.635 | 0.123 |
| | | PWCDI | 0.852 | 0.892 | 0.857 | 0.856 | 0.826 | 0.611 | 0.099 |
| 10 | High | KL | 0.591 | 0.989 | 0.753 | 0.928 | 0.842 | 0.337 | |
| | | ACDI | 0.719 | 0.991 | 0.937 | 0.947 | 0.714 | 0.534 | |
| | | CDI | 0.939 | 0.992 | 0.939 | 0.933 | 0.924 | 0.771 | |
| | | PWKL | 0.997 | 0.981 | 0.977 | 0.980 | 0.971 | 0.909 | |
| | | MI | 0.981 | 0.973 | 0.967 | 0.978 | 0.976 | 0.900 | −0.09 |
| | | MPWKL | 0.983 | 0.984 | 0.984 | 0.980 | 0.974 | 0.928 | 0.019 |
| | | PWACDI | 0.976 | 0.980 | 0.978 | 0.971 | 0.972 | 0.921 | 0.012 |
| | | PWCDI | 0.983 | 0.983 | 0.984 | 0.979 | 0.973 | 0.926 | 0.017 |
| | Low | KL | 0.578 | 0.981 | 0.825 | 0.9 | 0.655 | 0.277 | |
| | | ACDI | 0.688 | 0.972 | 0.883 | 0.891 | 0.685 | 0.460 | |
| | | CDI | 0.877 | 0.966 | 0.885 | 0.878 | 0.869 | 0.618 | |
| | | PWKL | 0.935 | 0.946 | 0.927 | 0.939 | 0.922 | 0.768 | |
| | | MI | 0.933 | 0.941 | 0.936 | 0.944 | 0.949 | 0.781 | 0.013 |
| | | MPWKL | 0.937 | 0.949 | 0.944 | 0.935 | 0.930 | 0.798 | 0.030 |
| | | PWACDI | 0.925 | 0.939 | 0.933 | 0.920 | 0.928 | 0.800 | 0.032 |
| | | PWCDI | 0.938 | 0.947 | 0.944 | 0.936 | 0.930 | 0.800 | 0.032 |

*Note.* ACCR = attribute correct classification rate; A = attribute; PCCR = pattern correct classification rate; KL = Kullback–Leibler index method; ACDI = attribute-level cognitive diagnostic model discrimination index; CDI = cognitive diagnostic model discrimination index; PWKL = posterior-weighted Kullback–Leibler information method; MI = mutual information method; MPWKL = modified PWKL; PWACDI = posterior-weighted attribute-level cognitive diagnostic model discrimination index; PWCDI = posterior-weighted cognitive diagnostic model discrimination index.

**Table 3.** Average Test Administration Time per Examinees (10 Items, High-Quality Bank).

| Algorithm | Time (old, in ms) | Time (new, in ms) |
|---|---|---|
| PWKL | 20 | 19 |
| MPWKL | 142 | 22 |
| PWCDI | 144 | 23 |
| MI | 55 | |
| SHE | 28 | |

*Note.* "Old" refers to the calculation strategy in which the matrix D is not pre-calculated and "new" is otherwise. PWKL = posterior-weighted Kullback–Leibler information method; MPWKL = modified posterior-weighted Kullback–Leibler information method; PWCDI = posterior-weighted cognitive diagnostic model discrimination index; MI = mutual information method simplified as in Wang (2013).

the test when the probability of the cognitive pattern with the largest probability reaches a pre-specified value, such as 0.7, 0.8, and 0.9, in the current study.

*Evaluation criteria.* The efficiency of an algorithm in a variable-length test can be measured by the mean test length. Other descriptive statistics of the test length including the maximum, minimum, and standard deviation were also reported.

*Results.* All of the descriptive statistics for three algorithms under various combinations of item banks and different criteria for the stopping rule are summarized in Table 4. Regardless of the item quality and stopping rule criterion, the mean test length for MI, PWACDI, and PWCDI is smaller than that of the PWKL, except that in the low item quality bank, MI produces a larger mean test length when the stopping rule criterion is conservative (i.e., 0.8, 0.9). Item bank quality and stopping rule criteria have some effect on MI and PWACDI, but under all of the conditions, PWCDI uniformly has about 0.5 items fewer than is the case for the PWKL.

## Discussion

The PWKL is a well-established efficient Bayesian item selection algorithm in CD-CAT. It can achieve satisfactory measurement accuracy with a relative long test. Some applications of CD, such as in interim assessment, however, aim to reaching the goal of accuracy in a short test. MI, as a computationally intensive SHE approach method, provided a partial solution to this need.

The current study pursues the response distribution–based approach, developing two Bayesian methods based on CDI/ACDI, PWACDI, and PWCDI. The key to the improvement is the information on all of the other possible cognitive patterns besides the estimated cognitive pattern. This is particularly important during the early stage of CD-CAT. Inaccuracy of estimating the latent trait at the early stage of CAT is well recognized (Chang & Ying, 1996). Thus, some item selection methods are not efficient during the early stage because the cognitive pattern estimate plays an important role in the calculation. The PWKL remedied this issue by incorporating the posterior distribution of the cognitive patterns, which is the usual Bayesian solution. The proposed methods provide a further improvement by taking advantage of all of the pairwise comparison of all possible cognitive patterns in the CDI, together with the Bayesian solution. Two simulation studies demonstrate that the new algorithms can improve the PCCR greatly in a short test and can satisfy the pre-specified stopping rule with fewer items in a variable-length test.

It is worth noting that there might be an issue of Q-matrix completeness for the short test length conditions in Study 1. Chiu, Douglas, and Li (2009) stated that the necessary and sufficient condition for a complete Q-matrix was that it contained all the unit vectors. More specifically, the Q-matrix for an examinee in the short test length conditions is complete if it is a 5-by-5 identity matrix after some necessary column swapping. According to this rule, the completeness of the Q-matrices produced by all of the algorithms may be empirically checked. The Q-matrices produced by MI, MPWKL, PWCDI, and PWCDI are complete whereas those produced by PWKL might not be, which can be additional evidence of the superiority of the new algorithms.

Among all the questions that deserve further studies, the most interesting one is to investigate the efficiency of the two new methods if they are combined with the item exposure control mechanism. In practice, some statistical and non-statistical constraints are important such as the item exposure rates. Wang et al. (2011) proposed two restrictive stochastic item selection methods for addressing the issue of the trade-off between measurement precision and item security based on the PWKL, namely, RP and the restrictive threshold (RT) method. The PWACDI and PWCDI can be easily generalized into the RT and RP methods, and replace the PWKL index in

**Table 4.** The Descriptive Statistics of the Test Length for the Variable-Length Test.

| Item quality | Stopping rule | PWKL | | | | MI | | | | PWACDI | | | | PWCDI | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Maximum | Minimum | M | SD | Maximum | Minimum | M | SD | Maximum | Minimum | M | SD | Maximum | Minimum | M | SD |
| High | 0.7 | 14 | 4 | 5.75 | 1.13 | 13 | 5 | 5.52 | 0.76 | 17 | 5 | 5.56 | 1.24 | 12 | 5 | 5.37 | 0.86 |
| | 0.8 | 20 | 4 | 7.19 | 1.87 | 16 | 5 | 6.74 | 1.57 | 22 | 5 | 6.77 | 1.84 | 20 | 5 | 6.74 | 1.51 |
| | 0.9 | 23 | 5 | 9.21 | 2.36 | 23 | 6 | 8.90 | 2.29 | 24 | 6 | 8.65 | 2.24 | 22 | 6 | 8.72 | 2.12 |
| Low | 0.7 | 24 | 4 | 8.75 | 2.81 | 27 | 6 | 8.54 | 2.70 | 32 | 5 | 8.56 | 3.27 | 23 | 6 | 8.31 | 2.55 |
| | 0.8 | 31 | 5 | 10.33 | 3.27 | 32 | 6 | 10.39 | 3.37 | 39 | 6 | 10.04 | 3.50 | 30 | 6 | 9.85 | 3.04 |
| | 0.9 | 29 | 5 | 12.10 | 3.58 | 35 | 6 | 12.87 | 4.24 | 50 | 6 | 11.95 | 4.28 | 36 | 7 | 11.60 | 3.57 |

*Note.* PWKL = posterior-weighted Kullback–Leibler information method; MI = mutual information method; PWACDI = posterior-weighted attribute-level cognitive diagnostic model discrimination index; PWCDI = posterior-weighted cognitive diagnostic model discrimination index.

RP and RT. It would be interesting to investigate whether the RP and RT based on the MI and the two proposed methods can still maintain this advantage against the original RP and RT.

## Declaration of Conflicting Interests

## Funding

## References

Chang, H. H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, *20*, 213-229.

Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika*, *74*, 619-632.

Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, *74*, 633-665.

DeCarlo, L. T. (2010). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, *35*, 8-26.

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

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*, 301-321.

Hartz, S. M. C. (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, Champaign, IL.

Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement*, *29*, 262-277.

Henson, R., Roussos, L., Douglas, J., & He, X. (2008). Cognitive diagnostic attribute-level discrimination indices. *Applied Psychological Measurement*, *32*, 275-288.

Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation*, *15*(3), 1-7.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*, 258-272.

Kaplan, M., de la Torre, J., & Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, *39*, 167-188.

Liu, J., Ying, Z., & Zhang, S. (2015). A rate function approach to computerized adaptive testing for cognitive diagnosis. *Psychometrika*, *80*, 468-490.

Lord, F. M. (1968). *ETS Research Bulletin Series: Some test theory for tailored testing*. Princeton, NJ: Educational Testing Service.

Mislevy, R., Almond, R., Yan, D., & Steinberg, L. (2000). *Bayes nets in educational assessment: Where do the numbers come from?* Princeton, NJ: CRESST/Educational Testing Service.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York, NY: 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*, 337-350.

Tatsuoka, C., & Ferguson, T. (2003). Sequential classification on partially ordered sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*, 143-157.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*, 345-354.

Wang, C. (2013). Mutual information item selection method in cognitive diagnostic computerized adaptive testing with short test length. *Educational and Psychological Measurement*, *73*, 1017-1035.

Wang, C., Chang, H. H., & Huebner, A. (2011). Restrictive stochastic item selection methods in cognitive diagnostic computerized adaptive testing. *Journal of Educational Measurement*, *48*, 255-273.

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, *6*, 473-492.

Xu, X., Chang, 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.