Original Research

A Note on the Relationship of the Shannon Entropy Procedure and the Jensen–Shannon Divergence in Cognitive Diagnostic Computerized Adaptive Testing

Wenyi Wang[1] , Lihong Song[1], Teng Wang[1], Peng Gao[1], and Jian Xiong[1]

Abstract

The purpose of this study is to investigate the relationship between the Shannon entropy procedure and the Jensen–Shannon divergence (JSD) that are used as item selection criteria in cognitive diagnostic computerized adaptive testing (CD-CAT). Because the JSD itself is defined by the Shannon entropy, we apply the well-known relationship between the JSD and Shannon entropy to establish a relationship between the item selection criteria that are based on these two measures. To understand the relationship between these two item selection criteria better, an alternative way is also provided. Theoretical derivations and empirical examples have shown that the Shannon entropy procedure and the JSD in CD-CAT have a linear relation under cognitive diagnostic models. Consistent with our theoretical conclusions, simulation results have shown that two item selection criteria behaved quite similarly in terms of attribute-level and pattern recovery rates under all conditions and they selected the same set of items for each examinee from an item bank with item parameters drawn from a uniform distribution $U(0.1, 0.3)$ under post hoc simulations. We provide some suggestions for future studies and a discussion of relationship between the modified posterior-weighted Kullback–Leibler index and the G-DINA (generalized deterministic inputs, noisy "and" gate) discrimination index.

Summative assessments are typically used for grading and accountability purposes, and formative assessments are often used for supporting student learning (Laveault & Allal, 2016). Researchers and practitioners began to focus on formative assessments for student learning, rather than focus solely on summative assessments because many evidences showed that formative assessments produce significant and often substantial learning gains and improve student confidence and achievement (Black & Wiliam, 1998; Laveault & Allal, 2016). Cognitive diagnosis assessment (CDA) can be regarded as a kind of formative assessments because it is intended to promote assessment for learning to modify instruction and learning in classrooms by providing the formative diagnostic information about students' cognitive strengths and weaknesses (Jang, 2008; Leighton & Gierl, 2007). CDA has received increasing attention in recent years (Leighton & Gierl, 2007; Rupp et al., 2010; K. K. Tatsuoka, 2009), especially since the No Child Left Behind Act of 2001

mandated the selection and use of diagnostic assessments to improve teaching and learning and the new federal grant program known as "Race to the Top" (RTTT) has led into a new era of K–12 assessments which emphasized both accountability and instructional improvement (Chang, 2012).

Computerized adaptive testing (CAT) has become a popular mode of many summative and formative assessments (Quellmalz & Pellegrino, 2009). As a method of administering test items, CAT tailors the item difficulty to the ability level of the individual examinee (Chang & Ying, 2007). It is attractive to practitioners because it yields a high measurement precision with a short test. In the framework of CAT,

[1]Jiangxi Normal University, Nanchang, China

Corresponding Author:
Lihong Song, Elementary Education College, Jiangxi Normal University, 99 Ziyang Avenue, Nanchang, Jiangxi 330022, China.
Email: viviansong1981@163.com

cognitive diagnostic computerized adaptive testing (CD-CAT) is also a popular mode of online testing for cognitive diagnosis, as it can help one make informed decisions about the next steps in instruction for each student and greatly facilitate individualized learning (Chang, 2015) and provide many benefits to support formative assessments (Gierl & Lai, 2018). Particularly, the U.S. National Education Technology Plan 2017 with the title of "Reimagining the Role of Technology in Education" (U.S. Department of Education, 2017) emphasizes that technology can help us redefine assessment to meet the needs of the learner in a variety of ways. For technology-based formative assessments or CAT, test items are adapted to learner's ability and knowledge during the testing process. Thus, CAT can provide real-time reporting of results during the instructional process, which is crucial for personalized learning (Chen & Chang, 2018).

A key ingredient in CD-CAT is the item selection index. Researchers have attempted to investigate many item selection indices. The first type of index is based on the Kullback–Leibler (KL) information, such as the KL index (Cheng, 2009; McGlohen & Chang, 2008; C. Tatsuoka & Ferguson, 2003; Xu et al., 2003), the likelihood- or posterior-weighted KL (LWKL or PWKL) index and the hybrid KL index (Cheng, 2009), the restrictive progressive or threshold PWKL index (Wang et al., 2011), the aggregate ranked information index and the aggregate standardized information index (Wang et al., 2014), the modified PWKL (MPWKL) index (Kaplan et al., 2015), the KL expected discrimination index (W. Y. Wang et al., 2015), the posterior-weighted cognitive diagnostic model (CDM) discrimination index and the posterior-weighted attribute-level CDM discrimination index (Zheng & Chang, 2016), and the information product index (Zheng et al., 2018). The second is based on the Shannon entropy, called the Shannon entropy (SHE) procedure (Cheng, 2009; McGlohen & Chang, 2008; C. Tatsuoka, 2002; C. Tatsuoka & Ferguson, 2003; Xu et al., 2003, 2016). The third is based on the mutual information, including the expected mutual information index (Wang, 2013) and the Jensen–Shannon divergence (JSD) index (Kang et al., 2017; Minchen & de la Torre, 2016; Yigit et al., 2018). There are other indices, such as the generalized deterministic inputs, noisy "and" gate (G-DINA; de la Torre, 2011) model discrimination index (GDI; Kaplan et al., 2015), the rate function approach (Liu et al., 2015), the halving algorithm (C. Tatsuoka & Ferguson, 2003; W. Y. Wang et al., 2015; Zheng & Wang, 2017), and so on. Yigit et al. (2018) has proved that the mutual information index and the JSD index are equivalent. Although the previous simulation studies showed that the SHE and the JSD or mutual information perform quite similarly, the main purpose of this study is to describe the theoretical relationship between the SHE procedure and the JSD index.

CDMs

Before introducing item selection indices for CD-CAT, the general concept of CDMs and the G-DINA model as a kind of general CDMs are described here. CDMs have been defined by Rupp and Templin (2008) as "probabilistic, confirmatory multidimensional latent variable models with a simple or complex loading structure" (p. 226). The loading structure for a CDM is represented by its Q-matrix (K. K. Tatsuoka, 1983, 2009). The entries of a Q-matrix indicate 1 or 0, in which $q_{jk} = 1$ when item j involves attribute k for answering item j correctly and $q_{jk} = 0$ otherwise. CDMs define an item response function of a Q-matrix, examinee's discrete latent variables, and item parameters to predict the probability of an observable categorical response to an item. This study only focuses on CDMs designed to handle dichotomous responses. For a dichotomous CDM, the form of an item response function for a binary response variable is denoted by $P_{ju}(\alpha_i) = P(U_{ij} = u \mid \alpha_i, q_j, \beta_j)$, where $u \in \{0,1\}$, q_j is the jth row of a Q-matrix, and β_j is item parameters for item j. Note that this study will use $P_{ju}(\alpha_i)$ to discuss the theoretical relationship between the SHE procedure and the JSD index. While the G-DINA model and other CDMs are described below only for showing details about different item response functions for dichotomous CDMs.

Let $K_j^* = \sum_{k=1}^{K} q_{jk}$ denote the number of required attributes for item j, where K is the number of attributes of a test. For K attributes, there are 2^K distinct attribute patterns in the universal set of knowledge states, attribute patterns, or latent classes. Let α_i denote an attribute pattern from the universal set of knowledge states. In the G-DINA model, item j that measures K_j^* attributes partitions the 2^K distinct attribute patterns into $2^{K_j^*}$ latent groups. To simplify the notation, let α_{ij}^* be the reduced attribute pattern of the full attribute pattern α_i with respect to the required attributes for item j. Let U_{ij} be the response of examinee i to item j. We define $P_{ju}(\alpha_i) = P(U_{ij} = u \mid \alpha_{ij}^*)$ to be the probability distribution of the binary random variable U_{ij}, where $P_{j1}(\alpha_i)$ and $P_{j0}(\alpha_i) = 1 - P_{j1}(\alpha_i)$ are the probabilities of getting a right answer and wrong answer on item j by examinee i with full attribute pattern α_i or reduced attribute pattern α_{ij}^*. For the G-DINA model, the probability of correctly answering item j by examinee i is given by de la Torre (2011) and Ma et al. (2016).

$$g\left[P_{j1}(\alpha_i)\right] = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{j1}\alpha_{ik}^*$$
$$+ \sum_{k=1}^{K_j^*-1}\sum_{k'=k+1}^{K_j^*} \delta_{jkk'} \cdot \alpha_{ik}^* \alpha_{ik'}^* + \dots + \delta_{j12\cdots K_j^*}\prod_{k=1}^{K_j^*}\alpha_{ik}^*, \quad (1)$$

where δ_{j0} is the intercept for item j, δ_{j1} is the main effect due to α_{ik}^*, $\delta_{jkk'}$ is the interaction effect due to α_{ik}^* and $\alpha_{ik'}^*$, and $\delta_{j12\cdots K_j^*}$ is the interaction effect due to $\alpha_{i1}^*, \dots, \alpha_{iK_j^*}^*$. In addition, link function $g[P_{ju}(\alpha_i)]$ can be formulated using the logit, log, and identity links. The logit link results in a general model which is equivalent to the log-linear CDM (Henson et al., 2009) and can be viewed as a special case of the general diagnostic model (GDM; von Davier, 2005,

2008). The resulting model from the log link function is referred to as the log CDM (de la Torre, 2011).

For the identity link, that is $g[P_{ju}(\alpha_i)] = P_{ju}(\alpha_i)$, the "deterministic input; noisy 'and' gate" (DINA) model (de la Torre & Douglas, 2004; Haertel, 1989; Junker & Sijtsma, 2001), the "deterministic input; noisy 'or' gate" (DINO) model (Templin & Henson, 2006), and the additive CDM (A-CDM) can be obtained from the G-DINA model when appropriate constraints are applied. For example, the item response function of the DINA model is $P_{j1}(\alpha_i) = \delta_{j0} + \delta_{j12\cdots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{ik}^*$, by setting all lower-order interaction terms to zero and by taking $\delta_{j0} = g_j$ and $\delta_{j12\cdots K_j^*} = 1 - s_j - g_j$. The parameter g_j is the probability of correctly guessing the answer if an examinee lacks at least one of the required attributes, and the parameter s_j refers to the probability of slipping and incorrectly answering the item if an examinee has mastered all the required attributes. The DINA model is a parsimonious and interpretable model that requires only two parameters for each item regardless of the number of attributes being considered.

Overview of Two Item Selection Indices for CD-CAT

SHE Procedure

After an item bank has been calibrated with a CDM, one must determine how to choose items for examinees from the item bank. CD-CAT employs algorithms to select items sequentially on the basis of examinee's responses, which is designed to classify student's attribute pattern accurately with a short test. The SHE procedure (Cheng, 2009) and the JSD index (Minchen & de la Torre, 2016; Yigit et al., 2018) are described below.

Suppose that the prior is chosen as $\pi_0(\alpha_c)$ for attribute pattern α_c, where $c = 1, 2, \ldots, 2^K$. For examinee i, suppose that t items are selected, the vector of corresponding item responses is denoted as $\mathbf{u}_i^{(t)}$, and a set $R_i^{(t)}$ represents the set of available items at this stage. The posterior distribution $\pi(\alpha_c \mid \mathbf{u}_i^{(t)})$ then becomes

$$\pi_{i,t}(\alpha_c) = \pi(\alpha_c \mid \mathbf{u}_i^{(t)}) = \frac{\pi_0(\alpha_c) L(\mathbf{u}_i^{(t)} \mid \alpha_c)}{\sum_{c=1}^{2^K} \pi_0(\alpha_c) L(\mathbf{u}_i^{(t)} \mid \alpha_c)}, \quad (2)$$

where $L(\mathbf{u}_i^{(t)} \mid \alpha_c)$ is the likelihood function, and it is the product of each item response function if the assumptions of local independence are satisfied. The SHE of the posterior distribution $\pi_{i,t}$ can then be written as

$$H(\pi_{i,t}) = -\sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c) \log \pi_{i,t}(\alpha_c). \quad (3)$$

Assuming that $\pi_{i,t}(\alpha_c)$ is an updated prior probability distribution and $U_{ij} = u$ is an item response for candidate item j in $R_i^{(t)}$, the posterior distribution $\pi_{i,t+1}$ then becomes

$$\pi_{i,t+1}(\alpha_c) = \pi_{i,t+1}(\alpha_c \mid \mathbf{u}_i^{(t)}, U_{ij} = u) = \frac{\pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)}{\sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)}. \quad (4)$$

From the last term, $\pi_{i,t+1}(\alpha_c \mid \mathbf{u}_i^{(t)}, U_{ij} = u)$ can be rewritten as $\pi_{i,t+1}(\alpha_c^{(t)} \mid U_{ij} = u)$, where $\alpha_c^{(t)}$ has an updated prior distribution $\pi_{i,t}$.

By considering the uncertainty of item response U_{ij}, the marginal probability distribution of item response on item j given the probability distribution $\pi_{i,t}$ can be computed as follows

$$P\left(U_{ij} = u \mid \mathbf{u}_i^{(t)}\right) = \sum_{c=1}^{2^K} P\left(U_{ij} = u, \alpha_c \mid \mathbf{u}_i^{(t)}\right)$$
$$= \sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c) P_{ju}(\alpha_c), \quad (5)$$

where the second term follows directly from $P(U_{ij} = u \mid \mathbf{u}_i^{(t)}, \alpha_c) = P(U_{ij} = u \mid \alpha_c)$, which can be derived from the assumption of local independence; as the current posterior distribution $\pi_{i,t}$ can be viewed as a new prior for $\alpha_c^{(t)}$ after having seen test data $\mathbf{u}_i^{(t)}$, $P(U_{ij} = u \mid \mathbf{u}_i^{(t)})$ can be simplified to $P(U_{ij} = u) = \sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c^{(t)}) P_{ju}(\alpha_c^{(t)})$ when $\alpha_c^{(t)}$ is substituted for α_c in the last term.

The next item to be selected for examinee i by the SHE is the one in $R_i^{(t)}$ that minimizes the expected SHE:

$$SHE_{ij} = \sum_{u=0}^{1} P\left(U_{ij} = u \mid \mathbf{u}_i^{(t)}\right) H\left(\pi_{i,t+1}\left(\alpha_c \mid \mathbf{u}_i^{(t)}, U_{ij} = u\right)\right). \quad (6)$$

As shown above, $P(U_{ij} = u \mid \mathbf{u}_i^{(t)})$ and $\pi_{i,t+1}(\alpha_c \mid \mathbf{u}_i^{(t)}, U_{ij} = u)$, respectively, become $P(U_{ij} = u)$ and $\pi_{i,t+1}(\alpha_c^{(t)} \mid U_{ij} = u)$. Let $H(\alpha_c^{(t)} \mid U_{ij})$ be the conditional entropy of $\alpha_c^{(t)}$ given U_{ij}. From the definition of conditional entropy (Cover & Thomas, 2006), $H(\alpha_c^{(t)} \mid U_{ij})$ is defined as the weighted sum of $H(\alpha_c^{(t)} \mid U_{ij} = u)$ over each possible value of u taken by the random variable U_{ij}, using $P(U_{ij} = u)$ as the weights. Thus, SHE_{ij} can be considered as the conditional entropy $H(\alpha_c^{(t)} \mid U_{ij})$ of $\alpha_c^{(t)}$ given U_{ij}.

JSD Index

The JSD as a new class of information measures based on the SHE was introduced by Lin (1991) to measure the overall difference of any finite number of distributions. Let $P_{ju}(\alpha_1), P_{ju}(\alpha_2), \ldots, P_{ju}(\alpha_{2^K})$ be 2^K item response functions with weights of prior probabilities $\pi_{i,t}(\alpha_1), \pi_{i,t}(\alpha_2), \ldots,$

$\pi_{i,t}(\alpha_{2^K})$, respectively. By the definition of the generalized JSD in Equation 5.1 of the paper of Lin (1991), or from Equations A.3, A.4, and A.5 in online appendices of the paper of Yigit et al. (2018), the JSD for item j can be written as

$$JSD_{ij} = H\left(P\left(U_{ij} = u \mid \mathbf{u}_i^{(t)}\right)\right) - \sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c) H\left(P_{ju}(\alpha_c)\right), \quad (7)$$

where

$$H\left(P\left(U_{ij} = u \mid \mathbf{u}_i^{(t)}\right)\right) = -\sum_{u=0}^{1}\left[\sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)\right]\log$$
$$\left[\sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)\right], \quad (8)$$

and

$$\sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c) H\left(P_{ju}(\alpha_c)\right) = \quad (9)$$
$$-\sum_{c=1}^{2^K}\left[\pi_{i,t}(\alpha_c)\left[\sum_{u=0}^{1} P_{ju}(\alpha_c)\log P_{ju}(\alpha_c)\right]\right].$$

The next item to be administered for examinee i is the one in $R_i^{(t)}$ that maximizes JSD_{ij}. Yigit et al. (2018) have proved that JSD_{ij} can be considered as the mutual information $I(\alpha_i; U_{ij}) = I(U_{ij}; \alpha_i)$ between the two discrete random variables of α_i and U_{ij}. Thus, the JSD index is also a measure of the amount of information one random variable α_i contains about another U_{ij}.

Similar results have been observed by Kang et al. (2017) within the framework of dual-objective CD-CAT (Kang et al., 2017; McGlohen & Chang, 2008; Wang et al., 2014; Zheng et al., 2018). For simultaneously estimating examinees' α_i and general ability θ_i, the dual-objective CD-CAT is based on a CDM and an item response theory model. In other words, item response U_{ij} has two Bernoulli distributions $P_{ju}(\alpha_i)$ and $P_{ju}(\theta_i)$. Take the two-parameter logistic model as an example, the probability of responding correctly to item j is defined as $P_{j1}(\theta_i) = 1/(1 + \exp(-Da_j(\theta_i - b_j)))$, that is, $P_{ju}(\theta_i) = \exp(uDa_j(\theta_i - b_j))/(1 + \exp(Da_j(\theta_i - b_j)))$. Here, D is a constant, a_j is the discrimination parameter, and b_j is the difficulty parameter. The JSD of dual-objective CD-CAT was defined as mutual information between the two discrete random variables U_{ij} and Z, where U_{ij} has a mixture distribution between $P_{ju}(\alpha)$ and $P_{ju}(\theta)$, and Z is the binary indicator variable for each distribution. For detailed information about relationship between the JSD, KL information, and Fisher information, please refer to Kang et al. (2017).

Relationship Between the SHE and the JSD

The purpose of this section is to establish the statement that the SHE and the JSD as two item selection criteria in CD-CAT are linearly related. Because the JSD itself is defined by the SHE, we apply the well-known relationship between the mutual information (or JSD) and SHE to establish a relationship between the item selection criteria that are developed using these two measures. The mutual information and SHE satisfy two well-known equations 2.43 and 2.44 from Theorem 2.4.1 in Cover and Thomas (2006, p. 21); that is, $I(X;Y) = H(X) - H(X \mid Y)$ and $I(X;Y) = H(Y) - H(Y \mid X)$, where $I(X;Y)$ and $H(X \mid Y)$ are respectively mutual information and conditional entropy for two random variables X and Y. Let $X = U_{ij}$ and $Y = \alpha_c^{(t)}$, with probability distributions $P(U_{ij} = u)$ and $\pi_{i,t}$. As $JSD_{ij} = I(U_{ij}; \alpha_c^{(t)})$ was proved in Yigit et al. (2018), we have $JSD_{ij} = I(U_{ij}; \alpha_c^{(t)}) = H(\alpha_c^{(t)}) - H(\alpha_c^{(t)} \mid U_{ij})$, which follows directly from the second well-known equation. As shown in the "SHE Procedure" section, SHE_{ij} can be written as the conditional entropy $H(\alpha_c^{(t)} \mid U_{ij})$. Thus, we have $JSD_{ij} = H(\pi_{i,t}) - SHE_{ij}$.

Next, we will provide an alternative way to prove the above statement, which would be useful for a better understanding of the relation. For simplicity, let the denominators or the normalizing constants of Equations 2 and 4 be $C_1 = \sum_{c=1}^{2^K} \pi_0(\alpha_c) L(\mathbf{u}_i^{(t)} \mid \alpha_c)$ and $C_2 = \sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)$. Note that the right-hand side of C_2 is relevant to u. The detailed mathematical steps are described below. Substituting Equations 4 and 5 into Equation 6, the SHE_{ij} can be written in the following equivalent form

$$SHE_{ij} = \sum_{u=0}^{1}\left[\sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)\right] H\left(\frac{\pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)}{C_2}\right). \quad (10)$$

By the definition of SHE, Equation 10 can be computed by

$$SHE_{ij} = \sum_{u=0}^{1}\left[\sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)\right]$$
$$\left[-\sum_{c=1}^{2^K} \frac{\pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)}{C_2} \log \frac{\pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)}{C_2}\right] \quad (11)$$

Recall from two basic logarithmic properties that the log of a quotient is equal to the difference between the logs of the numerator and denominator, and the log of a product is equal to the sum of the logs of the factors. Equation 11 can be written as

$$SHE_{ij} = \sum_{u=0}^{1}\left[\sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)\right]$$
$$\left[-\sum_{c=1}^{2^K} \frac{\pi_{i,t}(\alpha_c) P_{ju}(\alpha_c)}{C_2}\left(\log \pi_{i,t}(\alpha_c) + \log P_{ju}(\alpha_c) - \log C_2\right)\right].$$
$$(12)$$

Notice C_2 and $\log C_2$ can be factored out from the third summation, as they remain constant over the summation index c from 1 to 2^K. Hence, Equation 12 has the form

$$SHE_{ij} = \sum_{u=0}^{1}\left[-\sum_{c=1}^{2^K}\left(\pi_{i,t}\left(\alpha_c\right)P_{ju}\left(\alpha_c\right)\left(\log \pi_{i,t}\left(\alpha_c\right)+\log P_{ju}\left(\alpha_c\right)\right)\right)\right]$$
$$+ \sum_{u=0}^{1}C_2 \log C_2,$$

$$(13)$$

since $C_2 = \sum_{c=1}^{2^K}\pi_{i,t}\left(\alpha_c\right)P_{ju}\left(\alpha_c\right)$.

After changing the order of the summation and factoring two constant terms (i.e., $\pi_{i,t}(\alpha_c)\log \pi_{i,t}(\alpha_c)$ and $\pi_{i,t}(\alpha_c)$) out from the new second summation, the first term on the right-hand side of the Equation 13 can be written as in the following equivalent form:

$$-\sum_{c=1}^{2^K}\left[\pi_{i,t}\left(\alpha_c\right)\log \pi_{i,t}(\alpha_c)\sum_{u=0}^{1}P_{ju}\left(\alpha_c\right)\right]$$
$$-\sum_{c=1}^{2^K}\left[\pi_{i,t}\left(\alpha_c\right)\left[\sum_{u=0}^{1}P_{ju}\left(\alpha_c\right)\log P_{ju}\left(\alpha_c\right)\right]\right]$$
$$= -\sum_{c=1}^{2^K}\left[\pi_{i,t}\left(\alpha_c\right)\log \pi_{i,t}\left(\alpha_c\right)\right]+\sum_{c=1}^{2^K}\pi_{i,t}\left(\alpha_c\right)H\left(P_{ju}\left(\alpha_c\right)\right),$$

$$(14)$$

which follows from $\sum_{u=0}^{1}P_{ju}(\alpha_c)=1$ and Equation 9.

Based on Equations 3, 13, and 14, the SHE_{ij} can be written as

$$SHE_{ij} = H\left(\pi_{i,t}\right)+\sum_{c=1}^{2^K}\pi_{i,t}\left(\alpha_c\right)H\left(P_{ju}\left(\alpha_c\right)\right)+\sum_{u=0}^{1}C_2 \log C_2.$$

$$(15)$$

From Equation 8, the third term on the right-hand side of Equation 15 is equal to $-H(P(U_{ij}=u\mid \mathbf{u}_i^{(t)}))$, since $C_2 = \sum_{c=1}^{2^K}\pi_{i,t}\left(\alpha_c\right)P_{ju}\left(\alpha_c\right)$. Then based on Equations 7 and 15, the SHE_{ij} can be written as

$$SHE_{ij} = H\left(\pi_{i,t}\right)-JSD_{ij}, \qquad (16)$$

which can be rewritten as

$$JSD_{ij} = H\left(\pi_{i,t}\right)-SHE_{ij}. \qquad (17)$$

The meaning of Equation 17 is consistent with the fact that the JSD or mutual information is a special case of a more general quantity called relative entropy. As $H(\pi_{i,t})$ is not a function of both item parameters and item responses of the candidate item j, it is a constant for examinee i. Thus,

minimizing SHE_{ij} is equivalent to maximizing JSD_{ij} because they select the same item for administration at the $(t+1)$ th stage of testing. So far, this completes the proof that the SHE and the JSD have a linear relationship under the G-DINA model. Note that two proofs of the relationship between the SHE and JSD just rely on the form of an item response function for a binary response variable and do not depend on any particular CDM, because the item response function $P_{ju}(\alpha_i)$, the prior distribution $\pi_0(\alpha_c)$, and the current item response vector $\mathbf{u}_i^{(t)}$ are all we need to calculate SHE and mutual information.

Tables A1 and A2 in the appendix are presented for the illustration of computation of values of indices for the SHE and JSD. Here, the SHE and JSD are computed for two items with different item response distributions or different item parameters, where a discrete uniform prior distribution for attribute patterns was used. From these two tables, the relationship between the SHE and JSD for the two items satisfied Equation 16 or Equation 17, and minimizing SHE_{ij} is equivalent to maximizing JSD_{ij}. Empirical examples show that the two item selection criteria are expected to behave similarly in CD-CAT.

Simulation Study

Design

A small-scale simulation study was conducted to compare the performance of the SHE and JSD. Following a design similar to that in Cheng (2009) and Xu et al. (2016), the DINA model and five independent attributes were considered in the simulation study. For the generation of four item banks, a Q-matrix for 300 items should be first simulated. The entries of the Q-matrix were generated item by item and attribute by attribute. Each item has 20% chance of measuring each attribute. Four item banks were considered: (a) slipping and guessing parameters were fixed as one of the three levels, such as 0.05, 0.1, or 0.2, and (b) both slipping and guessing parameters were randomly draw from a uniform distribution on the interval [0.1, 0.3]. Test length was either fixed at 5 or 10 items. The sample size of examinees was set to 2,000. Attribute patterns for all examinees were randomly drawn from all possible attribute patterns with equal probability. Details of simulation design are presented in Table A3 in the appendix.

To consider the impact of the simulation of item responses in CD-CAT on the performance of the SHE and JSD, two types of CAT simulation were considered: full simulations or post hoc simulations (Magis et al., 2017). In case of a full CAT simulation, an item response for examinee i on item j was randomly drawn from a Bernoulli distribution, denoted by Bernoulli($P_{j1}(\alpha_i)$). Full CAT simulations imply that item responses for examinee i on the same set of items may be different. Under the post hoc simulation scenario, a complete item response matrix was created first for all examinees on each item bank before CD-CAT administrations and the

Table 1. Mean and Standard Deviation (in brackets) of Attribute and Pattern Recovery Rate for Slipping and Guessing Parameters of 0.05.

			Attribute					
Simulations	Test length	Method	1	2	3	4	5	Pattern
Post hoc	5	SHE	0.950 (0.008)	0.950 (0.007)	0.950 (0.008)	0.950 (0.008)	0.950 (0.008)	0.774 (0.011)
		JSD	0.950 (0.008)	0.950 (0.007)	0.950 (0.008)	0.950 (0.008)	0.950 (0.008)	0.774 (0.011)
	10	SHE	0.985 (0.005)	0.987 (0.006)	0.988 (0.005)	0.989 (0.006)	0.986 (0.007)	0.941 (0.009)
		JSD	0.984 (0.005)	0.990 (0.006)	0.987 (0.005)	0.989 (0.006)	0.986 (0.007)	0.942 (0.009)
Full	5	SHE	0.950 (0.008)	0.951 (0.007)	0.950 (0.008)	0.951 (0.007)	0.951 (0.008)	0.775 (0.012)
		JSD	0.951 (0.007)	0.950 (0.007)	0.950 (0.008)	0.950 (0.007)	0.950 (0.009)	0.773 (0.009)
	10	SHE	0.985 (0.005)	0.987 (0.006)	0.988 (0.006)	0.989 (0.006)	0.986 (0.007)	0.942 (0.009)
		JSD	0.984 (0.005)	0.989 (0.005)	0.987 (0.005)	0.989 (0.005)	0.986 (0.007)	0.943 (0.011)

Note. SHE = Shannon entropy; JSD = Jensen–Shannon divergence.

Table 2. Mean and Standard Deviation (in brackets) of Attribute and Pattern Recovery Rate for Slipping and Guessing Parameters of 0.1.

			Attribute					
Simulations	Test length	Method	1	2	3	4	5	Pattern
Post hoc	5	SHE	0.901 (0.008)	0.900 (0.007)	0.898 (0.008)	0.902 (0.008)	0.898 (0.008)	0.589 (0.011)
		JSD	0.901 (0.008)	0.900 (0.007)	0.898 (0.008)	0.902 (0.008)	0.898 (0.008)	0.589 (0.011)
	10	SHE	0.951 (0.005)	0.955 (0.006)	0.954 (0.005)	0.963 (0.006)	0.952 (0.007)	0.816 (0.009)
		JSD	0.949 (0.005)	0.955 (0.006)	0.952 (0.005)	0.962 (0.006)	0.955 (0.007)	0.814 (0.009)
Full	5	SHE	0.897 (0.008)	0.901 (0.007)	0.898 (0.008)	0.902 (0.007)	0.900 (0.008)	0.591 (0.012)
		JSD	0.899 (0.007)	0.900 (0.007)	0.900 (0.008)	0.898 (0.007)	0.898 (0.009)	0.587 (0.009)
	10	SHE	0.950 (0.005)	0.955 (0.006)	0.954 (0.006)	0.962 (0.006)	0.954 (0.007)	0.815 (0.009)
		JSD	0.947 (0.005)	0.958 (0.005)	0.952 (0.005)	0.961 (0.005)	0.957 (0.007)	0.816 (0.011)

Note. SHE = Shannon entropy; JSD = Jensen–Shannon divergence.

responses to the selected items for the SHE or JSD were drawn from the complete item response matrix. First of all, post hoc simulations were considered to make use of exactly the same item responses for two item selection methods (SHE and JSD) under each item bank in CD-CAT. We replicated each type of simulation process 100 times under each condition and recorded final estimates of attribute patterns for all examinees.

Results

The attribute-level recovery rate is defined as the proportion of each attribute that is correctly identified. The pattern recovery rate is defined as the proportion of entire attribute pattern that is correctly recovered. Mean and standard deviation of attribute-level and pattern recovery rates for each level of item parameters are shown in Tables 1–4. For the SHE, our results are consistent with the results of Xu et al. (2016). Consistent with our theoretical conclusions, the SHE and JSD behaved quite similarly, because their attribute-level and pattern recovery rates were very close to each other under all conditions.

Figure 1 presents pattern recovery rates for different test lengths and simulation types under slipping and guessing parameters of U(0.1, 0.3). From the two top panels of Figure 1, it can be observed that pattern recovery rates obtained by the SHE and JSD are the same for each replication under post hoc simulations. While for the full simulations, pattern recovery rates for the SHE and JSD were different for each replication because different item responses had an impact on test item selection. When the test length was 5, 43% pattern recovery rates of the SHE were higher than the mean of pattern recovery rates of the JSD, and 51% pattern recovery rates of the JSD were higher than the mean of pattern recovery rates of the SHE. The percentages became 49% and 60%, respectively, when the test length was 10. This result is consistent with the previous finding: "The mutual information item selection algorithm generates nearly the most accurate attribute pattern recovery in more than half of the conditions" (Wang, 2013, p. 1030).

We also checked whether two item selection algorithms selected the same set of items for each examinee under post hoc simulations. For the first three item banks, two item

Table 3. Mean and Standard Deviation (in brackets) of Attribute and Pattern Recovery Rate for Slipping and Guessing Parameters of 0.2.

| Simulations | Test length | Method | Attribute | | | | | Pattern |
			1	2	3	4	5	
Post hoc	5	SHE	0.800 (0.008)	0.807 (0.007)	0.800 (0.008)	0.797 (0.008)	0.799 (0.008)	0.329 (0.011)
		JSD	0.800 (0.008)	0.807 (0.007)	0.800 (0.008)	0.797 (0.008)	0.799 (0.008)	0.329 (0.011)
	10	SHE	0.845 (0.005)	0.870 (0.006)	0.854 (0.005)	0.871 (0.006)	0.853 (0.007)	0.508 (0.009)
		JSD	0.847 (0.005)	0.870 (0.006)	0.851 (0.005)	0.868 (0.006)	0.855 (0.007)	0.509 (0.009)
Full	5	SHE	0.801 (0.008)	0.802 (0.007)	0.803 (0.008)	0.799 (0.007)	0.800 (0.008)	0.330 (0.012)
		JSD	0.801 (0.007)	0.801 (0.007)	0.797 (0.008)	0.798 (0.007)	0.803 (0.009)	0.329 (0.009)
	10	SHE	0.844 (0.005)	0.865 (0.006)	0.850 (0.006)	0.873 (0.006)	0.855 (0.007)	0.505 (0.009)
		JSD	0.850 (0.005)	0.865 (0.005)	0.847 (0.005)	0.867 (0.005)	0.853 (0.007)	0.504 (0.011)

Note. SHE = Shannon entropy; JSD = Jensen–Shannon divergence.

Table 4. Mean and Standard Deviation (in brackets) of Attribute and Pattern Recovery Rate for Slipping and Guessing Parameters of $U(0.1, 0.3)$.

| Simulations | Test length | Method | Attribute | | | | | Pattern |
			1	2	3	4	5	
Post hoc	5	SHE	0.884 (0.008)	0.882 (0.007)	0.848 (0.008)	0.897 (0.008)	0.838 (0.008)	0.530 (0.011)
		JSD	0.884 (0.008)	0.882 (0.007)	0.848 (0.008)	0.897 (0.008)	0.838 (0.008)	0.530 (0.011)
	10	SHE	0.953 (0.005)	0.946 (0.006)	0.939 (0.005)	0.937 (0.006)	0.913 (0.007)	0.747 (0.009)
		JSD	0.953 (0.005)	0.946 (0.006)	0.939 (0.005)	0.937 (0.006)	0.913 (0.007)	0.747 (0.009)
Full	5	SHE	0.884 (0.008)	0.882 (0.007)	0.846 (0.008)	0.898 (0.007)	0.838 (0.008)	0.529 (0.012)
		JSD	0.884 (0.007)	0.880 (0.007)	0.846 (0.008)	0.898 (0.007)	0.838 (0.009)	0.529 (0.009)
	10	SHE	0.953 (0.005)	0.946 (0.006)	0.939 (0.006)	0.938 (0.006)	0.914 (0.007)	0.747 (0.009)
		JSD	0.953 (0.005)	0.946 (0.005)	0.940 (0.005)	0.938 (0.005)	0.915 (0.007)	0.749 (0.011)

Note. SHE = Shannon entropy; JSD = Jensen–Shannon divergence.

selection algorithms based on the SHE and JSD indeed selected the same set of items but with slightly different orders. Because all test items in these item banks have the same values of item parameters, some items presented in different positions have the same value of SHE or JSD. For example, two items with same item parameters but a single distinct attribute may have the same value of SHE or JSD. For the fourth item bank, we found that two item selection algorithms based on the SHE and JSD selected the same set of items.

Discussion

In this study, we complete the proof that the SHE procedure and the JSD are linearly related under CDMs. In other words, we showed that minimizing JSD and maximizing SHE can be used interchangeably because they will select the same items in CD-CAT. The two measures are linearly related but they are not equal, meaning that two measures have the form $JSD_{ij} = H(\pi_{i,t}) - SHE_{ij}$. Although they are not equal, item selection methods based on the SHE and JSD will randomly select one of test items with the same value of the SHE or JSD index for administration at the next stage of testing, because $H(\pi_{i,t})$ is a constant for examinee i and minimizing SHE_{ij} is equivalent to maximizing JSD_{ij} at the current stage of testing.

This study is not without limitations. Theoretically, SHE, KL information, and mutual information are three ways to measure the uncertainty, and they are related to each other. It would be interesting to further investigate relationships of item selection indices based on the KL information, the SHE, the JSD, and other indices under general dichotomous or polytomous CDMs. For example, the GDI and MPWKL might be related, because they perform similarly and better than the PWKL in terms of correct attribute classification rates or test lengths. We believe GDI is simply a weighted variance of the probabilities of success of an item associated with attribute patterns given an attribute pattern distribution, and therefore we can start with comparing the weighted KL with the weighted variance to show a relationship. The GDI is defined as follows (Kaplan et al., 2015):
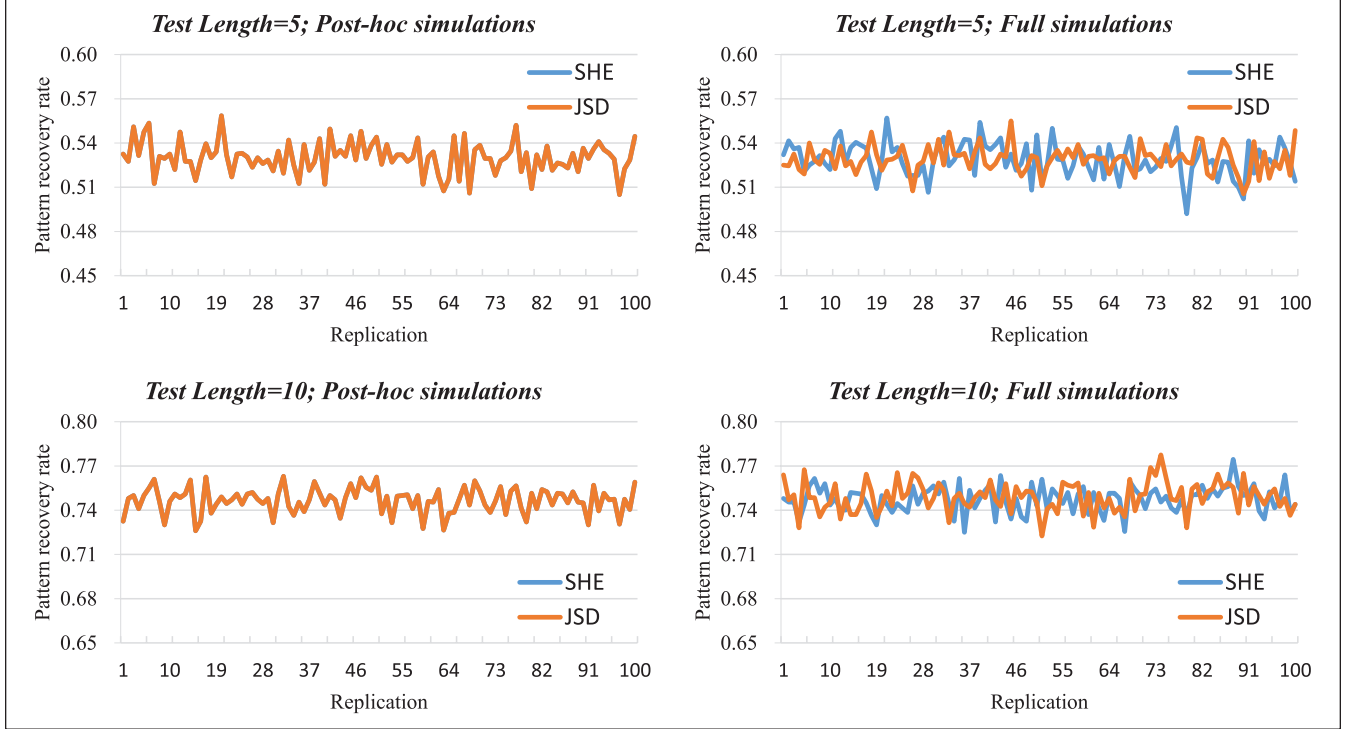
Figure 1. Pattern recovery rate for different test lengths and simulation types under slipping and guessing parameters of $U(0.1, 0.3)$.

$$\zeta_{ij}^2 = \sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c)\left[P_{j1}(\alpha_c) - \bar{P}_{ij}\right]^2, \tag{18}$$

where $\bar{P}_{ij} = \sum_{c=1}^{2^K} \pi_{i,t}(\alpha_c)P_{j1}(\alpha_c)$. The MPWKL is defined as follows (Kaplan et al., 2015):

$$MPWKL_{ij} =$$
$$\sum_{d=1}^{2^K}\left[\sum_{c=1}^{2^K}\left[\sum_{u=0}^{1}\left(\log\frac{P_{ju}(\alpha_d)}{P_{ju}(\alpha_c)}\right)P_{ju}(\alpha_d)\pi_{i,t}(\alpha_c)\right]\pi_{i,t}(\alpha_d)\right]. \tag{19}$$

The following algebraic procedures will simplify the calculations in Equation 19 above

$$MPWKL_{ij} =$$
$$\sum_{d=1}^{2^K}\sum_{c=1}^{2^K}\sum_{u=0}^{1}\left(\log P_{ju}(\alpha_d) - \log P_{ju}(\alpha_c)\right)P_{ju}(\alpha_d)\pi_{i,t}(\alpha_c)\pi_{i,t}(\alpha_d) \tag{20}$$

$$= \sum_{d=1}^{2^K}\sum_{c=1}^{2^K}\sum_{u=0}^{1}P_{ju}(\alpha_d)\pi_{i,t}(\alpha_c)\pi_{i,t}(\alpha_d)\log P_{ju}(\alpha_d)$$
$$-\sum_{d=1}^{2^K}\sum_{c=1}^{2^K}\sum_{u=0}^{1}P_{ju}(\alpha_d)\pi_{i,t}(\alpha_c)\pi_{i,t}(\alpha_d)\log P_{ju}(\alpha_c) \tag{21}$$

$$= \sum_{d=1}^{2^K}\sum_{u=0}^{1}\pi_{i,t}(\alpha_d)P_{ju}(\alpha_d)\log P_{ju}(\alpha_d)$$
$$-\sum_{c=1}^{2^K}\sum_{u=1}^{1}\sum_{d=0}^{2^K}P_{ju}(\alpha_d)\pi_{i,t}(\alpha_d)\pi_{i,t}(\alpha_c)\log P_{ju}(\alpha_c) \tag{22}$$

$$= \sum_{c=1}^{2^K}\sum_{u=0}^{1}\pi_{i,t}(\alpha_c)P_{ju}(\alpha_c)\log P_{ju}(\alpha_c)$$
$$-\sum_{c=1}^{2^K}\sum_{u=0}^{1}\left[\pi_{i,t}(\alpha_c)\log P_{ju}(\alpha_c)\left(\sum_{d=1}^{2^K}P_{ju}(\alpha_d)\pi_{i,t}(\alpha_d)\right)\right] \tag{23}$$

$$= \sum_{c=1}^{2^K}\pi_{i,t}(\alpha_c)\sum_{u=0}^{1}\left[P_{ju}(\alpha_c) - \sum_{d=1}^{2^K}P_{ju}(\alpha_d)\pi_{i,t}(\alpha_d)\right]\log P_{ju}(\alpha_c) \tag{24}$$

$$= \sum_{c=1}^{2^K}\pi_{i,t}(\alpha_c)\left[P_{j1}(\alpha_c) - \bar{P}_{ij}\right]\left[\log P_{j1}(\alpha_c) - \log\left(1 - P_{j1}(\alpha_c)\right)\right], \tag{25}$$

where Equation 20 follows from the quotient rule of logarithmic properties and Equation 25 follows directly from the complement rule in probability, as expressed by the equation $P_{j0}(\alpha_c) = 1 - P_{j1}(\alpha_c)$. An easy way to calculate the MPWKL

is provided by Equation 25. From Equations 18 and 25, both the GDI and MPWKL are a function of $\pi_{i,t}(\alpha_c)(P_{j1}(\alpha_c) - \bar{P}_{ij})$, but with different weights $\log P_{j1}(\alpha_c) - \log(1 - P_{j1}(\alpha_c))$ and $P_{ju}(\alpha_c) - \bar{P}_{ij}$. Thus, we know that the GDI is closely related to the MPWKL.

The findings of this study may contribute to the growing literature on formative assessments. First, theoretical derivations and empirical examples have shown that both indices (SHE and JSD) are expected to select the same next item given item response pattern of the same set of previous test items in CD-CAT. Consistent with our theoretical conclusions, simulation results have shown that the SHE and JSD behaved quite similarly in terms of attribute-level and pattern recovery rates. This finding can possibly be useful to help practitioners to choose an effective item selection algorithm (SHE or JSD) in the development and application of CD-CAT system in the field of educational and psychological measurement. Second,

the effectiveness of item selection algorithm in CD-CAT will impact the quality of curriculum delivery and the outcomes of learning. If individual diagnosis results with a high measurement precision can be provided by using an effective item selection algorithm of CD-CAT, then diverse instructional materials can cater to the diverse needs or specific knowledge status of all learners (Lashley, 2019). Finally, information-based indices are now not only widely applied in CD-CAT, but also useful for any test construction stage where test items are selected based on their statistical characteristics (e.g., Henson & Douglas, 2005; Henson et al., 2008; Kuo et al., 2016). For example, the cognitive diagnostic index, the attribute-level discrimination index, and their modified indices as KL information based measures have been used for the construction of diagnostic tests. Future research on automated test assembly for cognitive diagnosis will expand the scope of the application of the current finding.

Appendix

Table A1. Example 1 for the Illustration of Computation of Values of Indices for the SHE and JSD.

Function	$\alpha_1 = (0,0)$	$\alpha_2 = (1,0)$	$\alpha_3 = (0,1)$	$\alpha_4 = (1,1)$	SUM
$\pi_{i,0}(\alpha)$	0.25	0.25	0.25	0.25	1.00
$-\pi_{i,0}(\alpha)\log(\pi_{i,0}(\alpha))$	0.15	0.15	0.15	0.15	$H(\pi_{i,0}) = 0.60$
$P_{j1}(\alpha)$	0.12	0.95	0.33	0.95	2.35
$P_{j0}(\alpha) = 1 - P_{j1}(\alpha)$	0.88	0.05	0.67	0.05	1.65
$H(P_{j.}(\alpha)) = -\sum_{u=0}^{1} P_{ju}(\alpha)\log(P_{ju}(\alpha))$	0.16	0.09	0.28	0.09	0.61
$\pi_{i,0}(\alpha)P_{j1}(\alpha)$	0.03	0.24	0.08	0.24	$C_{21} = 0.59$
$\pi_{i,0}(\alpha)P_{j0}(\alpha)$	0.22	0.01	0.17	0.01	$C_{20} = 0.41$
$\pi_{i,0}(\alpha)H(P_{j.}(\alpha))$	0.04	0.02	0.07	0.02	$S_1 = 0.15$
$\pi_{i,1}(\alpha \mid U_{ij} = 1) = \pi_{i,0}(\alpha)P_{j1}(\alpha) / C_{21}$	0.05	0.40	0.14	0.40	1.00
$\pi_{i,1}(\alpha \mid U_{ij} = 0) = \pi_{i,0}(\alpha)P_{j1}(\alpha) / C_{20}$	0.53	0.03	0.41	0.03	1.00
$-\pi_{i,1}(\alpha \mid U_{ij} = 1)\log(\pi_{i,1}(\alpha \mid U_{ij} = 1))$	0.07	0.16	0.12	0.16	$H_1 = 0.50$
$-\pi_{i,1}(\alpha \mid U_{i1} = 0)\log(\pi_{i,1}(\alpha \mid U_{i1} = 0))$	0.15	0.05	0.16	0.05	$H_0 = 0.40$

$SHE_{ij} = C_{21}H_1 + C_{20}H_0 = 0.46$

$JSD_{ij} = -(C_{21}\log(C_{21}) + C_{20}\log(C_{20})) - S_1 = 0.14$

$JSD_{ij} = H(\pi_{i,0}) - SHE_{ij} = 0.60 - 0.46 = 0.14$

Note. The base of the logarithm is 10.

Table A2. Example 2 for the Illustration of Computation of Values of Indices for the SHE and JSD.

Function	$\alpha_1 = (0,0)$	$\alpha_2 = (1,0)$	$\alpha_3 = (0,1)$	$\alpha_4 = (1,1)$	SUM
$\pi_{i,0}(\alpha)$	0.25	0.25	0.25	0.25	1.00 $H(\pi_{i,0}) = 0.60$
$-\pi_{i,0}(\alpha)\log(\pi_{i,0}(\alpha))$	0.15	0.15	0.15	0.15	
$P_{j1}(\alpha)$	0.33	0.47	0.33	0.77	1.90
$P_{j0}(\alpha) = 1 - P_{j1}(\alpha)$	0.67	0.53	0.67	0.23	2.10
$H(P_{j.}(\alpha)) = -\sum_{u=0}^{1} P_{ju}(\alpha)\log(P_{ju}(\alpha))$	0.28	0.30	0.28	0.23	1.09 $C_{21} = 0.48$
$\pi_{i,0}(\alpha)P_{j1}(\alpha)$	0.08	0.12	0.08	0.19	$C_{20} = 0.53$
$\pi_{i,0}(\alpha)P_{j0}(\alpha)$	0.17	0.13	0.17	0.06	$S_1 = 0.27$
$\pi_{i,0}(\alpha)H(P_{j.}(\alpha))$	0.07	0.08	0.07	0.06	
$\pi_{i,1}(\alpha \mid U_{ij} = 1) = \pi_{i,0}(\alpha)P_{j1}(\alpha) / C_{21}$	0.17	0.25	0.17	0.41	1.00
$\pi_{i,1}(\alpha \mid U_{ij} = 0) = \pi_{i,0}(\alpha)P_{j1}(\alpha) / C_{20}$	0.32	0.25	0.32	0.11	1.00 $H_1 = 0.57$
$-\pi_{i,1}(\alpha \mid U_{ij} = 1)\log(\pi_{i,1}(\alpha \mid U_{ij} = 1))$	0.13	0.15	0.13	0.16	$H_0 = 0.57$
$-\pi_{i,1}(\alpha \mid U_{i1} = 0)\log(\pi_{i,1}(\alpha \mid U_{i1} = 0))$	0.16	0.15	0.16	0.11	

$$SHE_{ij} = C_{21}H_1 + C_{20}H_0 = 0.57$$

$$JSD_{ij} = -(C_{21}\log(C_{21}) + C_{20}\log(C_{20})) - S_1 = 0.03$$

$$JSD_{ij} = H(\pi_{i,0}) - SHE_{ij} = 0.60 - 0.57 = 0.03$$

Note. The base of the logarithm is 10.

Table A3. Details of Simulation Design.

Factors	Details
Attribute structure	Independent structure with five attributes
CDM	The DINA model
Examinees	Sample size is 2,000
	Attribute patterns are generated by taking one of the 2^5 possible patterns with equal probability
Item banks	Each of four item banks consists of 300 items
	Each item has 20% chance of measuring each attribute
	Item parameters are set to $s = g = 0.05$, $s = g = 0.1$, $s = g = 0.2$, or $s \sim U(0.1, 0.3)$ and $g \sim U(0.1, 0.3)$
CD-CAT	Test length is either fixed at 5 or 10 items
	Two item selection indices are the SHE and JSD with a prior uniform distribution
	MLE method is used to estimate attribute patterns of examinees
Simulations	Full simulations or post hoc simulations are used to generate item responses

Note. CDM = cognitive diagnostic model; DINA = deterministic inputs, noisy "and" gate; CD-CAT = cognitive diagnostic computerized adaptive testing; SHE = Shannon entropy; JSD = Jensen–Shannon divergence; MLE = maximum likelihood estimation.

ORCID iD

Wenyi Wang https://orcid.org/0000-0002-2677-4405

References

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa*, *80*(2), 1–13.

Chang, H.-H. (2012). Making computerized adaptive testing diagnostic tools for schools. In R. W. Lissitz & H. Jiao (Eds.), *Computers and their impact on state assessment: Recent history and predictions for the future* (pp. 195–226). Information Age Publisher.

Chang, H.-H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika*, *80*(1), 1–20.

Chang, H.-H., & Ying, Z. (2007). Computerized adaptive testing. In N. J. Salkind & K. Rasmussen (Eds.), *Encyclopedia of measurement and statistics* (Vol. 1, pp. 170–173). Sage.

Chen, Y., & Chang, H.-H. (2018). Psychometrics help learning: From assessment to learning. *Applied Psychological Measurement*, *42*(1), 3–4.

Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika*, *74*(4), 619–632.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). John Wiley.

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.

Gierl, M. J., & Lai, H. (2018). Using automatic item generation to create solutions and rationales for computerized formative testing. *Applied Psychological Measurement*, *42*(1), 42–57.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*(4), 301–321.

Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement*, *29*(4), 262–277.

Henson, R., Roussos, L., Douglas, J., & He, X. (2008). Cognitive diagnostic attribute-level discrimination indices. *Applied Psychological Measurement*, *32*(4), 275–288.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191–210.

Jang, E. E. (2008). A framework for cognitive diagnostic assessment. In C. A. Chapelle, Y. R. Chung, & J. Xu. (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 117–131), Iowa State University.

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., & 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., & Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, *39*(3), 167–188.

Kuo, B.-C., Pai, H.-S., & de la Torre, J. (2016). Modified cognitive diagnostic index and modified attribute-level discrimination index for test construction. *Applied Psychological Measurement*, *40*(5), 315–330.

Lashley, L. (2019). A reflective analysis of the selection and production of instructional material for curriculum delivery at the primary level in postcolonial Guyana. *SAGE Open*, *9*(2), 1–15.

Laveault, D., & Allal, L. (2016). Implementing assessment for learning: Theoretical and practical issues. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 1–20). Springer.

Leighton, J. P., & Gierl, M. J. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.

Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, *37*(1), 145–151.

Liu, J., Ying, Z., & Zhang, S. (2015). A rate function approach to computerized adaptive testing for cognitive diagnosis. *Psychometrika*, *80*(2), 468–490.

Ma, W., Iaconangelo, C., & de la Torre, J. (2016). Model similarity, model selection, and attribute classification. *Applied Psychological Measurement*, *40*(3), 200–217.

Magis, D., Yan, D., & von Davier, A. A. (2017). *Computerized adaptive and multistage testing with R: Using packages catR and mstR*. Springer.

McGlohen, M. K., & Chang, H.-H. (2008). Combining computer adaptive testing technology with cognitively diagnostic assessment. *Behavior Research Methods*, *40*(3), 808–821.

Minchen, N., & de la Torre, J. (2016, July). *The continuous G-DINA model and the Jensen–Shannon divergence* [Paper presentation]. Paper presented at the International Meeting of the Psychometric Society, Asheville, NC, United States.

Quellmalz, E. S., & Pellegrino, J. W. (2009). Technology and testing. *Science*, *323*(5910), 75–79.

Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement*, *6*(4), 219–262.

Rupp, A. A., Templin, J. L., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. 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.

Tatsuoka, C., & Ferguson, T. (2003). Sequential classification on partially ordered sets. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *65*(1), 143–157.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345–354.

Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. Taylor & Francis.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287–305.

U.S. Department of Education. (2017). *Reimagining the role of technology in education: 2017 National Education Technology Plan update*. https://tech.ed.gov/files/2017/01/NETP17.pdf

von Davier, M. (2005). *A general diagnostic model applied to language testing data* (ETS Research Report RR-05-16). Educational Testing Service.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 287–307.

Wang, C. (2013). Mutual information item selection method in cognitive diagnostic computerized adaptive testing with short test length. *Educational and Psychological Measurement*, *73*(6), 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*(3), 255–273.

Wang, C., Zheng, C., & 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.

Wang, W. Y., Ding, S. L., & Song, L. H. (2015). New item-selection methods for balancing test efficiency against item-bank usage efficiency in CD-CAT. *Springer Proceedings in Mathematics & Statistics*, *89*, 133–151.

Xu, G., Wang, C., & Shang, Z. (2016). On initial item selection in cognitive diagnostic computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, *69*(3), 291–315.

Xu, X. L., Chang, H. H., & Douglas, J. (2003). *A simulation study to compare CAT strategies for cognitive diagnosis* [Paper presentation]. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, United States.

Yigit, H. D., Sorrel, M. A., de la Torre, J. (2018). Computerized adaptive testing for cognitively based multiple-choice data. *Applied Psychological Measurement*, *43*(5), 388–401. https://doi.org/10.1177/0146621618798665

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

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

Author Biographies

Wenyi Wang is an associate professor at Jiangxi Normal University. His primary research interests include cognitive diagnostic assessment and computerized adaptive testing.

Lihong Song is an associate professor at Jiangxi Normal University. Her primary research interests include cognitive diagnostic assessment and application of statistical methods to education.

Teng Wang is a postgraduate student at Jiangxi Normal University. His primary research interests include cognitive diagnostic assessment and computerized adaptive testing.

Peng Gao is a postgraduate student at Jiangxi Normal University. His primary research interests include cognitive diagnostic assessment and computerized adaptive testing.

Jian Xiong is a postgraduate student at Jiangxi Normal University. His primary research interests include item response theory and computerized adaptive testing.