Article

Applied Psychological Measurement
2015, Vol. 39(6) 465–479

DOI: 10.1177/0146621615577087
apm.sagepub.com
SAGE

Consistency of Cluster Analysis for Cognitive Diagnosis: The DINO Model and the DINA Model Revisited

Chia-Yi Chiu[1] and Hans-Friedrich Köhn[2]

Abstract

The Asymptotic Classification Theory of Cognitive Diagnosis (ACTCD) developed by Chiu, Douglas, and Li proved that for educational test data conforming to the Deterministic Input Noisy Output "AND" gate (DINA) model, the probability that hierarchical agglomerative cluster analysis (HACA) assigns examinees to their true proficiency classes approaches 1 as the number of test items increases. This article proves that the ACTCD also covers test data conforming to the Deterministic Input Noisy Output "OR" gate (DINO) model. It also demonstrates that an extension to the statistical framework of the ACTCD, originally developed for test data conforming to the Reduced Reparameterized Unified Model or the General Diagnostic Model (a) is valid also for both the DINA model and the DINO model and (b) substantially increases the accuracy of HACA in classifying examinees when the test data conform to either of these two models.

Keywords

cognitive diagnosis, classification, cluster analysis, consistency, DINA model, DINO model, heuristics

Cognitive diagnosis models (CDMs; Rupp, Templin, & Henson, 2010) of educational test performance decompose an examinee's overall ability into a set of specific discrete skills, called *attributes*, each of which he or she may or may not have mastered, thereby providing a detailed description, or *attribute profile*, of his or her strengths and weaknesses in the ability domain of the test. The entire set of possible attribute profiles for a given test defines classes of intellectual proficiency to which examinees can be assigned.

Current methods of fitting CDMs to educational test data generally use maximum likelihood estimation (MLE) procedures such as Expectation Maximization (EM) or Markov chain Monte Carlo (MCMC) to estimate model parameters that are then used to assign examinees to proficiency classes. These procedures often encounter difficulties in practice. For example, to obtain reliable parameter estimates, MLE procedures typically require large samples of examinees that

[1]Rutgers, The State University of New Jersey, New Brunswick, USA
[2]University of Illinois at Urbana–Champaign, USA

Corresponding Author:
Hans-Friedrich Köhn, University of Illinois at Urbana–Champaign, 603 East Daniel St., Champaign, IL 61820, USA.
Email: hkoehn@cyrus.psych.illinois.edu

may not be available in small- or medium-sized testing programs. In addition, the iterative MLE procedures are sensitive to the starting values, and do not guarantee optimal solutions despite often consuming considerable amounts of computer time. But perhaps most important, MLE procedures are vulnerable to the misspecification of the CDM supposedly underlying the data (recall that the true model is never known). Model misspecification causes examinees to be assigned to proficiency classes to which they do not belong.

In response to these difficulties, a number of researchers (Ayers, Nugent, & Dean, 2008; Chiu, 2008; Chiu & Douglas, 2013; Chiu, Douglas, & Li, 2009; Park & Lee, 2011; Willse, Henson, & Templin, 2007) have explored the potential of nonparametric classification techniques as heuristic or approximate methods for assigning examinees to proficiency classes. (A heuristic uses clever computational shortcut strategies to obtain a solution that is very close, if not identical, to the optimal solution.) Software for implementing these techniques can be developed from efficient cluster analysis programs that are readily available in the major statistical packages.

The Asymptotic Classification Theory of Cognitive Diagnosis (ACTCD; Chiu et al., 2009) provides the theoretical foundation for using hierarchical agglomerative cluster analysis (HACA) as a heuristic for assigning examinees to proficiency classes for educational data conforming to the Deterministic Input Noisy Output ''AND'' gate (DINA) model (Junker & Sijtsma, 2001; Macready & Dayton, 1977).

This article extends the ACTCD to item responses conforming to the Deterministic Input Noisy Output ''OR'' gate (DINO) model (Templin & Henson, 2006). It also demonstrates that an addition to the statistical framework of the ACTCD, *augmented attribute sum-score profile* (Chiu, 2008), which was originally developed to accommodate item responses conforming to the Reduced Reparameterized Unified Model (Reduced RUM; Hartz, 2002; Hartz & Roussos, 2008) or the General Diagnostic Model (GDM; von Davier, 2005, 2008), is valid also for the DINA model and the DINO model and increases the accuracy of HACA in classifying examinees when the item responses conform to either of these two models. A brief review of relevant definitions and key technical concepts precedes the theoretical development. The results of two simulation studies using item responses generated from the DINO model and the DINA model and an application of the studied methods to a real-world data set are then reported to provide empirical support for these extensions of the ACTCD. Finally, the ''Discussion'' section summarizes some of their practical implications.

Technical Background

Models for Cognitive Diagnosis

Let Y_{ij} be the observed response of examinee i, $i = 1, \ldots, N$, to binary item j, $j = 1, \ldots, J$. Consider N examinees who belong to M distinct latent classes of intellectual proficiency. CDMs constrain the relation between the observed item response and the latent variable, so that the mastery of cognitive attributes characteristic for distinct latent proficiency classes determines the observed response (correct or incorrect) to a test item. Suppose that K latent binary attributes constitute a certain ability domain; there are then 2^K distinct attribute profiles composed of these K attributes representing $M = 2^K$ distinct latent proficiency classes. Note that an attribute profile for a proficiency class can consist of all zeroes, because it is possible for an examinee not to have mastered any attributes at all. Let the K-dimensional vector, $\boldsymbol{\alpha}_m = (\alpha_1, \ldots, \alpha_K)^T$, be the binary attribute profile of proficiency class $\mathcal{C}_m$, where the kth entry indicates whether the respective attribute has been mastered. (Throughout the text, the superscript T denotes the transpose of vectors; the ''prime notation'' is reserved for distinguishing between vectors or their scalar entries.)

Consider a test of J items assessing ability in the domain. Each individual item j is associated with a binary attribute profile that specifies or constrains the particular skills required for answering it correctly. Given K attributes, there are at most $2^K - 1$ distinct item-attribute profiles; item-attribute profiles that consist entirely of zeroes are inadmissible, because they correspond to items whose answers require no skills at all. The entire set of constraints specifying the associations between J items and K attributes constitutes the Q-matrix, $\mathbf{Q} = \{q_{jk}\}_{(J \times K)}$, where $q_{jk} = 1$ if a correct answer to the jth item requires mastery of the kth attribute, and 0 otherwise; thus, the rows of $\mathbf{Q}$, $\mathbf{q}_j$, are the item-attribute profiles (K. K. Tatsuoka, 1985).

Nonparametric Classification Using HACA Adapted for Cognitive Diagnosis

For input into HACA, Chiu and collaborators (Chiu, 2008; Chiu et al., 2009) aggregated each examinee's set of item responses, $\boldsymbol{Y}_i$, into a K-dimensional profile of attribute sum-scores, $\boldsymbol{W}_i$, defined as $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{iK})' = \boldsymbol{Y}_i\mathbf{Q}$, where $W_{ik} = \sum_{j=1}^{J} Y_{ij}q_{jk}$. Because each cell entry of $\mathbf{Q} = \{q_{jk}\}$ represents the association between an item and an attribute, each element of $\boldsymbol{W}_i$ is the sum of the correct answers of examinee i to all items requiring mastery of the kth attribute. Items that require mastery of more than one attribute for their solution contribute to multiple elements of $\boldsymbol{W}_i$. Across examinees, the attribute sum-score profiles, $\boldsymbol{W}_i$, form the rows of a rectangular $N \times K$ matrix, $\boldsymbol{W}$. (Item response matrices are symbolized by non-italicized boldface capital letters, for example, $\boldsymbol{W}$, but their rows by subscripted italicized boldface capital letters, for example, $\boldsymbol{W}_i$, to emphasize that these are vectors of random variables. For brevity, the examinee index is omitted when the context permits.) These attribute sum-score profiles served as input to HACA, which assigned examinees to groups, or *clusters*, that served as proxies for the proficiency classes in cognitive diagnosis.

Many techniques exist for the nonparametric classification of a set of objects (such as the rows of a matrix). The principal objective shared by all of these techniques is to identify maximally homogeneous clusters that are maximally separated. (The classic reference is Hartigan, 1975.) To adapt HACA for cognitive diagnosis requires first transforming the $N \times K$ matrix of the examinees' attribute sum-score profiles into an $N \times N$ symmetric matrix of inter-examinee Euclidean distances. Popular HACA algorithms include single-, complete-, and average-link clustering (Johnson, 1967) and Ward's (1963) minimum-variance method. The link algorithms all sequentially merge or agglomerate examinees (or groups of examinees) closest to each other at each step into an inverted tree-shaped hierarchy of nested clusters that represents the relationships between examinees. The inter-examinee distances are updated after each merger to reflect the latest status of examinee/cluster cohesion as input for the next agglomeration step; the specific method of updating these distances distinguishes the various link algorithms. Ward's (1963) method differs from the link algorithms in that it updates the within-cluster sum of squared errors rather than the inter-examinee distances.

The ACTCD Developed for the DINA Model

The development of the ACTCD was inspired by a search for legitimate nonparametric classification heuristics for item responses conforming to the DINA model (Junker & Sijtsma, 2001; Macready & Dayton, 1977), a conjunctive non-compensatory CDM. The conjunction parameter η_{ij} indicates whether examinee i has mastered all the attributes needed to answer item j correctly; η_{ij} is defined as $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$. The item response function (IRF) of the DINA model for item j and examinee i is $P(Y_{ij} = 1|\boldsymbol{\alpha}) = (1 - s_j)^{\eta_{ij}} g_j^{(1-\eta_{ij})}$, where $s_j = P(Y_{ij} = 0 \mid \eta_{ij} = 1)$ and $g_j = P(Y_{ij} = 1 \mid \eta_{ij} = 0)$ are item parameters formalizing the probabilities of "slipping" (failing to answer item j correctly despite having the skills required to do so) and ''guessing'' (answering item j correctly despite lacking the skills required to do so), respectively.

The ACTCD (Chiu, 2008; Chiu et al., 2009) consists of three lemmas, each of which specifies a condition necessary for a consistency theorem to hold. (For formal proofs, consult Chiu et al., 2009.) Like MLE procedures, heuristic classification techniques using attribute sum-score profiles as input require that the Q-matrix be *complete*. A Q-matrix is said to be complete if it allows identification of all possible attribute profiles. This definition translates into the formal expression tailored to the DINA model, $\boldsymbol{\eta}(\boldsymbol{\alpha}) = \boldsymbol{\eta}(\boldsymbol{\alpha}^*) \Rightarrow \boldsymbol{\alpha} = \boldsymbol{\alpha}^*$, as it was used by Chiu et al. (2009). However, here, a more general definition of completeness of the Q-matrix is required that also covers the DINO model. Formally, this general definition is based on the conditional expectation of an examinee's item response profile, $\boldsymbol{Y}$, given attribute profile α, denoted by $\boldsymbol{S}(\boldsymbol{\alpha}) = E(\boldsymbol{Y} \mid \boldsymbol{\alpha})$. For the DINA model, the jth entry of $\boldsymbol{S}(\boldsymbol{\alpha})$ is defined as $E(Y_j|\boldsymbol{\alpha}) = (1 - s_j)^{\eta_j} g_j^{(1-\eta_j)}$.

Definition: $\boldsymbol{S}(\boldsymbol{\alpha}) = \boldsymbol{S}(\boldsymbol{\alpha}^*) \Rightarrow \boldsymbol{\alpha} = \boldsymbol{\alpha}^*$.

Lemma 1 of the ACTCD states that, for item responses conforming to the DINA model, **Q** is complete if and only if each attribute is represented by at least one single-attribute item (Chiu et al., 2009). The more general definition of **Q**-completeness means that the proof of Lemma 1 for the DINA model requires a modification, which is presented next.

Let $\mathbf{e}_k$ denote a unit vector with the kth element equal to 1 and all other elements equal to 0.

Lemma 1 (DINA): Assume that $0 < g_j < 1 - s_j < 1$ for all items. A $J \times K$ matrix **Q** is complete if and only if it includes the K vectors, $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_K$, among its J rows.

Proof (Modified):
($\Rightarrow$) *Assume* that some $\mathbf{e}_k$ is missing from **Q**. Let attribute profiles $\boldsymbol{\alpha} = (0, 0, \ldots, 0)^T$ and $\boldsymbol{\alpha}^* = \mathbf{e}_k$, with the kth element equal to 1 and all other elements equal to 0. The definition of $\boldsymbol{\eta}$ implies that $\eta_j(\boldsymbol{\alpha}) = \eta_j(\boldsymbol{\alpha}^*) = 0$ for all j. Hence, $S_j(\boldsymbol{\alpha}) = S_j(\boldsymbol{\alpha}^*) = g_j$ for all j. Therefore, $\boldsymbol{S}(\boldsymbol{\alpha}) = \boldsymbol{S}(\boldsymbol{\alpha}^*)$, so **Q** is not complete.

($\Leftarrow$) *Assume* that K of the J rows of $\boldsymbol{Q}$ consist of $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_K$. Reorder the rows of **Q** by moving these K rows to the first K row positions. For any $\boldsymbol{\alpha}$, the first K entries of $\boldsymbol{\eta}(\boldsymbol{\alpha})$, written as $\boldsymbol{\eta}_{1:K}(\boldsymbol{\alpha})$, are then identical to $\boldsymbol{\alpha}$. Of course, if $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$, then $\boldsymbol{\eta}_{1:K}(\boldsymbol{\alpha}) \neq \boldsymbol{\eta}_{1:K}(\boldsymbol{\alpha}^*)$. Hence, if $\boldsymbol{\eta}_{1:K}(\boldsymbol{\alpha}) \neq \boldsymbol{\eta}_{1:K}(\boldsymbol{\alpha}^*)$, then $\boldsymbol{\eta}(\boldsymbol{\alpha}) \neq \boldsymbol{\eta}(\boldsymbol{\alpha}^*)$ holds for all items in **Q**, regardless of whether $\boldsymbol{\eta}_{(K+1):J}(\boldsymbol{\alpha})$ and $\boldsymbol{\eta}_{(K+1):J}(\boldsymbol{\alpha}^*)$ are identical. Therefore, if $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$, then $\boldsymbol{\eta}(\boldsymbol{\alpha}) \neq \boldsymbol{\eta}(\boldsymbol{\alpha}^*)$, which implies that $\boldsymbol{S}(\boldsymbol{\alpha}) \neq \boldsymbol{S}(\boldsymbol{\alpha}^*)$; thus, **Q** must be complete.

Let $\boldsymbol{T}(\boldsymbol{\alpha}) = E(\boldsymbol{W}|\boldsymbol{\alpha})$ be the conditional expectation of the attribute sum-score profile, $\boldsymbol{W}$, given attribute profile $\boldsymbol{\alpha}$, where the kth element of $\boldsymbol{T}(\boldsymbol{\alpha})$ is defined as $T_k(\boldsymbol{\alpha}) = E(W_k|\boldsymbol{\alpha}) = \sum_{j=1}^{J} E(Y_j|\boldsymbol{\alpha}) q_{jk}$. Loosely speaking, $\boldsymbol{T}(\boldsymbol{\alpha})$ can be regarded as the center of the proficiency class characterized by $\boldsymbol{\alpha}$. Lemma 2 of the ACTCD states that, if the item responses conform to the DINA model and the Q-matrix is complete, then for two attribute profiles, $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$, $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^* \Rightarrow \boldsymbol{T}(\boldsymbol{\alpha}) \neq \boldsymbol{T}(\boldsymbol{\alpha}^*)$ always holds (Chiu et al., 2009).

Thus, Lemma 2 justifies using $\boldsymbol{W}$ as a statistic for α because the centers of the different proficiency classes are guaranteed to be distinct. That is, the proficiency classes are well-separated, which is a requirement for proving the consistency of $\boldsymbol{W}$.

If a finite mixture model with M latent classes underlies the item responses, then Lemma 3 of the ACTCD establishes that complete-link HACA accurately assigns examinees to their true proficiency classes, provided that the resulting classification hierarchy is ''cut'' at M clusters (Chiu et al., 2009).

Building on these three lemmas, the consistency theorem of classification states that, for item responses conforming to the DINA model, the probability that HACA assigns examinees correctly to their true proficiency classes based on $\boldsymbol{W}$ approaches 1 as the length of a test (i.e., the number of items, J) increases (Chiu et al., 2009).

Extension of the ACTCD to the DINO Model

The DINO model (Templin & Henson, 2006) is a disjunctive CDM. Define the disjunction parameter $\omega_{ij} = 1 - \prod_{k=1}^{K} (1 - \alpha_{ik})^{q_{jk}}$ that indicates whether at least one of the attributes associated with item j has been mastered. (Like η_{ij} in the DINA model, ω_{ij} represents the ideal item response when neither slipping nor guessing occurs.) The IRF of the DINO model is $P(Y_{ij} = 1|\boldsymbol{\alpha}) = (1 - s_j)^{\omega_{ij}} g_j^{(1-\omega_{ij})}$. The expected item response is $S_j(\boldsymbol{\alpha}) = E(Y_j|\boldsymbol{\alpha}) = (1 - s_j)^{\omega_j} g_j^{(1-\omega_j)}$.

Lemmas 1 and 2 of the ACTCD for the DINA model specify regularity conditions that must be satisfied by any CDM for the consistency theorem of classification to hold. Hence, extension of the ACTCD to the DINO model requires proving that it is covered by these two lemmas. These proofs rely on the ''duality'' of the DINO model and the DINA model (Liu, Xu, & Ying, 2011, 2012).

The Duality of the DINA Model and the DINO Model

As Y. Liu et al. (2011) discovered and proved, the DINA model and the DINO model are technically identical under certain transformations of (a) the examinees' attribute profiles, (b) their observed item scores, and (c) the model parameters. This means that one model can be expressed in terms of the other and both models can be fitted by the same software. (As an aside, note that the characterization of the special relation between the DINA model and the DINO model as ''dual'' deviates from the well-defined meaning of this term in operations research; for details, consult Papadimitriou & Steiglitz, 1998.) The proof of the duality of the DINA model and the DINO model presented here is a modification of the original proof by Y. Liu et al. (2011) tailored to the ACTCD.

Transformation of Examinees' Attribute Profiles

Consider the attribute profile $\boldsymbol{\alpha}' = \mathbf{1} - \boldsymbol{\alpha}$. Then,

$$\eta_j = \prod_{k=1}^{K} \alpha_k^{q_{jk}} = \prod_{k=1}^{K} (1 - \alpha_k')^{q_{jk}} = 1 - \left(1 - \prod_{k=1}^{K} (1 - \alpha_k')^{q_{jk}}\right) = 1 - \omega_j',$$

where $\omega'_j = 1 - \prod_{k=1}^{K} (1 - \alpha_k')^{q_{jk}}$ is the disjunctive parameter associated with the attribute profile $\boldsymbol{\alpha}'$. Thus, the conditional expectation of Y_j, given α', for the DINO model is,

$$E_{\text{DINO}}(Y_j|\boldsymbol{\alpha}') = (1 - ss_j)^{\omega'_j} gg_j^{1-\omega'_j}$$

(for emphasis, the slipping and guessing parameters of the DINO model are henceforth denoted by ss_j and gg_j, respectively).

Transformation of Examinees' Observed Item Response Scores

Let $Y_j' = 1 - Y_j$; then, the conditional expectation of Y'_j, given α', for the DINO model is,

$$\begin{aligned} E_{\text{DINO}}\left(Y'_j|\boldsymbol{\alpha}'\right) &= 1 - E_{\text{DINO}}\left(Y_j|\boldsymbol{\alpha}'\right) \\ &= 1 - \left(1 - ss_j\right)^{\omega'_j} gg_j^{1-\omega'_j} \\ &= ss_j^{\omega'_j}\left(1 - gg_j\right)^{1-\omega'_j}, \end{aligned} \tag{1}$$

because

$$1 - \left(1 - ss_j\right)^{\omega'_j} gg_j^{1-\omega'_j} = \begin{cases} 1 - gg_j & \text{if} \quad \omega'_j = 0 \\ ss_j & \text{if} \quad \omega'_j = 1 \end{cases}.$$

Transformation of the Model Parameters

In replacing ω'_j by $1 - \eta_j$ and $1 - \omega'_j$ by η_j, Equation 1 can be rewritten as,

$$E_{\text{DINO}}\left(Y'_j|\boldsymbol{\alpha}'\right) = ss_j^{1-\eta_j}\left(1 - gg_j\right)^{\eta_j} = g_j^{1-\eta_j}\left(1 - s_j\right)^{\eta_j} = E_{\text{DINA}}\left(Y_j|\boldsymbol{\alpha}\right), \tag{2}$$

by setting $g_j = ss_j$ and $s_j = gg_j$. This completes the proof.

Results Based on Equation 2

First, the observed item responses, Y'_j, can be fitted by the DINO model via software for the DINA model using the transformed item responses, $Y_j = 1 - Y'_j$, as input. Second, the estimates of examinees' attribute profiles in terms of the DINO model, $\hat{\boldsymbol{\alpha}}'$, are then computed from the DINA attribute profile estimates, $\hat{\boldsymbol{\alpha}}$, through the transformation $\hat{\boldsymbol{\alpha}}' = \mathbf{1} - \hat{\boldsymbol{\alpha}}$. Third, the estimates of the slipping and guessing parameters obtained by the DINA software, $\hat{s}_j$ and $\hat{g}_j$, are identical to the estimates of the guessing and slipping parameters for the DINO model—that is, $\hat{s}_j = \widehat{gg}_j$ and $\hat{g}_j = \widehat{ss}_j$. Obviously, these transformations ''work both ways'': Y'_j can be fitted by the DINA model via software for the DINO model using the transformed item responses, $Y_j = 1 - Y'_j$, as input, and so on. From a practical point of view, this is merely a curiosity: No one would consider fitting the DINO model by using the DINA software with the transformed data while the implementation of the EM algorithm for fitting the DINO model is readily available in R through the package CDM (Robitzsch, Kiefer, George, & Uenlue, 2014). The importance of these results is rather theoretical: If the two models can be shown to be identical under certain transformations, then they should share the same theoretical properties. Thus, any theoretical result discovered for one model must automatically hold for the other model too. As indicated earlier, the subsequent proofs of Lemmas 1 and 2 for the DINO model use this duality.

Lemma 1 (DINO): Assume that $0 < gg_j < 1 - ss_j < 1$ for all items. A $J \times K$ matrix $\mathbf{Q}$ is complete if and only if it includes the K vectors, $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_K$, among its J rows.

Proof:
($\Rightarrow$) The proof of Lemma 1 for the DINA model demonstrated that if $\mathbf{e}_k$ is missing from $\mathbf{Q}$, then $\boldsymbol{S}_{\text{DINA}}(\boldsymbol{\alpha}) = \boldsymbol{S}_{\text{DINA}}(\boldsymbol{\alpha}^*)$, where $\boldsymbol{\alpha} = (0, 0, \ldots, 0)^T$ and $\boldsymbol{\alpha}^* = \mathbf{e}_k$. Define $\boldsymbol{\alpha}' = \mathbf{1} - \boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^{*\prime} = \mathbf{1} - \boldsymbol{\alpha}^*$; because $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$, $\boldsymbol{\alpha}' \neq \boldsymbol{\alpha}^{*\prime}$. From the duality of the DINA model and the DINO model, then follows:

$$\mathbf{S}_{\text{DINO}}(\boldsymbol{\alpha}') = E_{\text{DINO}}(\mathbf{Y}'|\boldsymbol{\alpha}') = E_{\text{DINA}}(\mathbf{Y}|\boldsymbol{\alpha}) = E_{\text{DINA}}(\mathbf{Y}|\boldsymbol{\alpha}^*) = E_{\text{DINO}}\left(\mathbf{Y}'|\boldsymbol{\alpha}^{*'}\right) = \mathbf{S}_{\text{DINO}}\left(\boldsymbol{\alpha}^{*'}\right)$$

(recall that $\boldsymbol{S}_{\text{DINA}}(\boldsymbol{\alpha}) = E_{\text{DINA}}(\boldsymbol{Y}|\boldsymbol{\alpha})$). Therefore, $\boldsymbol{Q}$ is not complete.

($\Leftarrow$) Assume that $\mathbf{Q}$ contains among its J rows the K single-attribute item profiles $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_K$. The proof of Lemma 1 for the DINA model showed that under this condition, the Q-matrix is complete because $\boldsymbol{S}_{\text{DINA}}(\boldsymbol{\alpha}) \neq \boldsymbol{S}_{\text{DINA}}(\boldsymbol{\alpha}^*)$ if $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$. Recall that $\boldsymbol{\alpha}' \neq \boldsymbol{\alpha}^{*\prime}$ because $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$. From the duality of the DINA model and the DINO model, follows then $\boldsymbol{S}_{\text{DINO}}(\boldsymbol{\alpha}') \neq \boldsymbol{S}_{\text{DINO}}(\boldsymbol{\alpha}^{*\prime})$ because $\boldsymbol{S}_{\text{DINA}}(\boldsymbol{\alpha}) \neq \boldsymbol{S}_{\text{DINA}}(\boldsymbol{\alpha}^*)$. Therefore, $\boldsymbol{S}_{\text{DINO}}(\boldsymbol{\alpha}') \neq \boldsymbol{S}_{\text{DINO}}(\boldsymbol{\alpha}^{*\prime})$ if $\boldsymbol{\alpha}' \neq \boldsymbol{\alpha}^{*\prime}$. So, $\boldsymbol{Q}$ must be complete.

Lemma 2 (DINO): Assume that $\mathbf{Q}$ is complete and that $0 < gg_j < 1 - ss_j < 1$ for all items. If $\boldsymbol{\alpha}' \neq \boldsymbol{\alpha}^{*\prime}$, then $\boldsymbol{T}_{\text{DINO}}(\boldsymbol{\alpha}') \neq \boldsymbol{T}_{\text{DINO}}(\boldsymbol{\alpha}^{*\prime})$.

Proof:
Lemma 2 for the DINA model states $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^* \Rightarrow \boldsymbol{T}_{\text{DINA}}(\boldsymbol{\alpha}) \neq \boldsymbol{T}_{\text{DINA}}(\boldsymbol{\alpha}^*)$ provided $\mathbf{Q}$ is complete and $0 < g_j < 1 - s_j < 1$ for all items. Let $ss_j = g_j$ and $gg_j = s_j$. Due to the duality of the DINA model and the DINO model, the assumption $0 < gg_j < 1 - ss_j < 1$ is equivalent to $0 < s_j < 1 - g_j < 1$, which in turn is equivalent to $0 < g_j < 1 - s_j < 1$. Assume that $\boldsymbol{\alpha}' \neq \boldsymbol{\alpha}^{*\prime}$. Let $\boldsymbol{\alpha} = \mathbf{1} - \boldsymbol{\alpha}'$ and $\boldsymbol{\alpha}^* = \mathbf{1} - \boldsymbol{\alpha}^{*\prime}$. Therefore, $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$, which implies $\boldsymbol{T}_{\text{DINA}}(\boldsymbol{\alpha}) \neq \boldsymbol{T}_{\text{DINA}}(\boldsymbol{\alpha}^*)$ (i.e., Lemma 2 (DINA)). Now,

$$\mathbf{T}_{\text{DINO}}(\boldsymbol{\alpha}') = E_{\text{DINO}}(\mathbf{Y}'\mathbf{Q}|\boldsymbol{\alpha}') = E_{\text{DINO}}(\mathbf{Y}'|\boldsymbol{\alpha}')\mathbf{Q} = E_{\text{DINA}}(\mathbf{Y}|\boldsymbol{\alpha})\mathbf{Q} = \mathbf{T}_{\text{DINA}}(\boldsymbol{\alpha}),$$

and $\boldsymbol{T}_{\text{DINO}}(\boldsymbol{\alpha}^{*\prime}) = \boldsymbol{T}_{\text{DINA}}(\boldsymbol{\alpha}^*)$. Because $\boldsymbol{T}_{\text{DINA}}(\boldsymbol{\alpha}) \neq \boldsymbol{T}_{\text{DINA}}(\boldsymbol{\alpha}^*)$, $\boldsymbol{T}_{\text{DINO}}(\boldsymbol{\alpha}') \neq \boldsymbol{T}_{\text{DINO}}(\boldsymbol{\alpha}^{*\prime})$. Therefore, $\boldsymbol{\alpha}' \neq \boldsymbol{\alpha}^{*\prime} \Rightarrow \boldsymbol{T}_{\text{DINO}}(\boldsymbol{\alpha}') \neq \boldsymbol{T}_{\text{DINO}}(\boldsymbol{\alpha}^{*\prime})$ must be true and the proof is complete.

In summary, Lemma 1 (DINO) and Lemma 2 (DINO) can be used with Lemma 3 to prove that the consistency theorem of classification holds when the item responses conform to the DINO model. The proof is omitted because it is identical to that given by Chiu et al. (2009).

The Augmented Attribute Sum-Score Profile for the DINA Model and the DINO Model

The augmented attribute sum-score profile, $\boldsymbol{W}_{\text{aug}}$, was developed by Chiu (2008) as a remedy for the inability of the attribute sum-score profile, $\boldsymbol{W}$, to guarantee well-separated proficiency-class centers when the item responses conform to the Reduced RUM or the GDM. For these models, Chiu (2008) demonstrated that distinct proficiency classes can have identical conditional expectations of $\boldsymbol{W}$, which invalidates Lemma 2 (i.e., $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^* \nRightarrow \boldsymbol{T}(\boldsymbol{\alpha}) \neq \boldsymbol{T}(\boldsymbol{\alpha}^*)$); thus, $\boldsymbol{W}$ is no longer consistent. Chiu (2008) proved that replacing $\boldsymbol{W}$ by $\boldsymbol{W}_{\text{aug}}$ as a statistic for $\boldsymbol{\alpha}$ restores the guarantee of well-separated proficiency-class centers and that $\boldsymbol{W}_{\text{aug}}$ is consistent. Before computing $\boldsymbol{W}_{\text{aug}}$, the Q-matrix is augmented by a matrix of the same dimensionality in which the $\mathbf{q}_j$ corresponding to the single-attribute items are replicated and all other rows are zero vectors. That is, $\mathbf{Q}_{\text{aug}} = (\mathbf{Q}, \mathbf{Q}_{\mathbf{e}})$, where $\mathbf{Q}_{\mathbf{e}}$ is constructed by retaining the rows $\mathbf{q}_j = \mathbf{e}_k$ in $\mathbf{Q}$ and replacing the remaining $\mathbf{q}_j$ with $(0, 0, \ldots, 0)$. The definitions of the augmented attribute sum-score profile and its expectation can be derived from $\mathbf{Q}_{\text{aug}}$:

$$\begin{array}{ccccccc} \mathbf{W}_{aug} & = & \mathbf{Y}\mathbf{Q}_{aug} & = & \mathbf{Y}(\mathbf{Q}, \mathbf{Q}_e) & = & (\mathbf{W}, \mathbf{W}_e) \\ \mathbf{T}_{aug}(\boldsymbol{\alpha}) & = & E\left(\mathbf{W}_{aug}|\boldsymbol{\alpha}\right) & = & E((\mathbf{W}, \mathbf{W}_e)|\boldsymbol{\alpha}) & = & (\mathbf{T}(\boldsymbol{\alpha}), \mathbf{T}_e(\boldsymbol{\alpha})) \end{array}.$$

Note that $\boldsymbol{W}_{\mathbf{e}} = \boldsymbol{Y}\mathbf{Q}_{\mathbf{e}}$ and that $\boldsymbol{W}_{\text{aug}}$ is just the composite of $\boldsymbol{W}$ and $\boldsymbol{W}_{\mathbf{e}}$: $\boldsymbol{W}_{\text{aug}} = (\boldsymbol{W} \mid \boldsymbol{W}_{\mathbf{e}})$. Augmenting $\boldsymbol{W}$ by $\boldsymbol{W}_{\mathbf{e}}$ provides extra input to the clustering algorithm on test items that are particularly effective in differentiating proficiency classes (i.e., single-attribute items). Thus, the augmented attribute sum-score profile, $\boldsymbol{W}_{\text{aug}}$, enhances well-separated proficiency classes, which in turn ''boosts'' the assignment of examinees to proficiency classes.

Does the enhancement of well-separated proficiency classes by $\boldsymbol{W}_{\text{aug}}$ also benefit heuristic classification when applied to item responses conforming to the DINA model or the DINO model (for which $\boldsymbol{W}$ already guarantees well-separated proficiency-class centers)? This question must be answered empirically (by simulations, for example). First, however, it must be proven that Lemma 2 for the DINA model and the DINO model is still valid when $\boldsymbol{W}$ is replaced by $\boldsymbol{W}_{\text{aug}}$ as a statistic for α.

Lemma 2 (DINA: $\boldsymbol{W}_{\mathbf{aug}}$): Assume that $0 < g_j < 1 - s_j < 1$ for all items j, and that $\mathbf{Q}$ is complete. If $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$, then $\boldsymbol{T}_{\text{aug}}(\boldsymbol{\alpha}) \neq \boldsymbol{T}_{\text{aug}}(\boldsymbol{\alpha}^*)$.

Proof:
Lemma 2 in Chiu et al. (2009) states that, if the item responses conform to the DINA model and $\mathbf{Q}$ is complete, then $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^* \Rightarrow \boldsymbol{T}(\boldsymbol{\alpha}) \neq \boldsymbol{T}(\boldsymbol{\alpha}^*)$. Recall the definition $\boldsymbol{T}_{\text{aug}}(\boldsymbol{\alpha}) = (\boldsymbol{T}(\boldsymbol{\alpha}), \boldsymbol{T}_{\mathbf{e}}(\boldsymbol{\alpha}))$. Hence, $\boldsymbol{T}_{\text{aug}}(\boldsymbol{\alpha}) \neq \boldsymbol{T}_{\text{aug}}(\boldsymbol{\alpha}^*)$ regardless of whether $\boldsymbol{T}_{\mathbf{e}}(\boldsymbol{\alpha})$ equals $\boldsymbol{T}_{\mathbf{e}}(\boldsymbol{\alpha}^*)$.

Lemma 2 (DINO: $\boldsymbol{W}_{\text{aug}}$) is identical to Lemma 2 (DINA: $\boldsymbol{W}_{\text{aug}}$). The proof of Lemma 2 for $\boldsymbol{W}_{\text{aug}}$ for the DINO model is omitted because it follows the same logic as the proof for the DINA model (except that it relies on Lemma 1 and 2 for the DINO model, proven earlier).

In summary, the augmented attribute sum-score profile, $\boldsymbol{W}_{\text{aug}}$, is a legitimate statistic for $\boldsymbol{\alpha}$. The simulation studies described next compare the performance of $\boldsymbol{W}_{\text{aug}}$ to that of the attribute sum-score profile, $\boldsymbol{W}$, when serving as input to three clustering methods.

Simulation Studies

Two simulation studies were conducted to provide empirical support for the theoretical developments of the ACTCD proposed for the DINO model and the DINA model by demonstrating that (a) the consistency theorem of classification is valid for item responses conforming to the DINO model as well as for those conforming to the DINA model—that is, the accuracy of HACA in assigning examinees to their true proficiency classes increases with the number of test items; and that (b) use of the augmented attribute sum-score profile, $\boldsymbol{W}_{\text{aug}}$, rather than the attribute sum-score profile, $\boldsymbol{W}$, as input to HACA results in a more accurate proficiency-class assignment of examinees when the item responses conform to either of the two models. In Study 1, it is assumed that the true CDM underlying the data—the DINA model or the DINO model—is known. Hence, the performance of nonparametric heuristic classification techniques, as exemplified by HACA, is compared with the performance of the exact EM-based MLE method implemented in the R package CDM (Robitzsch et al., 2014). Study 2, in contrast, compares the performance of HACA with the exact EM-based MLE method when the true model underlying the data has been misspecified. Thus, data truly conforming to the DINA model are fitted with the DINO-EM algorithm, and data truly conforming to the DINO model are fitted with the DINA-EM algorithm.

Study 1

Experimental Design and Simulation of Item Responses

Examinee item responses conforming to the DINO model and to the DINA model were simulated according to the method described by Chiu et al. (2009). The experimental design included five variables: (a) the number of examinees, $N = 100, 500$; (b) the number of attributes, $K = 3, 4$; (c) the number of items, $J = 20, 40, 80$; (d) the distribution (discrete uniform, multivariate normal) underlying the attribute profiles; and (e) the distribution of the slipping and guessing item parameters, s_j and g_j (continuous uniform $\mathcal{U}(0, 0.15)$, continuous uniform $\mathcal{U}(0, 0.30)$). For the levels of variable K, $2^3 - 1 = 7$ and $2^4 - 1 = 15$ distinct binary item-attribute profiles (omitting those consisting of all zeroes) were generated to form the template Q-matrices of tests containing $J = 20$ items (Chiu et al., 2009, Table 2, p. 650), shown in Table 1 of the online appendix. The Q-matrices for tests containing 40 or 80 items were created by stacking the 20-item template Q-matrices.

Each examinee's attribute profile was generated from either a discrete uniform distribution (i.e., the attribute profile for each proficiency class, $\boldsymbol{\alpha}_m$, had the same probability, $1 / M$, where $M = 2^K$) or from the more realistic and complex multivariate normal threshold model (see Chiu et al., 2009).

The simulated item responses, Y_{ij}, were sampled from a Bernoulli distribution with π_{ij} defined by the IRF of the respective model and the slipping and guessing item parameters, s_j and g_j, drawn from the continuous uniform distributions specified previously. The values of the slipping and guessing parameters reflect the degree of error perturbation of the item responses, with $s_j, g_j \sim \mathcal{U}(0, 0.15)$ and $s_j, g_j \sim \mathcal{U}(0, 0.30)$ representing moderate and large deviations from the ideal item responses, respectively.

Completely crossing the levels of all five variables yielded for the DINO model and the DINA model experimental designs each with $2 \times 2 \times 3 \times 2 \times 2 = 48$ cells. Twenty-five data sets were generated for each cell, for a total of 2,400 simulated data sets.

Each simulated examinee's set of item responses was aggregated into an attribute sum-score profile and an augmented attribute sum-score profile to serve as input to complete-link, average-link, and Ward's method HACA; single-link HACA was not used because its well-known susceptibility to ''chaining'' (observations being added one-by-one to a cluster, with the resulting clusters resembling long strings or chains) makes it an inappropriate choice in this context. All three clustering methods are implemented in the hclust routine in R. The cluster tree produced by each of the three clustering methods was ''cut'' at level $M = 2^K$, the true number of proficiency classes.

Evaluation of the Clustering Results

For simulated item responses, the true proficiency-class memberships of examinees are known and provide a standard for evaluating the accuracy of the results of the cluster analyses (i.e., to which extent examinees are assigned to their true proficiency classes). However, because the clustering methods do not label the clusters representing the proficiency classes with their respective attribute profiles (i.e., HACA cannot identify the different proficiency classes in terms of $\boldsymbol{\alpha}$), it is not possible to compute rates of correct classification. Therefore, an alternative measure to assess the accuracy of examinees' classification is needed. Specifically, a measure of agreement between the known true classification of examinees and their classification obtained by each clustering method was computed using the Hubert–Arabie Adjusted Rand Index (ARI; Hubert & Arabie, 1985; Steinley, 2004), which is bounded by 0 and 1 indicating perfect disagreement and perfect agreement, respectively. Thus, a perfect ARI of 1 also indicates perfect accuracy in classifying examinees: Their true proficiency-class memberships were completely recovered by the classification method.

Study 1: Results

Tables 2 and 3 (see the online appendix) present the results of the first simulation study for each of the two distributions (discrete uniform, multivariate normal) used to generate examinees' attribute profiles for item responses conforming to the DINO model; Tables 4 and 5 (see the online appendix) present analogous results for item responses conforming to the DINA model. Each table is organized, so that each of its rows reports the simulation results for a specific combination of the variables *N, J*, and *K*, and the distribution of the slipping and guessing parameters. The columns of each table contain the simulation results for each of the three clustering methods when either $\boldsymbol{W}$ or $\boldsymbol{W}_{\text{aug}}$ served as input to that clustering method. The last column in each table, labeled *DINO-EM* or *DINA-EM*, presents the results when the simulated item responses were analyzed based on the true CDM underlying the data using the corresponding MLE method. Each entry in the tables reports the average (mean) ARI for the 25 data sets generated for a cell in the experimental design.

The simulation results for the DINO model and the DINA model are very similar, so they are summarized together here.

a. In accordance with the ACTCD, a larger number of test items generally leads to a more accurate assignment of examinees to their true proficiency classes by all three clustering methods, regardless of whether $\boldsymbol{W}$ or $\boldsymbol{W}_{\text{aug}}$ serves as input.
b. The performance of all three clustering methods in assigning examinees to their true proficiency classes is better when $\boldsymbol{W}_{\text{aug}}$ rather than $\boldsymbol{W}$ serves as input; therefore, the presentation of the remaining results will focus on those for which $\boldsymbol{W}_{\text{aug}}$ serves as input.
c. The distribution underlying the attribute profiles appears to have a differential effect on the performance of the clustering methods: With one exception, the performance of Ward's method HACA is better when attribute profiles are underlain by a discrete uniform distribution rather than by a multivariate normal distribution, and with one exception, the performance of average-link HACA is better when attribute profiles are underlain by a multivariate normal distribution rather than by a discrete uniform distribution.
d. When either the number of attributes or the level of error perturbation increases, the accuracy of the assignment of examinees to their true proficiency classes deteriorates—this is most obvious under those conditions where $J = 20$.
e. The number of examinees seems to have no effect on the accuracy of the assignment of examinees to their true proficiency classes.
f. When the level of error perturbation is moderate, or the number of test items is large, the ARIs computed for Ward's method HACA and average-link HACA are fairly comparable with those obtained using the MLE method.
g. The results obtained using the MLE method appear much less sensitive to variations in the number of attributes, the number of test items, the distribution of attribute profiles, and the level of error perturbation. This result was expected, of course, because the MLE method for the DINO model or the DINA model should outperform nonparametric heuristic classification techniques when used to analyze item responses generated from the corresponding model.

Study 2

The design of Study 2 is identical to that of Study 1, except for one modification: Only attribute profiles with an underlying multivariate normal distribution were considered. Thus, completely crossing the levels of all five variables yielded for the DINO model and the DINA model experimental designs each with $2 \times 2 \times 3 \times 1 \times 2 = 24$ cells. Twenty-five data sets were generated for each cell, for a total of 1,200 simulated data sets. Each simulated examinee's set of item responses was aggregated into an attribute sum-score profile and an augmented attribute sum-score profile to serve as input to the same clustering methods, as were used in Study 1. The recovery of the true proficiency classes was assessed by the ARI. Recall that the main purpose of Study 2 was to compare the performance of HACA with the exact EM-based MLE method when the true model underlying the data is unknown and has been misspecified (i.e., data truly conforming to the DINA model are fitted with the DINO model and vice versa).

Study 2: Results

The results of Study 2 are summarized in Tables 6 and 7 (see the online appendix), which follow the exact same lay-out like the previous tables. However, now, the last column in each table, labeled *DINA-EM* and *DINO-EM*, respectively, presents the results when the simulated item responses were analyzed based on the misspecified CDM using the corresponding MLE method. Each entry in the tables reports the average (mean) ARI for the 25 data sets generated for a cell in the experimental design. The results regarding the performance of the different HACA methods match those from Study 1 and are, therefore, not further commented on. The most remarkable finding of Study 2 concerns the performance of the HACA methods in comparison with the EM-based MLE methods when the CDM has been misspecified: For all 48 cells, the heuristic HACA methods outperformed the exact EM-based MLE methods in the recovery of the true proficiency classes as measured in terms of the average ARI scores.

Practical Application

The real-world fraction–subtraction data set (K. K. Tatsuoka, 1984) is one of the most thoroughly studied in cognitive diagnosis research (e.g., de la Torre, 2008, 2009; de la Torre & Douglas, 2008; DeCarlo, 2011; Mislevy, 1996; C. Tatsuoka, 2002). A subset of this data set consisting of the responses of 536 middle school students to a collection of 15 test items (de la Torre, 2008) is used here to compare the ability of the three clustering methods to assign examinees to proficiency classes when $\boldsymbol{W}$ and $\boldsymbol{W}_{\text{aug}}$ serve as input. The 15 fraction–subtraction test items and the five attributes required to answer them correctly are shown in the Q-matrix in Table 8 (see the online appendix). Using this Q-matrix, de la Torre (2008) found that the DINA model fit this subset of the fraction–subtraction data well (e.g., with one exception, all item-parameter estimates were less than 0.30). Therefore, for illustrative purposes, the DINA model is considered here to be the ''true'' model underlying the item responses, and the proficiency classes to which DINA-EM assigns the examinees are considered to be the ''true'' proficiency classes.

Twenty-four proficiency classes were identified using DINA-EM; these 24 classes are used here as the standard of comparison for evaluating the proficiency-class assignments produced by the three clustering methods. The results are reported in Table 9 of the online appendix. The ARI for complete-link HACA when $\boldsymbol{W}_{\text{aug}}$ serves as input exceeds that when $\boldsymbol{W}$ serves as input. (Recall that complete-link HACA is the clustering method for which the consistency theorem of classification was originally proven.) The two other clustering methods do not reap the same benefit from $\boldsymbol{W}_{\text{aug}}$. Presumably, the difference between the ARIs of $\boldsymbol{W}$ and $\boldsymbol{W}_{\text{aug}}$ for the fraction–subtraction data is not very large because the Q-matrix contains only two single-attribute items (out of 15 in total), which—as an additional complication—both ''target'' the same attribute. Finally, it should be kept in mind that the ACTCD is an asymptotic theory; and a test consisting of 15 items might simply be too short.

Still, the relatively low ARI scores obtained for the HACA clustering methods when used with the fraction–subtraction data call for some additional explanations. First, note that with five attributes, there are $2^5 = 32$ possible proficiency classes. However, the Q-matrix of the fraction–subtraction data is incomplete. Thus, not all of the 32 different proficiency classes are identifiable. (Recall that completeness of the Q-matrix is a universal requirement of cognitive diagnosis. If the Q-matrix is not complete, then the identification of all $2^K = M$ different proficiency classes is impossible regardless of whether MLE methods or nonparametric classification techniques are used.) It can be shown that the Q-matrix of the fraction–subtraction data allows only the identification of 10 out of the 32 proficiency classes. But the benchmark for computing the ARI scores for evaluating the performance of $\boldsymbol{W}$ and $\boldsymbol{W}_{\text{aug}}$ as input statistics to the HACA methods was defined by the 24 proficiency classes identified by the DINA-EM algorithm. (The over-extraction of 24 proficiency classes does not speak against the DINA-EM algorithm, but is just another indication that the fraction–subtraction data are ill-conditioned. Previous studies of this data set were usually concerned with item-parameter estimation and model fit, but not with the assignment of examinees to proficiency classes, which might explain why this problematic aspect of the fraction–subtraction data has barely been discussed in the literature; see, for example, DeCarlo, 2011). It can be demonstrated that the over-extraction of proficiency classes by the DINA-EM algorithm negatively affects the ARI scores computed for HACA. The ARIs were computed for partitions with the technically identifiable number of 10 or fewer proficiency classes. (Note that the true number of proficiency classes in the data set could actually be less than 10.) The results are reported in Table 10 of online appendix. As predicted, the ARIs increase considerably if the number of proficiency classes is 10 or less.

Discussion

The ACTCD developed by Chiu et al. (2009) proves that for educational test item responses conforming to the DINA model, the probability that HACA assigns examinees to their true proficiency classes approaches 1 as the number of test items increases. This article proves that the ACTCD also covers item responses conforming to the DINO model.

HACA requires as input a statistic for an examinee's attribute profile, $\boldsymbol{\alpha}$, that is based on his or her observed item responses. Lemma 2 of the ACTCD states that the attribute sum-score profile, $\boldsymbol{W}$, is a legitimate input to HACA because $\boldsymbol{W}$ is a consistent statistic for $\boldsymbol{\alpha}$. This article also demonstrates that an extension to the statistical framework of the ACTCD originally developed for item responses conforming to the Reduced RUM or the GDM, the augmented attribute sum-score profile, $\boldsymbol{W}_{\text{aug}}$, (a) is also a valid statistic for $\boldsymbol{\alpha}$ for both the DINO model and the DINA model and (b) increases the accuracy of HACA in classifying examinees when test item responses conform to either of these two models.

The empirical performance of the two statistics for $\boldsymbol{\alpha}$ was compared in two simulation studies. Their results confirm that using augmented attribute sum-score profiles, $\boldsymbol{W}_{\text{aug}}$, rather than attribute sum-score profiles, $\boldsymbol{W}$, as input to HACA increases the accuracy of assigning examinees to proficiency classes. This is especially the case when the number of test items is small, because the fewer the test items, the more dramatic the advantage of augmented attribute sum-score profiles over the attribute sum-score profiles. (For the real-world fraction–subtraction data, the advantage of $\boldsymbol{W}_{\text{aug}}$ over $\boldsymbol{W}$ was less obvious; but this is most likely attributable to the peculiarities of this data set, as was discussed earlier.) In summary, if educational data conform to the DINO model and the DINA model, then use of the augmented attribute sum-score profiles (rather than the attribute sum-score profiles) as input to HACA can be regarded as a precautionary measure against the possibility of ill-separated proficiency classes, which can always occur in a finite sample of examinees.

What recommendations can be made to the educational practitioner? Recall that originally nonparametric classification techniques were proposed as a heuristic alternative to the exact MLE procedures for assigning examinees to proficiency classes because these methods often encountered difficulties in practice and suitable computer software was mostly unavailable. But recently, Robitzsch et al. (2014) have developed the package CDM that provides an implementation of the EM algorithm for fitting the DINO model and the DINA model in R. Thus, educational practitioners should use these exact MLE methods whenever possible because they typically outperform HACA in assigning examinees to proficiency classes under regular conditions—that is, when the sample sizes of examinees are sufficient and the CDM underlying the data has been correctly identified.

However, nonparametric classification techniques can be useful in situations where MLE methods fail or are difficult to implement—most notably, when the true CDM underlying the data is unknown. The second simulation study demonstrates that if the CDM has been misspecified, then across all experimental conditions, the three HACA methods obtain substantially better recovery of examinees' true proficiency-class membership than the exact EM-based MLE procedure using the wrong model. One could argue, of course, that the (most likely) correct model could always be identified by fitting two different CDMs to the data set in question, and then choose the one with the better fit. But what to do if test item responses conform to multiple CDMs (a scenario that was not considered here due to space limitations)? In this situation, nonparametric classification techniques can offer a solution to the task of assigning examinees to proficiency classes.

Finally, it should be recalled that the examinee clusters obtained from nonparametric classification methods serve as proxies for the proficiency classes. But nonparametric classification methods cannot estimate the attribute profiles underlying the clusters. Hence, the clusters must be interpreted or *labeled*—that is, their underlying attribute profiles must be reconstructed from the chosen input data, which can be tedious if the number of examinees is large. For $\boldsymbol{W}$ as input to clustering, Chiu et al. (2009) developed an automatic cluster labeling algorithm that seeks to identify an optimal match between examinees' within-cluster sum-score profiles and candidate attribute profiles potentially underlying this cluster. This algorithm has been further developed so that now also $\boldsymbol{W}_{\text{aug}}$ as input to clustering can be accommodated. (Both versions of the algorithm are written in R.) As an example, the labeling algorithm was applied to K. K. Tatsuoka's (1984) fraction–subtraction data when $\boldsymbol{W}_{\text{aug}}$ was used as input to HACA and the cluster tree was cut at $M = 10$—the maximum number of theoretically identifiable proficiency classes. Remarkably, the empirically identified proficiency-class labels agree with the labels of the 10 proficiency classes that are theoretically identifiable.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

The online appendix is available at http://apm.sagepub.com/supplemental

References

Ayers, E., Nugent, R., & Dean, N. (2008). Skill set profile clustering based on student capability vectors computed from online tutoring data. In R. S. J. de Baker, T. Barnes, & J. E. Beck (Eds.), *Educational data mining 2008: Proceedings of the 1st International Conference on Educational Data Mining* (pp. 210-217). Montréal, Québec, Canada: International Data Mining Society.

Chiu, C.-Y. (2008). *Cluster analysis for cognitive diagnosis: Theory and applications* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3337778)

Chiu, C.-Y., & Douglas, J. A. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns. *Journal of Classification*, *30*, 225-250.

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

DeCarlo, L. T. (2011). 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. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, *45*, 343-362.

DE LA TORRE, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*, 115-130.

de la Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, *73*, 595-624.

Hartigan, J. A. (1975). *Clustering algorithms*. New York, NY: Wiley.

Hartz, S. M. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practicality* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3044108)

Hartz, S. M., & Roussos, L. A. (October 2008). *The Fusion Model for skill diagnosis: Blending theory with practicality* (Research Report No. RR-08-71). Princeton, NJ: Educational Testing Service.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193-218.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241-254.

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.

Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement*, *36*, 548-564.

Liu, Y., Xu, G., & Ying, Z. (2011). *Learning item-attribute relationship in Q-matrix based diagnostic classification models* (Report: arXiv:1106.0721). New York City, NY: Department of Statistics, Columbia University. Retrieved from http://arxiv.org/pdf/1106.0721v1.pdf

Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, *2*, 99-120.

Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, *33*, 379-416.

Papadimitriou, C. H., & Steiglitz, K. (1998). *Combinatorial optimization: Algorithms and complexity* (2nd ed.). Mineola, NY: Dover.

Park, Y. S., & Lee, Y.-S. (2011). Diagnostic cluster analysis of mathematics skills. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, *4*, 75-107.

Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2014). *CDM: Cognitive Diagnosis Modeling* (R package version 3.1-14). Retrieved from http://CRAN.R-project.org/package=CDM

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

Steinley, D. (2004). Properties of the Hubert-Arabie Adjusted Rand Index. *Psychological Methods*, *9*, 386-396.

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, K. K. (1984). *Analysis of errors in fraction addition and subtraction problems* (Report NIE-G-81-0002). Urbana: Computer-Based Education Research Library, University of Illinois. (ED257665). Retrieved from http://www.eric.ed.gov

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconception by the pattern classification approach. *Journal of Educational Statistics*, *10*, 55-73.

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

von Davier, M. (2005, September). *A general diagnostic model applied to language testing data* (Research Report No. RR-05-16). Princeton, NJ: Educational Testing Service.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*, 287-301.

Ward, J. H. Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236-244.

Willse, J., Henson, R., & Templin, J. (2007, April). *Using sum scores or IRT in place of cognitive diagnostic models: Can more familiar models do the job?* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.