JEM JOURNAL of EDUCATIONAL MEASUREMENT NCME national council on measurement in education

Journal of Educational Measurement
Summer 2010, Vol. 47, No. 2, pp. 227–249

Factors Affecting the Item Parameter Estimation and Classification Accuracy of the DINA Model

Jimmy de la Torre and Yuan Hong
Rutgers, the State University of New Jersey
Weiling Deng
Educational Testing Service

To better understand the statistical properties of the deterministic inputs, noisy "and" gate cognitive diagnosis (DINA) model, the impact of several factors on the quality of the item parameter estimates and classification accuracy was investigated. Results of the simulation study indicate that the fully Bayes approach is most accurate when the prior distribution matches the latent class structure. However, when the latent classes are of indefinite structure, the empirical Bayes method in conjunction with an unstructured prior distribution provides much better estimates and classification accuracy. Moreover, using empirical Bayes with an unstructured prior does not lead to extremely poor results as other prior-estimation method combinations do. The simulation results also show that increasing the sample size reduces the variability, and to some extent the bias, of item parameter estimates, whereas lower level of guessing and slip parameter is associated with higher quality item parameter estimation and classification accuracy.

For assessments to help inform classroom instruction and learning, they must be cognitively diagnostic in that they must provide information that is "interpretative, diagnostic, highly informative, and potentially prescriptive" (Pellegrino, Baxter, & Glaser, 1999, p. 335). Designing more informative and diagnostic assessments requires the use of appropriate tools to extract relevant information from the assessments. Although conventional unidimensional item response theory models (IRMs) employed in most educational assessment settings have been useful for scaling and ordering students on a proficiency continuum, they do not provide sufficient information to allow practical diagnosis of students' specific strengths and weaknesses that can be used to facilitate learning and instruction. In contrast to unidimensional IRMs, cognitive diagnosis models (CDMs) with underlying latent classes are developed specifically for identifying the presence or absence of multiple finer-grained skills in a particular domain. A more generic term for "skill" is *attribute*, which can also refer to a cognitive process, state of knowledge, or knowledge representation. A profile in the form of a vector of binary latent variables can be generated for each student to indicate which attributes the student does and does not possess. This profile provides important and specific information about the student's state of learning or understanding that is relevant for subsequent actions such as tailored instruction or remediation. In addition, because cognitive diagnosis modeling has the flexibility to integrate developments in other fields (e.g., cognitive and learning sciences) as a component of the model, the design, analysis, and use of educational and psychological assessments can be improved using this approach.

Despite the potential benefits of CMDs, several factors have limited their use in applied educational settings. These factors include the lack, if not total absence, of cognitively diagnostic assessments constructed from a CDM framework, the dearth of appropriate cognitive models that can be used to undergird item and Q-matrix development, and the relative novelty and complexity of CDMs as an alternative psychometric framework. In this study we aimed to encourage a wider application of CDMs by addressing the third factor by providing additional work that will help researchers and practitioners better understand some statistical properties of one particular CDM. Specifically, in this article we sought to investigate how different factors such as latent class structure, sample size, level of guessing and slip parameters, and estimation method affect the quality of item parameter estimates and attribute classification accuracy of the deterministic inputs, noisy "and" gate (DINA; Junker & Sijtsma, 2001) model. The impact of these factors on the DINA model was examined using simulated data.

Background

Q-matrix

Typically, CDMs require information about how each test item is related to each of the attributes. A Q-matrix (Embretson, 1984; Tatsuoka, 1983) can provide such information in a $J \times K$ matrix of zeros and ones, where J is the number of items, and K the number of attributes. In this matrix, each item is listed in a separate row, and each attribute in a separate column. The *jk*th element of the matrix, q_{jk}, is equal to 1 if the *k*th attribute is required to correctly answer the *j*th item; otherwise, this element is equal to 0. The Q-matrix plays an important role in test development in that it embodies the attribute blueprint or cognitive specifications for test construction (Leighton, Gierl, & Hunka, 2004). Under the most general formulation (i.e., when no constraint is imposed on the attribute structure), a Q-matrix with K columns can have $2^K - 1$ unique rows (i.e., 0 to 1 patterns) corresponding to the $2^K - 1$ possible nontrivial attribute patterns. Although the zero attribute pattern is interpretable (i.e., it represents the examinees who have not mastered an attribute), the Q-vector with all zeros is considered nonsensical because an item that measures none of the relevant attributes cannot be constructed. In many applications, particularly when K is large, only a subset of the attribute patterns can be represented in the Q-matrix. Fixing the impact of other factors (e.g., test length, level of guessing, and slip parameter), the resolution of the profiles that can be generated from a test given a particular CDM is determined by the structure of the Q-matrix. That is, depending on the Q-matrix a test is based on, some attribute patterns may be indistinguishable from one another. However, this does not constitute a serious limitation because in many applications, particularly when the attributes are structured in a certain fashion, some attribute patterns may only be of minor interest, if at all. Therefore, the Q-matrix, and consequently the test, can be designed to maximize the separation of specific attribute patterns of particular interest.

DINA Model

Let the vector of dichotomous item responses of student i to J items be denoted by $\mathbf{Y}_i$, and assume that the components of $\mathbf{Y}_i$ are statistically independent given the vector of K attributes required for the test, $\alpha_i = \{\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK}\}'$. The kth element of the attribute vector, $\boldsymbol{\alpha}_{ik}$, is a binary indicator which assumes the value $\alpha_{ik} = 1$ when student i possesses the kth attribute; otherwise, $\alpha_{ik} = 0$.

A simple but interpretable model that relates the distribution of $\mathbf{Y}_i$ to the attribute vector $\boldsymbol{\alpha}_i$ is the DINA model. The DINA model is a discrete latent variable CDM that allows for the modeling of both the cognitive information of items and the inferences about the cognitive attributes of examinees, and has been the foundation of several approaches to cognitive diagnosis and assessment (e.g., Doignon & Falmagne, 1999; Tatsuoka, 1995). Like many CDMs, implementation of the DINA model requires construction of a Q-matrix. The DINA model assumes that the deficiency of the student in one attribute cannot be compensated by the mastery of other attributes. This assumption is referred to as *noncompensatory*, which can be statistically represented by a *conjunctive condensation function* (Maris, 1995, 1999). Given the ith student's attribute vector $\boldsymbol{\alpha}_i$, and the jth row of the Q-matrix, the conjunctive nature of the model generates a latent response η_{ij} deterministically through the function $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$, $\eta_{ij} = 1$ or 0. η_{ij} is also called the ideal response, and represents a deterministic prediction of task performance from each examinee's knowledge or performance state. Hence, the model classifies students into two categories for each item: students who possess all the attributes required for a successful response to the item, and students who lack at least one of the required attributes. The conditional distribution of Y_{ij} given an attribute vector is determined solely by η_{ij}, thus, several attribute patterns may result in the same latent response. Additional items with different attribute requirements are necessary if the attributes patterns classified under the same group are to be further distinguished from one another.

The probabilistic component of the DINA model allows for the possibility that students who possess all the required attributes for an item may slip and answer the item incorrectly, whereas students who do not possess all the required attributes may guess and answer the item correctly. The probabilities of slip and guessing on item j are denoted by $s_j = P(Y_{ij} = 0 \mid \eta_{ij} = 1)$ and $g_j = P(Y_{ij} = 1 \mid \eta_{ij} = 0)$. Given s_j and g_j, the item response function can be written as

$$P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i) = (1 - s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}}. \tag{1}$$

As can be seen from Equation 1, the DINA model is a parsimonious conditional model and requires only two parameters for each item. Because of this property, the complexity of the DINA model is decided by the number of items, and is not related to the number of attributes that are presented in the Q-matrix. Invoking the assumption of conditional independence of the responses given the attributes, the likelihood function of the response vector of examinee i using the DINA model is given by

$$L(\boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i) = \prod_{j=1}^{J} P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i)^{y_{ij}} [1 - P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i)]^{1-y_{ij}}. \tag{2}$$

For other discussions and applications of the DINA model, see de la Torre (2009b), Doignon and Falmagne (1999), Haertel (1989), Junker and Sijtsma (2001), Macready and Dayton (1977), and Tatsuoka (1995). For extensions of the DINA model to multiple-choice, nominal, and continuous responses, see de la Torre (2009a), Templin, Henson, Rupp, and Jang (2008), and de la Torre and Liu (2008), respectively.

For estimation purposes, distinctions are made between incidental and structural parameters: the slip and guessing parameters are treated as the structural parameters, whereas the attributes are considered the incidental parameters. Consistent estimates of the structural parameters can be obtained when the incidental parameters are integrated out of Equation 2. By employing a marginalized maximum likelihood (MML) estimation approach, Equation 2 is replaced by the marginalized likelihood given by

$$L(\boldsymbol{Y}_i) = \prod_{l=1}^{2^K} L(\boldsymbol{Y}_i \mid \boldsymbol{\alpha}_l) p(\boldsymbol{\alpha}_l), \tag{3}$$

where $p(\boldsymbol{\alpha}_l)$ represents the prior probability of the attribute pattern $\boldsymbol{\alpha}_l$. In MML estimation, Equation 3 taken across the N examinees is maximized to obtain estimates of the item parameters. Maximization of the marginalized likelihood requires integration over the distribution of the latent classes, and when K is large the number of latent classes involved in each iteration can slow down the estimation process markedly. Additional assumptions can be brought to bear to simplify the structure of the latent classes to facilitate the model parameter estimation. Two such assumptions are discussed below. For additional discussion of the DINA model estimation algorithm, see de la Torre (2009b).

Latent Class Structures

Each attribute vector or pattern represents a specific combination of the attribute mastery and nonmastery, and therefore, can be viewed as a unique latent class. A switch from 0 to 1, or vice versa, on an attribute produces a completely different class. In addition, because the attributes are latent, a domain with K attributes has 2^K associated latent classes. The latent classes are considered unstructured if the attributes are independent, in that mastery of one attribute is not a prerequisite to the mastery of another attribute. In addition, the latent classes are deemed unstructured if they cannot be expressed independently of each other conditional on some traits or abilities. When the latent classes are unstructured, the 2^K attribute patterns or classes are all permissible.

Although many applications of cognitive diagnosis are based on the assumption that attributes are independent, cognitive research suggests that cognitive skills should not be investigated in isolation (Kuhn, 2001). In other applications, it is more reasonable to assume that attributes are dependent and follow some type of structure. In general, latent classes are considered structured if some constraints exist regarding the relationships among the attributes. These constraints theoretically determine whether some classes are remotely likely or completely unlikely under the assumed structure. In this article two specific dependent latent class structures are discussed, namely, the hierarchical and higher-order structures.

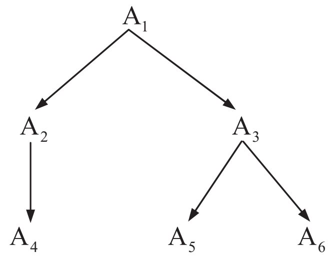

Figure 1. A divergent hierarchical structure for six attributes.

If the attributes are related such that the mastery of an attribute is a prerequisite to the mastery of another (i.e., attribute mastery is sequential), the structure of the latent classes is said to be hierarchical. The first and most important step in specifying a model with a hierarchical structure is the identification of the psychological ordering of attributes required to answer an item correctly. Because the attribute ordering directly influences the evaluation of the students' performance, and hence the accuracy of the classification of their cognitive ability, the process needs to be grounded theoretically.

Of the different hierarchical attribute structures presented by Leighton et al. (2004), the divergent structure is pursued in this article. Figure 1 gives a divergent structure involving six attributes, A_1 through A_6. A_1 represents the initial attribute that needs to be mastered and is the prerequisite for the mastery of all the other attributes. Going down the structure, A_2 is shown as a prerequisite to A_4, whereas A_3 is a prerequisite to both A_5 and A_6. Under this hierarchical structure, A_6 cannot be mastered until after both A_1 and A_3 have been mastered, but A_6 is not ordinally related to A_5. When the relationship among the attributes leading to successful performance is ordered and can be structured hierarchically, the number of latent classes can be reduced dramatically. In this example, only 16 of the original 64 latent classes need to be considered. Although a Q-matrix, hence item specifications, can be designed to correspond to the hierarchical structure of the attributes, it is not a necessary requirement for all applications. In some instances, items requiring more complex but not more elementary attributes can be constructed. For example, although "taking the derivative" presupposes knowledge of basic arithmetic operations, items measuring purely the ability to differentiate and not explicitly requiring basic arithmetic operations can still be constructed.

The latent classes are said to have a higher-order structure if they follow the model presented by de la Torre and Douglas (2004) where an examinee's overall ability θ determines the probability that the examinee masters each of the attributes. This formulation is reasonable when mastery of the specific attributes can be thought of as a function of a general ability. Under this type of latent class structure, attributes are assumed to be locally independent given the overall ability θ of an examinee. Here, the relationship between attributes and the overall ability θ is analogous to the relationship between items and latent proficiency in conventional IRMs.

Depending on the number of attributes and how stable they can be estimated, a choice can be made between several unidimensional and multidimensional IRMs to relate the overall ability θ and latent attribute vector $\boldsymbol{\alpha}$. However, in applications where only a few attributes are involved (e.g., $K \leq 8$), it is prudent to use simpler IRMs, and the one-parameter logistic model is one such model. In this formulation, the logit of the probability of possessing the kth attribute is a linear function of the overall ability, and is given by

$$p(\alpha_k \mid \theta) = \frac{\exp[1.7\lambda_1(\theta - \lambda_{0k})]}{1 + \exp[1.7\lambda_1(\theta - \lambda_{0k})]}, \tag{4}$$

where λ_{0k} is the location parameter of attribute k and λ_1 is the common slope parameter. Examinees with higher θ are more likely to possess the latent attributes compared to those with lower θ. In addition, the location parameter of an attribute is analogous to the item difficulty parameter of an item where a higher location parameter represents a higher degree of difficulty involved in mastering the attribute.

Although attributes that are conditionally independent given a higher-order latent trait are structured, the higher-order structure is considered indefinite. It differs from the hierarchical structure, which has a definite structure, in that, instead of designating some latent classes as not permissible (i.e., prior probabilities are zero), some latent classes are deemed improbable and are assigned low prior probabilities under the higher-order latent class structure. For example, with a reasonable spread in the location parameters, latent classes that show mastery of the more difficult attributes and nonmastery of easier attributes have low associated probabilities. Although both the hierarchical and higher-order structure assumption can simplify the joint distribution of the attributes, estimation of the model parameter in the hierarchical structure can proceed with minor modifications of the MML estimation algorithm for unstructured latent classes (i.e., ignoring the nonpermissible latent classes), whereas model parameter estimation in the higher-order structure has been implemented using an MCMC algorithm in a hierarchical Bayesian framework (see de la Torre & Douglas, 2004, for details).

Simulation Study

Design

To better understand how the DINA model is affected by various factors, a simulation study was designed to investigate the impact of the structure of the latent classes (unstructured, hierarchical, or higher-order), type of prior (unstructured or hierarchical), level of guessing and slip parameter (high or low), sample size (1,000, 2,000, or 4,000), and estimation method (fully or empirical Bayes) on the quality of item parameter estimates and on the accuracy of attributes classification. A potential mismatch between the true structure and posited structure (i.e., prior) of the latent classes can adversely affect the estimation of the item parameters, which in turn can affect the attribute classification accuracy. The adverse impact can be exacerbated when a more constrained method (i.e., fully Bayes) is used, and possibly when the sample size is smaller. Items with lower guessing and slip parameters are expected to directly contribute to higher attribute classification accuracy.

Table 1
Q-matrix for Unstructured Attributes

Item	Attribute 1	2	3	4	5	6
1	1	0	0	0	0	0
2	0	1	0	0	0	0
3	0	0	1	0	0	0
4	0	0	0	1	0	0
5	0	0	0	0	1	0
6	0	0	0	0	0	1
7	1	1	0	0	0	0
8	0	1	1	0	0	0
9	0	0	1	1	0	0
10	0	0	0	1	1	0
11	0	0	0	0	1	1
12	1	0	0	0	0	1
13	0	0	1	1	1	1
14	1	1	0	0	1	1
15	1	1	1	1	0	0

Although it would be ideal to design Q-matrices to correspond to the latent class structure to a large extent, in practice the fact that the Q-matrix is usually developed based on the existing items makes it hard to achieve a perfect match between the Q-matrix specification and latent class structure. Consequently, we only used an unstructured Q-matrix for the three types of latent class structures considered in this study. Allowing a Q-matrix not to be contingent on the psychological ordering of the attributes is a more flexible approach and is consistent with Tatsuoka's (1983) view of the Q-matrix. In addition, this design allows us to examine the quality of item parameter estimates and the accuracy of attribute classification not only when a correct Q-matrix is used (i.e., when the latent classes are unstructured), but also when some degree of Q-matrix mismatch is involved (i.e., under hierarchical and higher-order structure conditions). The Q-matrix used in this study is given in Table 1, which represents a subset of the 64 possible latent classes, and the rows were selected to ensure that all attributes are measured the same number of times.

In this study, six attributes were considered. Under the unstructured condition, all 64 latent classes were generated with equal probability; for the hierarchical structure, the divergent structure given in Figure 1 was used, and the 16 possible latent classes were generated also with equal probability; for the higher-order structure, the overall ability was generated from a standard normal distribution, and the higher-order structural parameters were set to $\lambda_1 = 1.5$ and $\lambda_0 = \{-1.5, -1.0, -1.0, -0.5, -0.5, -0.5\}'$ which in turn determined the probabilities of the latent classes. The values of the higher-order structural parameters were selected so that the structure of the latent classes was very disparate from the two previous structures. Under this design, the latent class with all the attributes mastered was most likely, and had a probability of .50. The latent classes where one of the three most difficult attributes (i.e., attributes 4, 5, and 6) was not mastered had a probability of .11, whereas the latent class where the three easiest attributes (attributes 1, 2, and 3) were not mastered while the most difficult attributes were mastered had a probability of less than .002. Incidentally, the last latent class example represented the least likely combination of the attribute mastery, and, although possible, was very unlikely to occur under this particular higher-order structure. The latent classes generated using the three attribute structures were used in conjunction with a Q-matrix to generate the response data. For a fixed class structure, data were generated for each of the sample size and level of guessing and slip parameter combinations.

The low level of guessing and slip items used to generate the responses were items with guessing and slip parameters ranging from .05 to .15; the high level of guessing and slip items refer to the items with guessing and slip parameters ranging from .20 to .30. The guessing and slip parameters were generated from a uniform distribution for each item and were fixed across the different levels of the other factors.

The generated data were analyzed using both flat unstructured and hierarchical prior distributions. Flat prior distribution refers to the assignment of equal probability for all the permissible latent classes under a particular structure. In a flat hierarchical prior, all permissible latent classes are given equal and nonzero probabilities (i.e., 1/16). The use of unstructured prior distribution represents the most general approach in test construction and analysis because no prior knowledge or constraint regarding the structure of the latent classes is necessary, but may provide suboptimal results when the latent classes have a specific structure that can be used in the analysis. The final step in estimating the model parameters involved the use of the code described above with two manipulations on the prior distributions. Using empirical Bayes estimation, the prior weights were updated after each iteration, whereas prior weights were kept at their initial values for all iterations using the fully Bayes approach. When the prior was unstructured, both estimation methods were used; when the prior was hierarchical, only the fully Bayes method was used.

One hundred data sets were generated and analyzed for each condition. The average bias of item parameter estimates across the replications was computed for each item under each condition. The standard deviation and root mean squared error (*RMSE*) of the estimates across the replications were also computed. In addition to the quality of the item parameter estimates, the correct attribute classification rates under each condition were also determined. Attribute classification was based on the posterior mean of each examinee (i.e., $\hat{\alpha}_{ik} = 1$ if $\hat{P}(\alpha_{ik} \mid \boldsymbol{Y}_i) \geq 0.5$, and zero otherwise), and the classification rates were computed at the individual attribute and attribute-vector levels.

Results

Item Parameter Estimates

Impact of sample size. First, the impact of sample size on the quality of parameter estimates was examined by focusing on a fixed latent class structure. The accuracy of the estimates was relatively unaffected by the sample size (i.e., the mean estimates remained the same across the different sample sizes), but the precision (i.e., the variability of the estimates across the 100 replications), and consequently the *RMSE* of the estimates across the replications, decreased as the sample size increased. This pattern can be observed in Table 2 which gives the results for conditions that involved items of low level of guessing and slip parameters, unstructured latent classes and prior distribution, and empirical Bayes estimation. As can be seen from the table, although the estimates using larger sample sizes were generally less biased, the improvement was not dramatic. This could be an indication that for the DINA model a sample size of 1,000 is sufficient in providing accurate parameter estimates. In comparison, the variability of the estimates showed a clear improvement with larger sample size, particularly in moving from a sample size of 1,000 to 4,000. Similar general patterns were observed across the various levels of the other factors. Thus, for this reason and due to the space limitation, only results for one specific sample size, $N = 1000$, are presented below.

Impact of item parameters. Second, the impact of level of guessing and slip parameter on item parameter estimation was investigated. On the whole, the results (not presented) indicate that data generated with low level of guessing and slip items resulted in more accurate (in terms of bias) and precise (in terms of *SD*) parameter estimates. These results were expected because, everything being equal, more accurate estimates of the slip and guessing parameters can be obtained when they are smaller, as extreme probabilities have less sampling variability. However, regardless of the level of guessing and slip parameter involved, similar patterns of the influences of the other factors on the parameter estimation can be observed from the results. Consequently, in the interest of space, only results for conditions involving low level of guessing and slip item parameters are discussed at length.

Impact of latent class structure. Aside from level of guessing and slip parameter and sample size, the impact of the latent class structure on item parameter estimation was a major focus of this article. This was investigated in detail by analyzing and comparing the parameter estimates under all the conditions generated by the three types of latent class structures. When the latent classes were unstructured, the unstructured prior distribution matched the latent class structure. Although item parameter estimates based on the fully Bayes method were better, the empirical Bayes estimates were highly similar. Table 3 shows that both the guessing and slip parameters of the 15 items can be accurately estimated with the maximum absolute bias of only .004 for the two parameters. In addition, the standard deviations of the estimates across the 100 replications, and the *RMSE*s of the estimates, were all less than or equal to .04. In contrast, using a hierarchical prior when the true latent classes were unstructured resulted in very poor item parameter estimates. Table 3 shows that almost all the estimated values for both guessing and slip parameters deviated markedly from the true value. It can be noted that the guessing parameter estimates of 10 of the 15 items had absolute biases of at least .10 and a maximum absolute bias of .32. The same can be said of the slip parameters except that the maximum bias was even larger (i.e., .68). As a whole, all the 15 items had either a guessing or slip parameter absolute bias of

Table 2
Estimates of Guessing and Slip Parameters with Unstructured Latent Class
(Low level guessing and slip parameters, Empirical Bayes)

	$N = 1{,}000$						$N = 2{,}000$						$N = 4{,}000$					
	g			*s*			*g*			*s*			*g*			*s*		
Item	Bias	*SD*	*RMSE*	Bias	*SD*	*RMSE*	Bias	*SD*	*RMSE*	Bias	*SD*	*RMSE*	Bias	*SD*	*RMSE*	Bias	*SD*	*RMSE*
1	.00	.03	.03	.00	.02	.02	.00	.02	.02	.00	.02	.02	.00	.01	.01	.00	.01	.01
2	.00	.03	.03	.00	.02	.02	.00	.02	.02	.00	.01	.01	.00	.01	.01	.00	.01	.01
3	.00	.02	.02	.00	.02	.02	.00	.02	.02	.00	.01	.01	.00	.01	.01	.00	.01	.01
4	.00	.02	.02	.00	.02	.02	.00	.02	.02	.00	.02	.02	.00	.01	.02	.00	.01	.01
5	.00	.03	.03	.00	.02	.02	.00	.02	.02	.00	.01	.01	.00	.02	.02	.00	.01	.01
6	.00	.03	.03	.00	.02	.02	.00	.02	.02	.00	.02	.02	.00	.01	.01	.00	.01	.01
7	.00	.01	.01	.00	.02	.02	.00	.01	.01	.00	.02	.02	.00	.01	.01	.00	.01	.01
8	.00	.01	.01	−.01	.03	.03	.00	.01	.01	.00	.02	.02	.00	.01	.01	.00	.01	.01
9	.00	.01	.01	.00	.03	.03	.00	.01	.01	.00	.02	.02	.00	.01	.01	.00	.01	.01
10	.00	.01	.01	.00	.04	.04	.00	.01	.01	.00	.03	.03	.00	.01	.01	.00	.02	.02
11	.00	.01	.02	.00	.03	.03	.00	.01	.01	.00	.02	.02	.00	.01	.01	.00	.02	.02
12	.00	.01	.01	.00	.04	.04	.00	.01	.01	.00	.03	.03	.00	.01	.01	.00	.02	.02
13	.00	.01	.01	.00	.05	.05	.00	.01	.01	.00	.03	.03	.00	.01	.01	.00	.03	.03
14	.00	.01	.01	.01	.07	.07	.00	.01	.01	.00	.04	.04	.00	.01	.01	.00	.02	.02
15	.00	.01	.01	−.01	.04	.04	.00	.01	.01	.00	.03	.03	.00	.00	.00	.00	.02	.02

Table 3
Item Parameter Estimates Involving Unstructured Latent Classes
(Low level guessing and slip parameters, N = 1,000)

		Unstructured Prior						Hierarchical Prior		
		Fully Bayes			Empirical Bayes			Fully Bayes		
	Item	Bias	*SD*	*RMSE*	Bias	*SD*	*RMSE*	Bias	*SD*	*RMSE*
	1	.00	.02	.02	.00	.03	.03	.05	.08	.09
	2	.00	.02	.02	.00	.03	.03	.13	.03	.13
	3	.00	.02	.02	.00	.02	.02	.04	.03	.05
	4	.00	.02	.02	.00	.02	.02	.32	.02	.32
	5	.00	.02	.02	.00	.03	.03	.24	.02	.24
	6	.00	.02	.02	.00	.03	.03	.23	.02	.23
	7	.00	.01	.01	.00	.01	.01	.01	.01	.02
g	8	.00	.01	.01	.00	.01	.01	.02	.01	.03
	9	.00	.01	.01	.00	.01	.01	.13	.02	.14
	10	.00	.01	.01	.00	.01	.01	.19	.02	.19
	11	.00	.01	.01	.00	.01	.02	.12	.02	.12
	12	.00	.01	.01	.00	.01	.01	.14	.02	.14
	13	.00	.01	.01	.00	.01	.01	.11	.01	.11
	14	.00	.01	.01	.00	.01	.01	.14	.01	.14
	15	.00	.01	.01	.00	.01	.01	.07	.01	.07
	1	.00	.02	.02	.00	.02	.02	.47	.02	.47
	2	.00	.02	.02	.00	.02	.02	.29	.03	.29
	3	.00	.02	.02	.00	.02	.02	.36	.02	.36
	4	.00	.02	.02	.00	.02	.02	.09	.03	.09
	5	.00	.02	.02	.00	.02	.02	.09	.03	.10
	6	.00	.02	.02	.00	.02	.02	.08	.02	.08
	7	.00	.02	.02	.00	.02	.02	.56	.02	.56
s	8	−.01	.03	.03	−.01	.03	.03	.45	.03	.45
	9	.00	.03	.03	.00	.03	.03	.19	.05	.20
	10	.00	.03	.03	.00	.04	.04	.03	.02	.03
	11	.00	.03	.03	.00	.03	.03	.02	.02	.02
	12	.00	.03	.03	.00	.04	.04	.43	.03	.43
	13	.00	.05	.05	.00	.05	.05	.21	.10	.23
	14	.01	.06	.06	.01	.07	.07	.53	.05	.54
	15	−.01	.04	.04	−.01	.04	.04	.68	.04	.68

at least .10 when unstructured latent classes were analyzed using the hierarchical prior distribution.

The results in Table 4 support the expectation that when the data were generated using a hierarchically structured latent classes, the use of a prior distribution that corresponded to the generating condition (i.e., hierarchical prior distribution) would yield the most accurate results. Moreover, Table 4 also shows that when the data were from hierarchical latent classes, using an unstructured prior and updating the prior

Table 4
Item Parameter Estimates Involving Hierarchical Latent Classes
(Low level guessing and slip parameters, N = 1,000)

		Unstructured Prior						Hierarchical Prior		
		Fully Bayes			Empirical Bayes			Fully Bayes		
	Item	Bias	*SD*	*RMSE*	Bias	*SD*	*RMSE*	Bias	*SD*	*RMSE*
	1	.73	.04	.73	.38	.15	.41	.12	.09	.15
	2	.03	.02	.04	.08	.02	.08	.09	.02	.09
	3	.35	.03	.35	.05	.03	.06	.06	.02	.06
	4	.03	.01	.04	.10	.01	.10	.10	.01	.11
	5	.07	.02	.07	.10	.02	.10	.11	.02	.11
	6	.03	.01	.03	.08	.01	.08	.09	.01	.09
	7	.26	.03	.26	.06	.01	.07	.06	.01	.07
g	8	.15	.02	.15	.08	.01	.08	.08	.01	.08
	9	.08	.01	.08	.10	.01	.10	.10	.01	.10
	10	.06	.01	.06	.08	.01	.08	.08	.01	.08
	11	.07	.01	.07	.09	.01	.09	.09	.01	.09
	12	.15	.02	.15	.12	.01	.12	.12	.01	.12
	13	.10	.01	.10	.10	.01	.11	.11	.01	.11
	14	.15	.01	.15	.14	.01	.14	.14	.01	.14
	15	.10	.01	.10	.08	.01	.08	.08	.01	.08
	1	.09	.02	.09	.09	.01	.09	.09	.01	.10
	2	.05	.01	.05	.08	.01	.08	.08	.01	.08
	3	.04	.01	.04	.05	.01	.06	.06	.01	.06
	4	.23	.03	.23	.11	.02	.11	.11	.02	.11
	5	.21	.03	.21	.12	.02	.12	.11	.02	.12
	6	.15	.03	.15	.09	.02	.09	.09	.02	.09
	7	.02	.01	.02	.06	.01	.06	.06	.01	.06
s	8	.04	.01	.04	.08	.01	.08	.08	.01	.08
	9	.12	.03	.12	.10	.02	.10	.10	.02	.10
	10	.26	.04	.26	.09	.04	.10	.08	.03	.09
	11	.15	.03	.15	.10	.03	.10	.09	.02	.09
	12	.08	.02	.08	.11	.02	.12	.11	.02	.12
	13	.14	.06	.16	.11	.05	.12	.11	.05	.12
	14	.13	.04	.13	.14	.04	.15	.14	.04	.14
	15	.02	.02	.02	.07	.02	.08	.08	.02	.08

via empirical Bayes estimation generally yielded more accurate and precise guessing and slip estimates compared to using the same prior but without updating it (i.e., fully Bayes estimates). The average absolute bias across 15 items was reduced using the empirical method by about .05 for the guessing parameter and by about .02 for the slip parameter. The same magnitude of reduction for the slip and guessing parameters was observed in comparing the average *RMSE*s, whereas comparison of the two methods based on the *SD* was not conclusive. In addition, updating the unstructured prior provided estimates that were highly similar to those using structured prior—with the exception of g_1, the estimates obtained using updated unstructured prior and structure prior had a maximum absolute difference of .01 across all items and parameters.

In comparing the best results in Tables 3 and 4, it can be observed that, although the best method in Table 4 (i.e., hierarchical prior) had lower mean variability than that in Table 3 (i.e., unstructured prior-fully Bayes; .01 and .02 vs. .02 and .03 for the guessing and slip parameters, respectively), it also had higher mean absolute bias (.09 vs. .00 for both parameters), leading to a higher mean *RMSE* (.10 and .10 vs. .02 and .03 for the guessing and slip parameters, respectively). These results indicate that, despite the match in the prior and latent class structures, item parameters under the hierarchical latent class structure can be estimated precisely, albeit with a high degree of inaccuracy. Furthermore, it can be conjectured that, in addition to the correct prior structure, a match between the structures of the Q-matrix and latent classes may be important in accurately estimating the parameters of the DINA model when the latent classes are highly structured.

When the latent classes had a higher-order structure, none of the priors used in this study matched the true underlying latent class structure. However, Table 5 shows that the item parameters under this latent class structure can be accurately and precisely estimated using an unstructured prior in conjunction with an empirical Bayes approach—the maximum absolute bias across all parameters was .02, whereas the maximum *RMSE* was .06, except for g_1 which had an *RMSE* of .12. In contrast, using a fully Bayes estimation method to analyze the same data in conjunction with unstructured and hierarchical prior distributions resulted in estimates in Table 5 that were inferior compared to the empirical Bayes estimates, and the difference was most obvious in comparing the mean absolute bias. Whereas the empirical Bayes estimates were relatively unbiased, the mean absolute biases of the fully Bayes guessing and slip parameters were .12 and .05 when the prior was unstructured, and .06 and .02 when the prior was structured. Compared to the slip parameters, the fully Bayes guessing parameter estimates were most biased for both prior distributions indicating that the guessing parameter is more susceptible to prior misspecification. Taken together, results in Tables 3 through 5 serve as a caution regarding the problems one may encounter when the prior distribution employed does not match the true underlying latent class structure, particularly when the fully Bayes method is employed.

Impact of prior and estimation method. Next, we focus on the impact of the type of prior and method of estimation on the item parameter estimates under various conditions. The above results show that when the structure of the prior distribution matched the structure of latent classes, the fully Bayes method provided better, albeit in many cases only slightly better, estimates compared to the empirical Bayes method; when the structure of the prior distribution does not match the latent class structure, the empirical Bayes method can improve the accuracy and precision of the estimates, and in some cases dramatic improvements can be observed. A large disparity between the assumed and true latent class structures can result in poor estimates for some if not all of the item parameters, and updating the prior weights can only provide limited improvements. Although not shown, the advantage of the

Table 5
Item Parameter Estimates Involving Higher-Order Latent Classes
(Low level guessing and slip parameters, N = 1,000)

		Unstructured Prior						Hierarchical Prior		
		Fully Bayes			Empirical Bayes			Fully Bayes		
	Item	Bias	*SD*	*RMSE*	Bias	*SD*	*RMSE*	Bias	*SD*	*RMSE*
	1	.43	.04	.43	−.02	.12	.12	−.08	.02	.09
	2	.24	.06	.25	−.01	.04	.04	.08	.03	.08
	3	.21	.05	.22	.00	.06	.06	.00	.03	.03
	4	−.02	.03	.04	−.01	.03	.03	.15	.02	.15
	5	.01	.02	.03	.00	.03	.03	.09	.02	.10
	6	−.02	.03	.04	−.01	.03	.03	.10	.03	.11
	7	.30	.04	.31	.00	.03	.03	.00	.02	.02
g	8	.21	.04	.22	.00	.02	.02	.01	.02	.02
	9	.09	.03	.09	.00	.02	.02	.09	.02	.10
	10	.03	.02	.03	.00	.02	.02	.12	.02	.12
	11	.04	.02	.05	.00	.01	.01	.07	.02	.07
	12	.15	.03	.15	.00	.02	.02	.07	.03	.07
	13	.07	.02	.08	.00	.01	.02	.06	.02	.07
	14	.15	.03	.15	.00	.02	.02	.05	.02	.05
	15	.20	.03	.20	.00	.02	.02	.02	.01	.02
	1	−.02	.02	.03	.00	.01	.01	.03	.01	.04
	2	−.05	.01	.05	.00	.01	.01	.00	.01	.01
	3	−.03	.01	.03	.00	.01	.01	.03	.01	.04
	4	−.02	.02	.02	.00	.01	.01	−.01	.01	.02
	5	−.01	.02	.02	.00	.01	.01	−.02	.02	.02
	6	−.02	.02	.02	.00	.01	.01	−.02	.01	.02
	7	−.07	.02	.07	.00	.02	.02	.02	.01	.02
s	8	−.07	.02	.08	.00	.02	.02	.01	.01	.02
	9	−.05	.02	.05	.00	.01	.02	−.01	.01	.02
	10	−.03	.02	.03	.00	.02	.02	−.03	.01	.03
	11	−.03	.01	.03	.00	.01	.01	−.03	.01	.03
	12	−.08	.02	.09	.00	.01	.01	−.02	.02	.02
	13	−.06	.02	.06	.00	.02	.02	−.04	.01	.04
	14	−.10	.02	.10	.00	.02	.02	−.03	.02	.03
	15	−.10	.02	.10	.00	.02	.02	−.01	.01	.02

empirical Bayes method is more apparent in estimating item parameters of high level of guessing and slip parameters.

Attribute Misclassification Rate

Attribute-level misclassification rate. As in the item parameter estimates, only the results for $N = 1,000$ are presented because the sample size had negligible impact on the attribute classification accuracy. Figures 2 and 3 show the misclassification rates for the six attributes individually under various conditions when high and

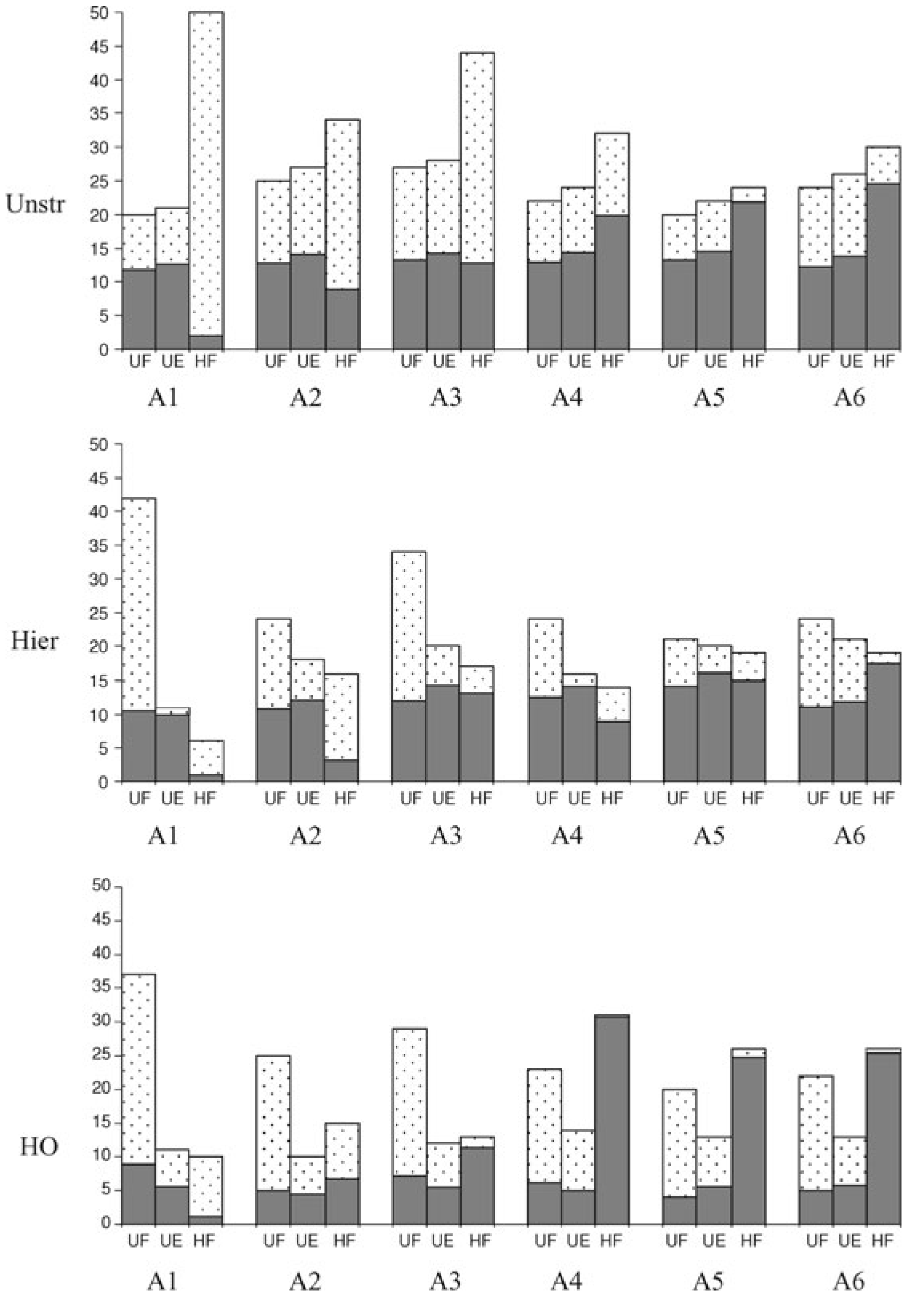

Figure 2. Percent of misclassified attributes (high level item parameters): Unstr = Unstructured; Hier = Hierarchical = H; HO = Higher-order; U = Unstructured; F = Fully Bayes; and E = Empirical Bayes.

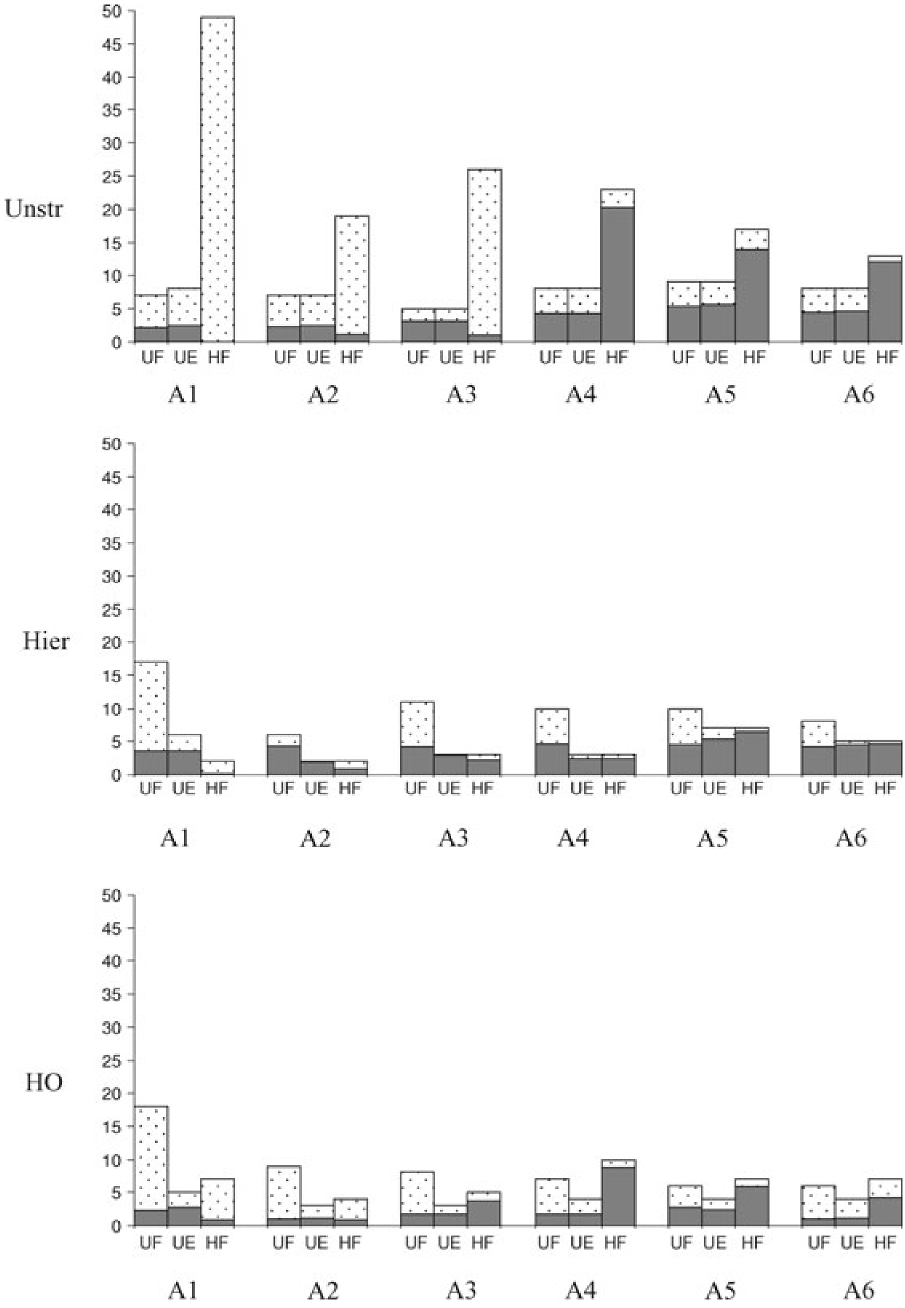

Figure 3. Percent of misclassified attributes (low level item parameters): Unstr = Unstructured; Hier = Hierarchical = H; HO = Higher-order; U = Unstructured; F = Fully Bayes; and E = Empirical Bayes.

low levels of item parameters were used, respectively. Each misclassification rate in these figures is also decomposed into its two components—false positive and false negative error rates, where the former is stacked on top of the latter. False positive error with respect to an attribute occurs when a student is classified as a master when the student has not mastered the attribute, whereas false negative error occurs when a student who has mastered the attribute is deemed a nonmaster.

The impact of level of guessing and slip parameters on the misclassification rate is very apparent when the two figures are compared panel by panel. As expected, data generated using low level of guessing and slip parameters provided lower misclassification rates, and this was true across all the conditions considered. For example, under the unstructured latent class condition (i.e., first panels of Figures 2 and 3) and unstructured prior distribution, the misclassification rates for individual attribute were below 9% when the guessing and slip parameters were low; these rates were dramatically higher (above 19%) when higher guessing and slip parameters were involved. To a large extent the general patterns of results in Figure 2 can also be observed in Figure 3, albeit involving different magnitudes. For one, the methods that produced the lowest misclassification rates when involving unstructured, hierarchical and higher-order latent classes were the unstructured-prior-fully-Bayes, hierarchical-prior-fully-Bayes, and unstructured-prior-empirical-Bayes methods, respectively. Moreover, updating an unstructured prior (i.e., the unstructured-prior-empirical-Bayes combination) did not result in a higher misclassification rate for any attribute across all conditions; rather, this method produced correct classification rates that were optimal or, in many cases, close to the optimal results. In addition, regardless of the levels of the item parameters, the attribute misclassification rate was affected by the underlying class structure. For example, attributes with unstructured latent classes had generally higher misclassification rates. Due to the similarities of the findings between figures, only Figures 2 and 3, which contrast the results more clearly, will be discussed in detail.

As noted above, the impact of prior type and estimation method depended on the structure of the latent classes. Figure 2 shows that when the structure of the latent classes was known (i.e., Panels 1 and 2), using a prior that matched the latent class structure and keeping the prior weights constant across the iterations produced the lowest misclassification rates. In addition, in comparing the results for conditions involving unstructured latent classes and prior (i.e., Panel 1), updating the correct prior weights using empirical Bayes resulted in slightly higher misclassification rates, which can be attributed to the additional noise incurred in updating a correct prior distribution. Finally, when the structure of the latent classes was not of a definite structure (i.e., Panel 3), using a general prior and updating the weights produced the lowest misclassification rates.

It should be noted that using an incorrect prior with fixed weights can produce higher and, in some cases, dramatically higher misclassification rates. This can be seen in the use of the hierarchical-prior-fully-Bayes method in Panel 1, the unstructured-prior-fully-Bayes method in Panel 2, and the same two methods in Panel 3. In contrast, results across the three panels indicate that misclassification rates from using an unstructured prior that allowed updates on the weights may not always be the lowest, but they were also never the highest. Thus, unless the latent class structure is known, a more general prior in conjunction with an empirical Bayes method may represent the safest approach in estimating the DINA model parameters and in classifying the examinee attributes.

Other things being equal, one can expect better item parameter estimation and lower attribute misclassification rates to go hand in hand, and vice versa. The findings of this study supported this expectation. In general, methods that led to better item parameters also resulted in lower misclassification rates (e.g., analyzing data generated from higher-order latent class structure using the unstructured-prior-empirical-Bayes method). In addition, methods that produced poor item parameter estimates also resulted in higher misclassification rates (e.g., analyzing data generated from unstructured latent class structure using the hierarchical-prior-fully-Bayes method).

Attribute-level false positive and negative error rates. Figure 2 also gives the decomposition of the misclassification rate into its false negative (lower portion of the bar) and false positive (upper portion of the bar) error components. The figure shows that the proportion of false positive error rate relative to false negative error rate of the different estimation methods exhibited contrasting patterns that varied considerably across the various attributes and latent class structures. For the unstructured-prior-fully-Bayes method, the misclassification rates were mostly false negative errors when it was the appropriate method (i.e., Panel 1); however, the misclassification rates were highly dominated by false positive errors when it was an inappropriate method (Panels 2 and 3). The structured-prior-fully-Bayes method showed a less discernible pattern: whether or not it was the appropriate method, the attribute misclassification rate can be largely false positive or false negative errors. For example, when it was the appropriate method (i.e., Panel 2), misclassifications in A_1 and A_2 were mostly false positive errors whereas misclassification in the last four attributes were mostly false negative errors. The same pattern across the attributes can be seen in Panel 3, albeit more extreme in the case of the last four attributes, where the method was not appropriate, whereas a different pattern can be seen in Panel 1 where the method was also inappropriate—the dominance of the false positive error extended to the third attribute. Finally, unstructured-prior-empirically-Bayes methods showed a predominance of false negative errors in Panels 1 and 2, and mostly false positive errors in Panel 3.

Attribute-pattern misclassification rate. Figure 4 presents the attribute vector misclassification rates of the different methods. In addition, Figure 4 also shows the number of misclassified individual attributes in these vectors. Correctly classifying all the attributes represents a more demanding task. Consequently, the misclassification rates for attribute vectors were also higher—the lowest and highest misclassification rates were 18% (Low condition of Panel 2), and 91% (High condition of Panel 1), respectively. Under the same conditions, the lowest and highest misclassification rates for individual attributes were only 2% and 50%, respectively.

As expected, the attribute vector misclassification rate was dramatically lower when items of low level of guessing and slip parameters were involved. The only exception was the structured-prior-fully-Bayes method under the unstructured latent class condition where smaller guessing and slip parameters did not result in a substantially lower misclassification rate. It can be noted that for all misclassification

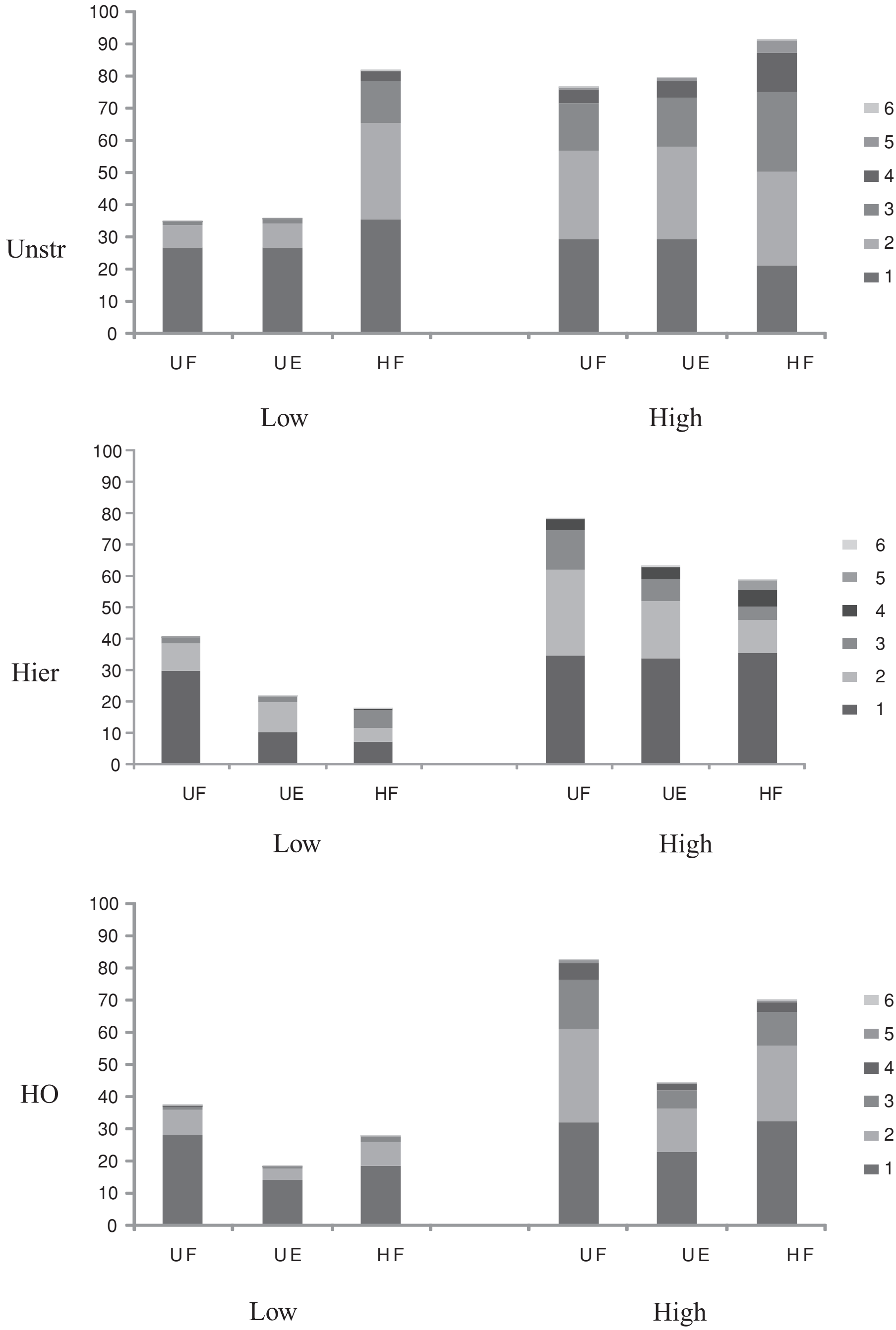

Figure 4. Percent of misclassified attribute vectors: Unstr = Unstructured; Hier = Hierarchical = H; HO = Higher-order; U = Unstructured; F = Fully Bayes; and E = Empirical Bayes.

rates, the number of misclassified attributes was generally small. For the most part, one or two misclassified attributes account for the overwhelming majority of the misclassified vectors. This was particularly true when the data were generated using low level of guessing and slip parameters and prior distribution was unstructured—nearly 95% of the misclassified vectors contain only one or two misclassified attributes whether or not the prior was updated.

Discussion

Cognitive diagnosis modeling has the potential of allowing educational assessment to be used as a tool that can facilitate instruction and learning. To encourage wider applications of this approach, this study investigated the impact of latent class structure, type of prior, sample size, level of guessing and slip parameters, and estimation method on the accuracy and precision of the slip and guessing parameter estimates and on the accuracy of attribute classification for the DINA model. Accurate estimation of item parameters and accurate classification of attributes are of paramount importance because they are necessary conditions in obtaining valid inferences in cognitive diagnosis.

The results of the simulation study indicate that in estimating the DINA model parameters and classifying the attributes, unless the latent classes have been established to be of a particular structure, the empirical Bayes method with fewer constraints on the prior distribution should be preferred over a more structured prior distribution particularly when the fully Bayes method is used. This is because the unstructured prior and empirical Bayes combination provides more robust results across various conditions. The simulation study also indicates that a relatively small sample size suffices to accurately estimate the DINA model parameters, and, in addition to the prior distribution and estimation method, level of guessing and slip parameter can have a profound impact in minimizing the attribute misclassification rate. Moreover, the simulation results show that the structure of the Q-matrix can potentially affect the quality of the item parameter estimates, and varying levels of false positive, false negative and overall misclassifications can be obtained for different underlying class structures.

The factors examined in this article represent but a subset of the factors affecting the quality of DINA model parameter estimation and attribute classification accuracy. In this regard, our study has several limitations that could be of interest in future studies. For one, the current study generated the data using specific instances of the latent class structures. That is, the permissible latent classes were equiprobable, and the higher-order structural parameters were based on some fixed values. Future studies can be undertaken to verify whether the same conclusions can be arrived at when the conditions are varied. Of particular interest is how the empirical Bayes estimates compare to the fully Bayes estimates when the structure of the prior distribution corresponds to the latent class structure but the prior weights do not match the weights used in generating the data. This particular design can underscore the usefulness of updating the prior weights even when the latent class structure (but not the relative weight) is known. Related to this, one may also examine the trade-off involved in empirically estimating the prior weights when the sample size is much smaller (e.g., $N = 100$).

In the simulation study, the Q-matrix and test length, two factors that have direct impact on the attribute classification rate, were fixed. Follow-up studies in the same vein should examine the ideal test lengths needed to achieve specific levels of attribute classification accuracy as a function of various factors (e.g., level of guessing and slip parameter, number of attributes, attribute structure). Future studies can also examine the interaction between the Q-matrix and the latent class structure. For example, using the method introduced by Henson and Douglas (2005), one can construct a test from an item pool designed to optimize correct classification rate. However, given the wide variety of latent class structures that can be considered, it is not clear that a single test of fixed length can provide optimal attribute classification rates across different latent class structures. It would also be interesting to investigate whether the impact of the Q-matrix misspecification on item parameter estimation and attribute classification that Rupp and Templin (2008) observed can be obtained when different attribute structures are involved.

Moreover, in our current design, we carefully selected the Q-matrix so that each attribute is measured by the same number of items (i.e., five), and the items measuring an attribute have the same composition (i.e., 1 one-attribute, 2 two-attribute, and 2 four-attribute items). However, research (e.g., Rupp & Templin, 2008) has shown that the proportion of items measuring an attribute and the number of attributes measured by an item can affect estimation accuracy. It would be instructive for future studies to investigate how these factors interact with the factors considered in this study to affect item parameter estimation and attribute classification accuracy.

Although this article looks at the impact of several factors on one specific latent-class CDM (i.e., the DINA model), the findings of the study can have implications on other CDMs that are based on the DINA model. These models include versions of the DINA model for continuous response, nominal response, cognitively-based multiple choice option, and multiple strategies, and a generalization of the DINA model (de la Torre, 2008, 2009a; de la Torre & Douglas, 2008; de la Torre & Liu, 2008; Templin et al., 2008). For even greater generality, future studies can consider other CDMs such as the NIDA and reparametrized unified models (de la Torre & Douglas, 2004; Hartz, 2002; Junker & Sijtsma, 2001). It would be worthwhile to examine whether the above factors have the same impact on the parameter estimation and attribute classification of these models.

Finally, it should be noted that the current implementation of the code in Ox (Doornik, 2003) for the DINA model can be considered highly efficient. For the condition involving low level of guessing and slipping items, 1,000 examinees, and unstructured latent classes with unstructured prior distribution, the average time required to estimate the model parameters using a fully or empirical Bayes method and a convergence criterion of .001 on a computer with a 3.2 GHz processor and 1 GB of memory was under 5 seconds.

References

de la Torre, J. (2008, July). *The generalized DINA model*. Paper presented at the International Meeting of the Psychometric Society, Durham, NH.

de la Torre, J. (2009a). A cognitive diagnosis model for cognitively-based multiple-choice options. *Applied Psychological Measurement*, *33*, 163–183.

de la Torre, J. (2009b). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*, 115–130.

de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333–353.

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

de la Torre, J., & Liu, Y. (2008, March). A *cognitive diagnosis model for continuous response*. Paper presented at the meeting of the National Council on Measurement in Education, New York, NY.

Doignon, J. P., & Falmagne, J. C. (1999). *Knowledge spaces*. New York, NY: Springer-Verlag.

Doornik, J. A. (2003). *Object-oriented matrix programming using Ox (Version 3.1)* [Computer software]. London, England: Timberlake Consultants Press.

Embretson, S. E. (1984). A general multicomponent latent trait model for response processes. *Psychometrika*, *49*, 175–186.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*, 333–352.

Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality*. Unpublished doctoral dissertation, University of Illinois, Champaign.

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

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.

Kuhn, D. (2001). Why development does (and does not) occur: Evidence from the domain of inductive reasoning. In J. L. McClelland & R. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives* (pp. 221–249). Hillsdale, NJ: Erlbaum.

Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement*, *41*, 205–236.

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

Maris, E. (1995). Psychometric latent response models. *Psychometrika*, *60*, 523–547.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*, 187–212.

Pellegrino, J. W., Baxter, G. P., & Glaser, R. (1999). Addressing the "two disciplines" problem: Linking theories of cognition and learning with assessment and instructional practices. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (pp. 307–353). Washington, DC: American Educational Research Association.

Rupp, A., & Templin, J. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, *68*, 78–96.

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

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327–359). Hillsdale, NJ: Erlbaum.

Templin, J., Henson, R., Rupp, A., & Jang, E. (2008). *Cognitive diagnosis models for nominal response data*. Paper presented at the meeting of the National Council on Measurement in Education, New York, NY.

Authors

JIMMY DE LA TORRE is an Associate Professor of Educational Psychology at Rutgers University, 10 Seminary Place, New Brunswick, NJ 08901; j.delatorre@rutgers.edu. His primary research interests include item response theory, cognitive diagnosis, Bayesian analysis, and the use of diagnostic assessments to support classroom instruction and learning.

YUAN HONG is a Research Associate at the Department of Educational Psychology, Rutgers University, 10 Seminary Place, New Brunswick, NJ 08901; yuanhong@eden.rutgers.edu. Her primary research interests include item response theory, cognitive diagnosis, value-added modeling and Bayesian analysis.

WEILING DENG is an Associate Psychometrician at Educational Testing Service, Mail Stop 13-P, 660 Rosedale Road, NJ 08541; wdeng@ets.org. Her primary research interests include psychometric methods.