Journal of Educational and Behavioral Statistics
March 2009, Vol. 34, No. 1, pp. 115–130
DOI: 10.3102/1076998607309474

DINA Model and Parameter Estimation: A Didactic

Jimmy de la Torre
Rutgers, The State University of New Jersey

Cognitive and skills diagnosis models are psychometric models that have immense potential to provide rich information relevant for instruction and learning. However, wider applications of these models have been hampered by their novelty and the lack of commercially available software that can be used to analyze data from this psychometric framework. To address this issue, this article focuses on one tractable and interpretable skills diagnosis model—the DINA model—and presents it didactically. The article also discusses expectation-maximization and Markov chain Monte Carlo algorithms in estimating its model parameters. Finally, analyses of simulated and real data are presented.

Keywords: *cognitive diagnosis; skills diagnosis; DINA; Markov chain Monte Carlo; expectation-maximization; parameter estimation*

Cognitive diagnosis models (CDMs) are psychometric models that can be used to evaluate students' strengths and weaknesses. These models provide specific information in the form of score profiles that can allow for effective measurement of student learning and progress, designing of better instruction, and possibly intervention to address individual and group needs. In contrast, traditional unidimensional item response models (IRMs) are primarily useful for scaling and ordering students on a latent proficiency continuum. However, the overall scores provided by IRMs do not contain sufficient information to aid in designing targeted instruction and tailored remediation.

As an alternative to unidimensional IRMs, cognitive and skills diagnosis models are developed for the purpose of identifying the presence or absence of multiple fine-grained skills required for solving problems on a test. In the literature, the presence and absence of skills are referred to as *skills mastery* and *non-mastery*, respectively, and are represented by a vector of binary latent variables. Thus, instead of a single score, a profile can be generated for a student or a group of students (e.g., class, district) to indicate which skills each student has or has not mastered. These profiles contain rich and relevant information that can have immense practical implications on classroom instruction and learning.

Although researchers and practitioners are becoming aware of CDMs and are starting to recognize their usefulness in providing rich information, this framework has remained underutilized because of two major limitations. First, as

The author thanks Kikumi Tatsuoka for generously sharing the fraction subtraction data.

compared to traditional IRMs, CDMs such as the DINA (deterministic inputs, noisy "and" gate), NIDA (noisy inputs, deterministic "and" gate), and reparametrized unified models (de la Torre & Douglas, 2004; Hartz, 2002; Junker & Sijtsma, 2001) are relatively novel and in some cases, more complex (for an extensive survey of various CDMs and their features, such as nature of underlying construct, response type, dimensionality, see Fu & Li, 2007). Consequently, many researchers lack familiarity with these models and their properties. Second, unlike traditional IRMs, which can be analyzed using commercially available software (e.g., BILOLG-MG; Zimowski, Muraki, Mislevy, & Bock, 1996), accessible computer programs for CDMs are not readily available. As a result, implementations of these models have been hampered. The goal of this article is to focus on one particular CDM that is tractable and interpretable—namely, the DINA model—and show how model parameters can be estimated using expectation-maximization (EM) and Markov chain Monte Carlo (MCMC) algorithms. Although computer code based on these algorithms can be obtained by request, the article provides a reference work for researchers who are interested in designing new algorithms.

1. The DINA Model

Let X_{ij} be the response of examinee i to item j, $i = 1, \ldots, I$, $j = 1, \ldots, J$, and let $\boldsymbol{\alpha}_i = \{\alpha_{ik}\}$ be the examinee's binary skills vector, $k = 1, \ldots, K$, where a 1 on the kth element denotes presence or mastery of skill k and 0, absence or nonmastery of the skill. When a general interpretation is intended, the generic term *attribute* can be used to subsume a skill, knowledge representation, or cognitive process. Implementation of most CDMs requires the construction of a Q-matrix (Embretson, 1984; K. Tatsuoka, 1985), which is a $J \times K$ matrix of zeros and ones, and the element on the jth row and kth column of the matrix, q_{jk}, indicates whether skill k is required to correctly answer item j. A Q-matrix can be viewed as a cognitive design matrix that explicitly identifies the cognitive specification for each item.

To illustrate, consider the mixed fraction subtraction domain at the middle school level. Mastery of this domain requires students to master the following set of five skills: subtract basic fractions, reduce and simplify, separate whole from fraction, borrow from whole, and convert whole to fraction. An item that has been used in this domain is $7\frac{3}{5} - \frac{4}{5}$; this item requires the first, third, and fourth skills to be answered correctly. The rows of the Q-matrix corresponding to this item would contain the vector (1, 0, 1, 1, 0).

In the DINA gate model, an examinee's skills vector and the Q-matrix produce a latent response vector $\boldsymbol{\eta}_i = \{\eta_{ij}\}$, where

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}. \tag{1}$$

The latent response Equation 1 assumes a value of 1 if examinee i possesses all the skills required for item j and a value of 0 if the examinee lacks at least one of the required skills. The *"and" gate* component of the model refers to the conjunctive process in determining η_{ij} in that a correct response to an item requires the presence all the prescribed skills for the item. In the example above, students who lack any one of the three required skills are not expected to answer the item correctly. Thus, the DINA model exhibits the same property as standard noncompensatory multidimensional IRMs. If the process is completely deterministic (i.e., error-free or nonstochastic), the latent response vector is identical to the manifest or observed response vector. However, because the underlying process is inherently stochastic, the latent response vector represents only an ideal response pattern. The noise introduced in the process is due to *slip* and *guessing* parameters—that is, examinees who possess all the required skills for an item can slip and miss the item, and examinees who lack at least one of the required skills can guess and still answer the item correctly with typically nonzero probabilities. In the DINA model, the slip and guessing parameters of item j are defined as $s_j = P(X_{ij} = 0 | \eta_{ij} = 1)$ and $g_j = P(X_{ij} = 1 | \eta_{ij} = 0)$, respectively. Therefore, the probability of examinee i with the skills vector $\boldsymbol{\alpha}_i$ answering item j correctly is given by

$$P_j(\boldsymbol{\alpha}_i) = P(X_{ij} = 1 | \boldsymbol{\alpha}_i) = g_j^{1-\eta_{ij}}(1 - s_j)^{\eta_{ij}}. \tag{2}$$

From this equation, answering an item correctly requires an examinee who has all the necessary skills to not slip and an examinee who lacks at least one of the required skills to guess correctly. Note that if there is no guessing and no slippage, the model probability of correct response to an item is either 0 or 1; that is, the response is solely determined by the interaction of $\boldsymbol{\alpha}$ and the Q-vector for the item. However, as noted by de la Torre and Douglas (2004), guessing in this context assumes a general interpretation; it is not confined to a correct response arrived through a random response but rather includes the use of alternative strategies not articulated in the Q-matrix. For example, if an item can be solved using a different set of skills, examinees who possess these skills but not those prescribed in the Q-matrix may appear to be guessing but in fact are systematically solving the problem using a different strategy.

A graphical representation of the DINA model is given in Figure 1. As the graph shows, the latent response η_{ij} is a function of the examinee's skills $\{\alpha_{ik}\}$ and the requisites of the item $\{q_{jk}\}$. Once η_{ij} has been determined, the probability that examinee i will give a correct response to item j is g_j if $\eta_{ij} = 0$ and $1 - s_j$ if $\eta_{ij} = 1$.

The DINA model is a parsimonious and interpretable model that requires only two parameters for each item (i.e., g_j and s_j) regardless of the number of attributes being considered, and despite its simplicity, it has been shown to provide good model fit (e.g., de la Torre & Douglas, 2004, 2005). De la Torre and Douglas (2004) and Junker and Sijtsma (2001) provide some applications of the DINA model. Although labeled differently, other discussions of the DINA

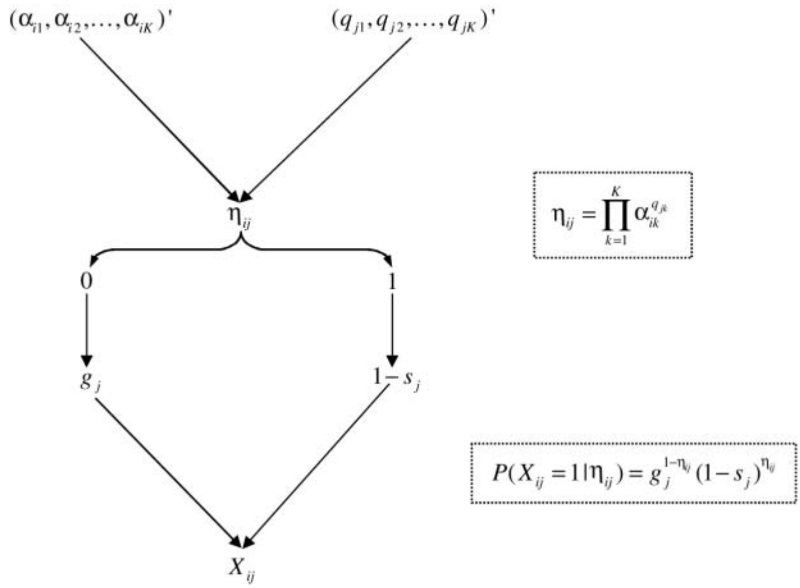

FIGURE 1. *A graphical representation of the examinee* i*'s response process to item* j.

model can be found in Doignon and Falmagne (1999), Haertel (1989), Macready and Dayton (1977), and C. Tatsuoka (2002). In is worth noting that although the DINA model under this formulation can be viewed as an extension of traditional IRMs (Embretson & Reise, 2000; van der Linden & Hambleton, 1997), it differs from the latter in some important respects. Primarily, the DINA model deals with multidimensional binary latent skills, whereas traditional IRMs deal with unidimensional continuous latent traits. Because the number of skills patterns in the DINA model is finite, each pattern can be viewed as a latent group or class. Consequently, the DINA model—and CDMs in general—can also be subsumed under multiple classification models (Maris, 1999) or restricted latent class models (Haertel, 1989).

2. Model Parameter Estimation

2.1. Joint Maximum Likelihood Estimation

The DINA model is a conditional distribution of X_{ij} given a skills vector $\boldsymbol{\alpha}_i$. Assuming conditional independence of the responses given the skills vector, the conditional likelihood of $\boldsymbol{X}_i$ can be written as

$$L(\boldsymbol{X}_i|\boldsymbol{\alpha}_i) = \prod_{j=1}^{J} P_j(\boldsymbol{\alpha}_i)^{X_{ij}}[1 - P_j(\boldsymbol{\alpha}_i)]^{1-X_{ij}}, \tag{3}$$

where $P_j(\boldsymbol{\alpha}_i)$ is as defined in Equation 2; assuming further that the examinees were randomly sampled, the conditional likelihood of the observed data $\boldsymbol{X}$ is equal to

$$L(\boldsymbol{X}|\boldsymbol{\alpha}) = \prod_{i=1}^{I} L(\boldsymbol{X}_i|\boldsymbol{\alpha}_i) = \prod_{i=1}^{I}\prod_{j=1}^{J} P_j(\boldsymbol{\alpha}_i)^{X_{ij}}[1 - P_j(\boldsymbol{\alpha}_i)]^{1-X_{ij}}. \tag{4}$$

If the skills vectors are known, estimating the model parameters $\boldsymbol{\beta} = (s_1, g_1, \ldots, s_J, \ g_J)'$ is straightforward. But in practice, the skills vectors are not known, and one way of estimating $\boldsymbol{\beta}$ is by joint maximum likelihood estimation, which allows the simultaneous estimation of the item parameters and the skills vectors. However, as in traditional IRMs, joint maximization of the structural parameter $\boldsymbol{\beta}$ and incidental parameter $\boldsymbol{\alpha}$ may lead to inconsistent $\hat{\beta}$ (Baker, 1992; Neyman & Scott, 1948).

2.2. Marginalized Maximum Likelihood Estimation

Instead of working with the conditional likelihood of $\boldsymbol{X}$ to obtain $\hat{\beta}$, the maximization can involve the marginalized likelihood of the data

$$L(\boldsymbol{X}) = \prod_{i=1}^{I} L(\boldsymbol{X}_i) = \prod_{i=1}^{I}\sum_{l=1}^{L} L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l)p(\boldsymbol{\alpha}_l), \tag{5}$$

where $L(\boldsymbol{X}_i)$ is the marginalized likelihood of the response vector of examinee i, $p(\boldsymbol{\alpha}_l)$ is the prior probability of the skills vector $\boldsymbol{\alpha}_l$, and $L = 2^K$. Because the joint distribution of $\boldsymbol{\alpha}_l$ is discrete, taking the weighted sum of the conditional likelihood across the 2^K possible skills patterns is equivalent to integration of the conditional likelihood over $g(\theta)$ in conventional IRMs when a unidimensional continuous latent proficiency is involved. Parameter estimation based on the marginalized likelihood (i.e., the marginal maximum likelihood estimation) can be implemented using EM algorithm. For details of an algorithm for estimating the DINA model parameters and their corresponding standard errors, see the appendix.

2.3. Higher-Order DINA Model and MCMC Estimation

To specify a complete latent variable model for cognitive diagnosis, both the conditional probability of a correct response given an attribution pattern and the joint distribution of the attribute patterns are needed. In the above specification, the conditional distribution is given by the DINA model, whereas the joint distribution is represented by a multinomial distribution. In the latter distribution, each attribute pattern corresponds to one category. Thus, in implementing the EM algorithm, the marginalization or expectation step involves updating $2^K - 1$

posterior probabilities corresponding to the number of possible skills patterns minus 1—that is, $\sum_{l=1}^{L} p(\boldsymbol{\alpha}_l|\boldsymbol{X}) = 1$. A model that considers all the possible latent classes represents the saturated model and is considered the most general formulation of the joint distribution. However, when the number of attributes is moderately large, the marginal maximum likelihood estimation involving the saturated model can be painstakingly slow, if not virtually impossible, because of memory issues.

A solution proposed by de la Torre and Douglas (2004) to reduce the computational burden associated with the estimation of the DINA model parameters involves a modification of the joint distribution of the attributes. Their solution stems from the observation that the components of $\boldsymbol{\alpha}$, which can represent knowledge states, may be associated with a notion of general intelligence. In addition, many of the tests used for cognitive diagnosis purposes can be viewed as measuring a small number of general abilities. Thus, the authors proposed a joint distribution of the attributes where the components of $\boldsymbol{\alpha}$ are assumed to be conditionally independent given a general higher-order latent trait θ. One formulation that they used for the probability of distribution of $\boldsymbol{\alpha}$ conditional on θ is

$$P(\boldsymbol{\alpha}|\theta) = \prod_{k=1}^{K} P(\alpha_k|\theta) = \prod_{k=1}^{K} \frac{\exp(\lambda_{0k} + \lambda_1\theta)}{1 + \exp(\lambda_{0k} + \lambda_1\theta)}, \tag{6}$$

with the assumption $\theta \sim N(0, 1)$. That is, the logit of $P(\alpha_k|\theta)$ is expressed as a linear function of θ. This formulation is analogous to modeling the probability of a correct response on an test item using the one-parameter logistic model. By definition, $\lambda_1 > 0$; thus, $P(\alpha_k = 1|\theta)$. As such, the probability that the attribute is present given the latent trait increases with θ.

In educational applications, the higher-order latent trait θ can be interpreted as a broadly defined general proficiency or overall aptitude in a particular domain. In the example above, a student's overall proficiency in fraction subtraction can correspond to θ. Under this setup, students with higher proficiency are also expected to have greater likelihood of mastering the five skills in this domain.

A higher-order latent trait in conjunction with the DINA model yields a model referred to as the higher-order DINA (HO-DINA) model. This formulation significantly reduces the computational complexity of the problem: Instead of estimating $2^K - 1$ parameters using the saturated model, only $K + 1$ parameters (K intercept and 1 slope parameters) are involved; that is, the number of parameters grows linearly, not exponentially, with K under this specification of the joint distribution of the attributes. Although the number of parameters is more manageable using the HO-DINA model, the EM algorithm does not readily lend itself to this formulation. Hence, MCMC has been used to estimate the parameters of the HO-DINA model. De la Torre and Douglas (2004) have shown that the MCMC algorithm that they proposed for the HO-DINA model can provide reliable parameter estimates.

TABLE 1
Q-Matrix for the Simulated Data

	Attribute						Attribute				
Item	1	2	3	4	5	Item	1	2	3	4	5
1	1	0	0	0	0	16	0	1	0	1	0
2	0	1	0	0	0	17	0	1	0	0	1
3	0	0	1	0	0	18	0	0	1	1	0
4	0	0	0	1	0	19	0	0	1	0	1
5	0	0	0	0	1	20	0	0	0	1	1
6	1	0	0	0	0	21	1	1	1	0	0
7	0	1	0	0	0	22	1	1	0	1	0
8	0	0	1	0	0	23	1	1	0	0	1
9	0	0	0	1	0	24	1	0	1	1	0
10	0	0	0	0	1	25	1	0	1	0	1
11	1	1	0	0	0	26	1	0	0	1	1
12	1	0	1	0	0	27	0	1	1	1	0
13	1	0	0	1	0	28	0	1	1	0	1
14	1	0	0	0	1	29	0	1	0	1	1
15	0	1	1	0	0	30	0	0	1	1	1

3. Examples

3.1. Simulated Data

Because the viability of the MCMC algorithm for the HO-DINA model has already been documented by de la Torre and Douglas (2004), this section focuses on the feasibility of an EM algorithm for the DINA model. A computer program based on the algorithm described in the appendix was written in Ox (Doornik, 2002), an object-oriented mathematical programming language. The console version of Ox can be downloaded free for academic research and teaching purposes, whereas the EM code can be made available by request.

The simulation data employed 2,000 examinees, 30 items, and 5 attributes, with all the slip and guessing parameters equal to 0.2. The Q-matrix for this data is given in Table 1. This setup is similar to that used by de la Torre and Douglas (2004), except that the joint distribution of the skills patterns were generated equiprobably from a multinomial distribution. One hundred data sets were simulated and analyzed. On a desktop computer with 3.0 GHz processor and 1 GB of memory, the estimation took an average of fewer than 18 seconds to run when the convergence criterion (i.e., maximum difference between previous and current parameter estimates) was pegged at 0.0001.

Table 2 shows the results for the simulated data, and it gives the mean estimate, root of the mean squared error, and the empirical standard error across the 100 replications. The results indicate that the algorithm can provide accurate

TABLE 2
Results of the Simulated Data Analysis

	Mean Estimate		Standard Error			
			Mean		Empirical	
Item	*g*	*s*	*g*	*s*	*g*	*s*
1	0.20	0.20	0.015	0.015	0.014	0.016
2	0.20	0.20	0.015	0.015	0.015	0.015
3	0.20	0.20	0.016	0.016	0.015	0.013
4	0.20	0.20	0.015	0.015	0.015	0.016
5	0.20	0.20	0.015	0.015	0.015	0.014
6	0.20	0.20	0.015	0.015	0.017	0.017
7	0.20	0.20	0.015	0.015	0.014	0.016
8	0.20	0.20	0.016	0.016	0.016	0.014
9	0.20	0.20	0.015	0.015	0.015	0.016
10	0.20	0.20	0.016	0.015	0.015	0.015
11	0.20	0.20	0.011	0.022	0.012	0.021
12	0.20	0.20	0.011	0.022	0.012	0.026
13	0.20	0.20	0.011	0.022	0.011	0.022
14	0.20	0.20	0.011	0.022	0.012	0.021
15	0.20	0.20	0.011	0.022	0.011	0.021
16	0.20	0.20	0.011	0.022	0.011	0.022
17	0.20	0.20	0.011	0.022	0.010	0.021
18	0.20	0.20	0.011	0.022	0.012	0.025
19	0.20	0.20	0.011	0.022	0.012	0.024
20	0.20	0.20	0.011	0.022	0.012	0.021
21	0.20	0.20	0.010	0.030	0.009	0.029
22	0.20	0.20	0.010	0.030	0.011	0.031
23	0.20	0.20	0.010	0.030	0.009	0.031
24	0.20	0.20	0.010	0.030	0.009	0.027
25	0.20	0.21	0.010	0.031	0.009	0.027
26	0.20	0.20	0.010	0.030	0.010	0.030
27	0.20	0.20	0.010	0.030	0.009	0.029
28	0.20	0.20	0.010	0.030	0.011	0.025
29	0.20	0.20	0.010	0.030	0.010	0.029
30	0.20	0.20	0.010	0.030	0.009	0.029

parameter estimates. All the parameter estimates, except that of $\hat{s}_{25} = 0.21$, were identical to the generating values of 0.20. In addition, the mean standard errors were close to the empirical standard deviations, indicating that the estimated standard errors faithfully reflect the variabilities of the parameter estimates across the samples. On the average, the estimated standard error is only about 2% more conservative than the empirical standard error.

TABLE 3
Q-Matrix for the Fraction Subtraction Data

Item		Attribute 1	2	3	4	5
1	$\frac{3}{4}-\frac{3}{8}$	1	0	0	0	0
2	$3\frac{1}{2}-2\frac{3}{2}$	1	1	1	1	0
3	$\frac{6}{7}-\frac{4}{7}$	1	0	0	0	0
4	$3-2\frac{1}{5}$	1	1	1	1	1
5	$3\frac{7}{8}-2$	0	0	1	0	0
6	$4\frac{4}{12}-2\frac{7}{12}$	1	1	1	1	0
7	$4\frac{1}{3}-2\frac{4}{3}$	1	1	1	1	0
8	$\frac{11}{8}-\frac{1}{8}$	1	1	0	0	0
9	$3\frac{4}{5}-3\frac{2}{5}$	1	0	1	0	0
10	$2-\frac{1}{3}$	1	0	1	1	1
11	$4\frac{5}{7}-1\frac{4}{7}$	1	0	1	0	0
12	$7\frac{3}{5}-\frac{4}{5}$	1	0	1	1	0
13	$4\frac{1}{10}-2\frac{8}{10}$	1	1	1	1	0
14	$4-1\frac{4}{3}$	1	1	1	1	1
15	$4\frac{1}{3}-1\frac{5}{3}$	1	1	1	1	0

3.2. Fraction Subtraction Data

The data analyzed in this article are responses of 2,144 middle school students to 15 fraction subtraction items measuring the five skills listed above. The data were originally described and used by K. Tatsuoka (1990) and more recently by C. Tatsuoka (2002) and de la Torre and Douglas (2004). EM and MCMC item parameter estimates were obtained for the DINA and HO-DINA models, respectively. For MCMC, the parameter estimates and the standard errors were obtained by computing the posterior means and standard deviations. Table 3 gives the fraction subtraction items and the Q-matrix, and Table 4 offers the EM and MCMC estimates.

Note that the two algorithms cannot be expected to provide identical results because of two important differences. First, the joint distribution of the skills in the DINA model is based on a multinomial distribution, whereas that in the HO-DINA is based on a higher-order latent proficiency. Second, the DINA estimates were based on the mode (i.e., maximum), whereas the HO-DINA estimates were based on the mean (expected value). Nonetheless, Table 4 shows that the algorithms for the two formulations of the DINA model provide markably similar estimates, with the exception of $\hat{g}_5$, $se(\hat{g}_1)$, and $se(\hat{g}_5)$. This indicates that the use of higher-order proficiency to constrain the joint distribution of the skills is reasonable for these data. In this example, students with Attributes 1, 3, and 4 can answer Item 12, $7\frac{3}{5}-\frac{4}{5}$, 87% of the time. In contrast,

TABLE 4
Results of the Fraction Subtraction Data Analysis (Standard Error in Parentheses)

	DINA				Higher-Order DINA			
Item	$\hat{g}$		$\hat{s}$		$\hat{g}$		$\hat{s}$	
1	0.00	(0.050)	0.28	(0.013)	0.00	(0.004)	0.28	(0.012)
2	0.21	(0.013)	0.12	(0.011)	0.21	(0.012)	0.12	(0.010)
3	0.13	(0.023)	0.04	(0.005)	0.13	(0.027)	0.04	(0.005)
4	0.12	(0.011)	0.13	(0.014)	0.13	(0.009)	0.13	(0.015)
5	0.30	(0.025)	0.25	(0.012)	0.23	(0.035)	0.25	(0.011)
6	0.03	(0.006)	0.23	(0.014)	0.03	(0.006)	0.23	(0.014)
7	0.07	(0.008)	0.08	(0.009)	0.07	(0.008)	0.08	(0.009)
8	0.15	(0.020)	0.05	(0.007)	0.15	(0.022)	0.05	(0.007)
9	0.08	(0.016)	0.06	(0.007)	0.09	(0.018)	0.06	(0.007)
10	0.17	(0.013)	0.07	(0.010)	0.17	(0.012)	0.07	(0.010)
11	0.10	(0.017)	0.11	(0.009)	0.11	(0.017)	0.11	(0.009)
12	0.03	(0.006)	0.13	(0.012)	0.04	(0.007)	0.13	(0.011)
13	0.13	(0.012)	0.16	(0.012)	0.14	(0.010)	0.16	(0.012)
14	0.02	(0.005)	0.20	(0.016)	0.02	(0.005)	0.20	(0.016)
15	0.01	(0.003)	0.18	(0.013)	0.01	(0.004)	0.18	(0.013)

students who lack at least one of these attributes can answer the item correctly only about 3% of the time.

4. Discussion

The article provides a detailed presentation of the DINA model, one of the most tractable and interpretable CDMs. In addition, it discusses two approaches in estimating its parameters. Results of simulation studies indicate that the EM algorithm and the MCMC algorithm discussed here and in de la Torre and Douglas (2004), respectively, can be used to obtain accurate parameter estimates of the model. Moreover, the computer codes developed based on these algorithms can be obtained and implemented free of charge. Thus, this research provides researchers with tools that can allow them to explore the practicability of the DINA model, which can in turn pave the way for the applications of CDMs in practical education settings to inform instruction and learning.

This article can be viewed as a first step in studying other issues concerning the estimation of the DINA model parameters, particularly, those using the EM algorithm. As noted earlier, this algorithm may become computationally expensive to implement with relatively large K. In addition to the HO-DINA model, another solution is to constrain the number of permissible skills patterns using theoretically based hierarchical skills structures, such as those described by Leighton, Gierl, and Hunka (2004). When such structures are applicable, the computational demand in the expectation cycle of the EM algorithm can be reduced dramatically by properly modifying the algorithm described in the appendix. For example, when the attributes have a linear structure, the number of permissible skills patterns reduces to $K+1$.

Another issue pertains to the use of a fixed distribution for $p(\boldsymbol{\alpha})$. That is, the same multinomial probabilities are used for all the expectation cycles of the algorithm. However, using empirical Bayes methods (Bradley & Louis, 2000), these probabilities can be updated in each iteration to more closely reflect the characteristics of the observed data. Replacing the fixed prior distribution with the empirical distribution is an option that can be found in calibration software for traditional IRMs (e.g., BILOG-MG; Zimowski et al., 1996) and provide more accurate estimates in situations where the assumed examinee distribution is markedly different from that of the examinee population in consideration. The EM algorithm in this article can be easily updated to accommodate such an option.

Finally, as in traditional IRMs, item parameter calibration is deemed only the first step in the application of CDMs. For CDMs to have greater impact, they should be able to provide information about students' knowledge states (i.e., which skills they have or have not mastered). However, classification of examinees involves myriad issues, such as attribute pattern identifiability, methods of classification (maximum likelihood estimation or expected a posteriori), test length requirement, and Q-matrix specification. To address these issues thoroughly, examinee classification in the context of the DINA model merits a separate and systematic study.

Appendix
An Expectation-Maximization Algorithm for the DINA Model

A1. Parameter Estimation

The probability of a correct response in Equation 2 can be reexpressed as

$$P_j(\boldsymbol{\alpha}_l) = \begin{cases} g_j & \text{if } \boldsymbol{\alpha}_l'\boldsymbol{q}_j < \boldsymbol{q}_j'\boldsymbol{q}_j \\ 1-s_j & \text{if } \boldsymbol{\alpha}_l'\boldsymbol{q}_j = \boldsymbol{q}_j'\mathbf{q}_j \end{cases}, \tag{A1}$$

where $\boldsymbol{q}_j$ is the transposed jth row of the Q-matrix. That is, $\eta_{lj}=0$ when $\boldsymbol{\alpha}_l'\boldsymbol{q}_j < \boldsymbol{q}_j'\boldsymbol{q}_j$ and $\eta_{lj}=1$ when $\boldsymbol{\alpha}_l'\boldsymbol{q}_j = \boldsymbol{q}_j'\boldsymbol{q}_j$.

To obtain the maximum likelihood estimate of $\beta_{j\eta}$, where $\beta_{j0}=g_j$ and $\beta_{j1}=s_j$, maximize

$$l(\boldsymbol{X}) = \log \prod_{i=1}^{I} L(\boldsymbol{X}_i) = \sum_{i=1}^{I} \log L(\boldsymbol{X}_i) \tag{A2}$$

with respect to $\beta_{j\eta}$:

$$\begin{aligned}\frac{\partial l(\boldsymbol{X})}{\partial \beta_{j\eta}} &= \sum_{i=1}^{I} \frac{\partial L(\boldsymbol{X}_i)}{\partial \beta_{j\eta}} / L(\boldsymbol{X}_i) \\ &= \sum_{i=1}^{I} \frac{1}{L(\boldsymbol{X}_i)} \sum_{l=1}^{L} p(\boldsymbol{\alpha}_l) \frac{\partial L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l)}{\partial \beta_{jn}}.\end{aligned} \tag{A3}$$

Now,

$$\frac{\partial L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l)}{\partial \beta_j \eta} = \prod_{j' \neq j} P_{j'}(\boldsymbol{\alpha}_l)^{X_{ij'}} [1 - P_{j'}(\boldsymbol{\alpha}_l)]^{1 - X_{ij'}} \frac{\partial P_j(\boldsymbol{\alpha}_l)^{X_{ij}} [1 - P_j(\boldsymbol{\alpha}_l)]^{1 - X_{ij}}}{\partial \beta_{j\eta}}. \tag{A4}$$

The derivative in the right-hand side of Equaton A4 is equal to

$$\begin{aligned}&[1 - P_j(\boldsymbol{\alpha}_l)]^{1 - X_{ij}} X_{ij} P_j(\boldsymbol{\alpha}_l)^{X_{ij}-1} \frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} + P_j(\boldsymbol{\alpha}_l)^{X_{ij}} (1 - X_{ij}) [1 - P_j(\boldsymbol{\alpha}_l)]^{1-X_{ij}-1} \frac{-\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} \\ &= P_j(\boldsymbol{\alpha}_l)^{X_{ij}} [1 - P_j(\boldsymbol{\alpha}_l)]^{1-X_{ij}} \frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} \left[\frac{X_{ij}}{P_j(\boldsymbol{\alpha}_l)} - \frac{1 - X_{ij}}{1 - P_j(\boldsymbol{\alpha}_l)}\right] \\ &= P_j(\boldsymbol{\alpha}_l)^{X_{ij}} [1 - P_j(\boldsymbol{\alpha}_l)]^{1-X_{ij}} \frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} \left[\frac{X_{ij} - P_j(\boldsymbol{\alpha}_l)}{P_j(\boldsymbol{\alpha}_l)(1 - P_j(\boldsymbol{\alpha}_l))}\right].\end{aligned} \tag{A5}$$

Substituting Equation A5 to the derivative in the right-hand side of Equation A4 will give us

$$\begin{aligned}\frac{\partial L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} &= \left[\prod_{j=1}^{J} P_j(\boldsymbol{\alpha}_l)^{X_{ij}} [1 - P_j(\boldsymbol{\alpha}_l)]^{1 - X_{ij}}\right] \frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} \left[\frac{X_{ij} - P_j(\boldsymbol{\alpha}_l)}{P_j(\boldsymbol{\alpha}_l)[1 - P_j(\boldsymbol{\alpha}_l)]}\right] \\ &= L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l) \frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} \left[\frac{X_{ij} - P_j(\boldsymbol{\alpha}_l)}{P_j(\boldsymbol{\alpha}_l)[1 - P_j(\boldsymbol{\alpha}_l)]}\right].\end{aligned} \tag{A6}$$

By substituting Equation A6 and interchanging the summations, Equation A3 can be written as

$$\begin{aligned}\frac{\partial l(\boldsymbol{X})}{\partial \beta_{j\eta}} &= \sum_{l=i}^{L} \frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} \left[\frac{1}{P_j(\boldsymbol{\alpha}_l)[1 - P_j(\boldsymbol{\alpha}_l)]}\right] \sum_{i=1}^{I} \frac{L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l) p(\boldsymbol{\alpha}_l)}{L(\boldsymbol{X}_i)} [X_{ij} - p_j(\boldsymbol{\alpha}_l)] \\ &= \sum_{l=i}^{L} \frac{P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} \left[\frac{1}{P_j(\boldsymbol{\alpha}_l)[1 - P_j(\boldsymbol{\alpha}_l)]}\right] \sum_{i=1}^{I} p(\boldsymbol{\alpha}_l|\boldsymbol{X}_i)[X_{ij} - P_j(\boldsymbol{\alpha}_l)]\end{aligned}$$

$$
\begin{aligned}
&= \sum_{l=i}^{L} \frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} \left[\frac{1}{P_j(\boldsymbol{\alpha}_l)[1-P_j(\boldsymbol{\alpha}_l)]}\right] \left[\sum_{i=1}^{I} p(\boldsymbol{\alpha}_l|\boldsymbol{X}_i)X_{ij} - P_j(\boldsymbol{\alpha}_l)\sum_{i=1}^{I} P(\boldsymbol{\alpha}_l|\boldsymbol{X}_i)\right] \\
&= \sum_{l=i}^{L} \frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} \left[\frac{1}{P_j(\boldsymbol{\alpha}_l)[1-P_j(\boldsymbol{\alpha}_l)]}\right] \left[R_{jl} - P_j(\boldsymbol{\alpha}_l)I_l\right],
\end{aligned}
\tag{A7}
$$

where $p(\boldsymbol{\alpha}_l|\ \boldsymbol{X}_i)$ is the posterior probability that examinee i has the attribute pattern $\boldsymbol{\alpha}_l$, $I_l = \sum_{i=1}^{I} P(\boldsymbol{\alpha}_l|\boldsymbol{X}_i)$ is the expected number of examinees with attribute pattern $\boldsymbol{\alpha}_l$, and $R_{jl} = \sum_{i=1}^{I} P(\boldsymbol{\alpha}_l|\boldsymbol{X}_i)X_{ij}$ is the expected number of examinees with attribute pattern $\boldsymbol{\alpha}_l$ answering item j correctly.

For item j, Equation A7 can be written as

$$
\begin{aligned}
\frac{\partial l(X)}{\partial \beta_{j\eta}} &= \sum_{\{\boldsymbol{\alpha}_l:\boldsymbol{\alpha}_l'\boldsymbol{q}_j < \boldsymbol{q}_j'\boldsymbol{q}_j\}} \frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} \left[\frac{1}{P_j(\boldsymbol{\alpha}_l)[1-P_j(\boldsymbol{\alpha}_l)]}\right] [R_{jl} - P_j(\boldsymbol{\alpha}_l)I_l] \\
&\quad + \sum_{\{\boldsymbol{\alpha}_l:\boldsymbol{\alpha}_l'\boldsymbol{q}_j = \boldsymbol{q}_j'\boldsymbol{q}_j\}} \frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}} \left[\frac{1}{P_j(\boldsymbol{\alpha}_l)[1-P_j(\boldsymbol{\alpha}_l)]}\right] [R_{jl} - P_j(\boldsymbol{\alpha}_l)I_l] \\
&= \frac{\partial g_j}{\partial \beta_{j\eta}} \left[\frac{1}{g_j[1-g_j]}\right] \sum_{\{\boldsymbol{\alpha}_l:\boldsymbol{\alpha}_l'\boldsymbol{q}_j < \boldsymbol{q}_j'\boldsymbol{q}_j\}} [R_{jl} - g_j I_l] \\
&\quad + \frac{\partial (1-s_j)}{\partial \beta_{j\eta}} \left[\frac{1}{(1-s_j)s_j}\right] \sum_{\{\boldsymbol{\alpha}_l:\boldsymbol{\alpha}_l'\boldsymbol{q}_j = \boldsymbol{q}_j'\boldsymbol{q}_j\}} [R_{jl} - (1-s_j)I_l]
\end{aligned}
\tag{A8}
$$

$$
\begin{aligned}
&= \frac{\partial g_j}{\partial \beta_{j\eta}} \left[\frac{1}{g_j[1-g_j]}\right] \left[R_{jl}^{(0)} - g_j I_{jl}^{(0)}\right] \\
&\quad + \frac{\partial (1-s_j)}{\partial \beta_{j\eta}} \left[\frac{1}{(1-s_j)s_j}\right] \left[R_{jl}^{(1)} - (1-s_j) I_{jl}^{(1)}\right],
\end{aligned}
\tag{A9}
$$

where $I_{jl}^{(0)}$ is the expected number of examinees lacking at least one of the required attributes for item j and where $R_{jl}^{(0)}$ is the expected number of examinees among $I_{jl}^{(0)}$ correctly answering item j. $I_{jl}^{(1)}$ and $R_{jl}^{(1)}$ have the same interpretation except that they pertain to the examinees with all the required attributes for item j. $I_{jl}^{(0)} + I_{jl}^{(1)}$ is equal to I_l for all j.

When $\eta = 0$ (i.e., $\beta_{j0} = g$), $\partial P_j(\boldsymbol{\alpha}_l)/\partial \beta_{j\eta}$ is 1 for the first term of Equation A8 and 0 for the second term. Thus, maximization of $\partial l(\boldsymbol{X})$ with respect to β_{j0} simplifies to solving for g_j in the equation

$$
\left[\frac{1}{g_j(1-g_j)}\right] \left[R_{jl}^{(0)} - g_j I_{jl}^{(0)}\right] = 0,
\tag{A10}
$$

which gives the estimator $\hat{g}_j = R_{jl}^{(0)}/I_{jl}^{(0)}$. Similarly, maximization of $\partial l(\boldsymbol{X})$ with respect to β_{j1} is equivalent to solving for s_j in the equation

$$-\left[\frac{1}{[1-s_j]s_j}\right]\left[R_{jl}^{(1)}-[1-s_j]I_{jl}^{(1)}\right]=0, \tag{A11}$$

which results in the estimator

$$\hat{s}_j=\left[I_{jl}^{(1)}-R_{jl}^{(1)}\right]/I_{jl}^{(1)}.$$

Step 1 of the algorithm starts with initial values for $\boldsymbol{g}$ and $\boldsymbol{s}$. In Step 2, $I_{jl}^{(0)}$, $R_{jl}^{(0)}$, $I_{jl}^{(1)}$ and $R_{jl}^{(1)}$ are computed based on the current values of $\boldsymbol{g}$ and $\boldsymbol{s}$. Step 3 involves finding $\boldsymbol{g}$ and $\boldsymbol{s}$ by solving Equations A10 and A11. Finally, Steps 2 and 3 are repeated until convergence.

A2. Computing the Standard Errors

The information matrix of the estimator of β is $\boldsymbol{I}(\boldsymbol{\beta})=-E[\partial^2 l(\boldsymbol{X})/\partial\boldsymbol{\beta}^2]$. The second derivative of the log-marginalized likelihood for item j with respect to the parameters $\beta_{j\eta}$ and $\beta_{j'\eta'}$ is given by

$$\begin{aligned}\frac{\partial^2 l(\boldsymbol{X})}{\partial\beta_{j\eta}\partial\beta_{j'\eta'}}&=\sum_{i=1}\left[\frac{1}{L(\boldsymbol{X}_i)}\frac{\partial L^2(\boldsymbol{X}_i)}{\partial\beta_{j\eta}\partial\beta_{j'\eta'}}-\frac{1}{L^2(X_i)}\frac{\partial L(\boldsymbol{X}_i)}{\partial\beta_{j\eta}}\frac{\partial L(\boldsymbol{X}_i)}{\partial\beta_{j'\eta'}}\right]\\&=-\sum_{i=1}^{I}\left[\frac{1}{L^2(\boldsymbol{X}_i)}\frac{\partial L(\boldsymbol{X}_i)}{\partial\beta_{j\eta}}\frac{\partial L(\boldsymbol{X}_i)}{\partial\beta_{j'\eta'}}\right].\end{aligned} \tag{A12}$$

because the first term vanishes when the expectation is taken. Taking the derivatives of Equation A12 results in

$$\begin{aligned}&-\sum_{i=1}^{I}\frac{1}{L^2(\boldsymbol{X}_i)}\left\{\sum_{l=1}^{L}p(\boldsymbol{\alpha}_l)L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l)\frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial\beta_{j\eta}}\left[\frac{X_{ij}-P_j(\boldsymbol{\alpha}_l)}{P_j(\boldsymbol{\alpha}_l)(1-P_j(\boldsymbol{\alpha}_l))}\right]\right\}\\&\times\left\{\sum_{l=1}^{L}p(\boldsymbol{\alpha}_l)L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l)\frac{\partial P_{j'}(\boldsymbol{\alpha}_l)}{\partial\beta_{j'\eta'}}\left[\frac{X_{ij'}-P_{j'}(\boldsymbol{\alpha}_l)}{P_{j'}(\boldsymbol{\alpha}_l)(1-P_j(\boldsymbol{\alpha}_l))}\right]\right\}.\end{aligned} \tag{A13}$$

Distributing $L^{-1}(\boldsymbol{X}_i)$, each factor above becomes

$$\begin{aligned}&\sum_{l=1}^{L}\frac{p(\boldsymbol{\alpha}_l)L(\boldsymbol{X}_i|\boldsymbol{\alpha}_L)}{L(\boldsymbol{X}_i)}\frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial\beta_{j\eta}}\left[\frac{X_{ij}-P_j(\boldsymbol{\alpha}_l)}{P_j(\boldsymbol{\alpha}_l)(1-P_j(\boldsymbol{\alpha}_l))}\right]\\&=\sum_{l=1}^{L}p(\boldsymbol{\alpha}_l|\boldsymbol{X}_i)\frac{\partial P_j(\boldsymbol{\alpha}_l)}{\partial\beta_{j\eta}}\left[\frac{X_{ij}-P_j(\boldsymbol{\alpha}_l)}{P_j(\boldsymbol{\alpha}_l)(1-P_j(\boldsymbol{\alpha}_l))}\right]\\&=p_j(\eta|\boldsymbol{X}_i)\frac{\partial P_j(\eta)}{\partial\beta_{j\eta}}\left[\frac{X_{ij}-P_j(\eta)}{P_j(\eta)(1-P_j(\eta))}\right],\end{aligned} \tag{A14}$$

where $p_j(\eta|\boldsymbol{X}_i) = \sum_{\{\boldsymbol{\alpha}_l:\eta_{lj}=\eta\}} p(\boldsymbol{\alpha}_l|\boldsymbol{X}_i)$ and $P_j(\eta) = P(X_j = 1|\eta_{lj} = \eta)$.

Therefore,

$$\frac{\partial^2 l(\boldsymbol{X})}{\partial\beta_{j\eta}\partial\beta_{j'\eta'}} = -\sum_{i=1}^{I}\left\{p_j(\eta|\boldsymbol{X}_i)\frac{\partial P_j(\eta)}{\partial\beta_{j\eta}}\left[\frac{X_{ij}-P_j(\eta)}{P_j(\eta)(1-P_j(\eta))}\right]\right\}\times \left\{p_{j'}(\eta'|\boldsymbol{X}_i)\frac{\partial P_{j'}(\eta')}{\partial\beta_{j'\eta'}}\left[\frac{X_{ij'}-P_{j'}(\eta')}{P_{j'}(\eta')(1-P_j(\eta'))}\right]\right\}, \tag{A15}$$

which is a sum of the products of expected values based on the examinees' posterior distributions. Instead of computing the expectation, the information matrix can be approximated by evaluating Equation A15 at $\hat{\boldsymbol{\beta}}$ using the observed $\boldsymbol{X}$ resulting in $I(\hat{\boldsymbol{\beta}})$. Finally, $I^{-1}(\hat{\boldsymbol{\beta}})$ provides an approximation of $\mathrm{Cov}(\hat{\boldsymbol{\beta}})$, and the root of its diagonal elements represents the $\mathrm{SE}(\hat{\boldsymbol{\beta}})$.

References

Baker, F. B. (1992). *Item response theory: Parameter estimation techniques*. New York: Dekker.

Bradley, C. P., & Louis, T. A. (2000). *Bayes and empirical Bayes method for data analysis* (2nd ed.). London: Chapman & Hall.

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. (2005, April). *Modeling multiple strategies in cognitive diagnosis*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montréal, Québec, Canada.

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

Doornik, J. A. (2002). Object-Oriented Matrix Programming Using Ox (Version 3.1) [Computer software]. London: Timberlake Consultants Press.

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

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

Fu, J., & Li, Y. (2007, April). *An integrative review of cognitively diagnostic psychometric models*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.

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

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.

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, 33,* 379–416.

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

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika, 16,* 661–679.

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. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics, 12,* 55–73.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Safto (Eds.), *Monitoring skills and knowledge acquisition* (pp. 453–488). Hillsdale, NJ: Erlbaum.

van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). *BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items*. Chicago: Scientific Software.

Author

JIMMY DE LA TORRE is Assistant Professor, Department of Educational Psychology at Rutgers Graduate School of Education, 10 Seminary Place, New Brunswick, NJ 08901; j.delatorre@rutgers.edu. His research interests include psychometrics, item response models, cognitive diagnosis, and the use of assessments to inform classroom instruction and learning.

Manuscript received October 26, 2006
Accepted September 12, 2007