PSYCHOMETRIKA—VOL. 69, NO. 3, 333–353
SEPTEMBER 2004

HIGHER-ORDER LATENT TRAIT MODELS FOR COGNITIVE DIAGNOSIS

JIMMY DE LA TORRE

RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY

JEFFREY A. DOUGLAS

UNIVERSITY OF ILLINOIS

Higher-order latent traits are proposed for specifying the joint distribution of binary attributes in models for cognitive diagnosis. This approach results in a parsimonious model for the joint distribution of a high-dimensional attribute vector that is natural in many situations when specific cognitive information is sought but a less informative item response model would be a reasonable alternative. This approach stems from viewing the attributes as the specific knowledge required for examination performance, and modeling these attributes as arising from a broadly-defined latent trait resembling the $\boldsymbol{\theta}$ of item response models. In this way a relatively simple model for the joint distribution of the attributes results, which is based on a plausible model for the relationship between general aptitude and specific knowledge. Markov chain Monte Carlo algorithms for parameter estimation are given for selected response distributions, and simulation results are presented to examine the performance of the algorithm as well as the sensitivity of classification to model misspecification. An analysis of fraction subtraction data is provided as an example.

Key words: cognitive diagnosis, item response theory, latent class model, Markov chain Monte Carlo.

1. Introduction

The introduction of multidimensional latent variable models for cognitive diagnosis has given hope that tests might reveal information with more diagnostic value than can possibly be revealed by the unidimensional latent trait of standard item response models. In these models mastery of particular skills or states of knowledge can be represented by a vector of binary latent variables, indicating mastery of each of a finite set of skills under diagnosis. A generic term for a psychological construct which might be a skill or knowledge state is *attribute*, but we may use more descriptive terms in the context of particular examples. The primary objective of cognitive diagnosis is to classify examinees into latent classes determined by vectors of binary skill indicators, and in the language of more general latent class modeling, models for doing this are called *multiple classification latent class models* (Maris, 1999).

The utility of cognitive diagnosis in settings where unidimensional item response modeling has been traditionally used is seen in Tatsuoka (1995), in which a test of fraction addition as well as an SAT mathematics examination are considered for cognitive diagnosis. Another example is given in Mislevy (1996), where probability models are developed for diagnosing mastery of seven rules required for mixed-number subtraction. In cases like these, much of the dependence in the items can be explained by a single continuous and broadly-defined latent trait. However, attributes that are related to this trait but have more specific interpretations can be used to achieve an even more precise fit and afford practitioners with the opportunity to diagnose subjects in a way that leads to tailored remediation.

This research was funded by National Institute of Health grant R01 CA81068. We would like to thank William Stout and Sarah Hartz for many useful discussions, three anonymous reviewers for helpful comments and suggestions, and Kikumi Tatsuoka and Curtis Tatsuoka for generously sharing data.

Correspondence should be sent to Jimmy de la Torre, Department of Educational Psychology, Rutgers, The State University of New Jersey, 10 Seminary Place, New Brunswick, NJ 08901, USA. E-Mail: jdelator@rci.rutgers.edu.

0033-3123/2004-3/2002-1010-A $00.75/0

The methods proposed here aim to address a difficult issue in fitting models for cognitive diagnosis with many attributes, when item response models appear to be a reasonable but less informative alternative. This is accomplished by viewing the attributes as the specific knowledge required for examination performance, and modeling these attributes as arising from a broadly-defined latent trait resembling the $\boldsymbol{\theta}$ of item response models. In this way we can specify a relatively simple model for the joint distribution of the attributes that is based on a plausible model for the relationship between general aptitude and specific knowledge. A potentially useful by-product of this is that attribute classification and estimation of general aptitude can be offered by the same analysis, in a single consistent model.

The primary aims of what follows are to propose a method for modeling the joint distribution of a latent attribute vector based on higher-order latent traits, and present a Markov chain Monte Carlo algorithm for parameter estimation. The model is motivated by the need for a relatively simple formulation of the joint distribution in settings where the notion of higher-order latent traits representing constructs of general aptitude defined more broadly than the specific attributes in the cognitive diagnosis model appear natural.

A secondary aim of this paper is to examine the sensitivity of correct classification rates to the correct specification of the model. Theories for responses in cognitive diagnosis often stem from imagining a sequence of latent responses to subtasks that must all be correct in order to correctly answer the item (Embretson, 1984, 1997; Maris, 1999). However, it is conceivable that competing models using the same list of attributes but derived from different cognitive theories may perform equally well, provided the distance between the models is not too great. This will be examined in simulation.

In the next section latent class models for cognitive diagnosis are discussed, and a method for parametrizing the joint distribution of K binary latent class indicators is proposed. The third section describes Markov chain Monte Carlo procedures for parameter estimation, which are applied in a simulation study in the fourth section where model fit using both correct and incorrect models is examined, with an aim of considering the robustness of subject classification when the model is not consistent with the cognitive theory for response generation. Section five provides an analysis of fraction subtraction data, and compares several models using various criteria and measures of model fit. Concluding remarks are given in the final section.

2. Model Specification

Let $\mathbf{Y}$ denote a vector of dichotomous item responses for J items. The components of $\mathbf{Y}$ are modeled as statistically independent given the attribute vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)'$. The kth element, α_k, of $\boldsymbol{\alpha}$ is a binary indicator of a subject's classification with regard to the kth attribute. For instance, in education α_k might indicate whether a subject has mastered a particular cognitive task or state of knowledge, such as converting a whole number to a fraction. In psychiatry it might be used to indicate a positive diagnosis for the kth of K psychiatric disorders under evaluation. To completely specify a latent variable model for $\mathbf{Y}$ we need to formulate the conditional distribution of $\mathbf{Y}$ given an attribute pattern $\boldsymbol{\alpha}$, as well as parametrize the joint distribution of $\boldsymbol{\alpha}$. First, we review selected models for the conditional distribution of $\mathbf{Y}$. Using the terminology of Junker and Sijstma (2001), we review the DINA and NIDA models as examples of conjunctive models, and a model with linear logistic item response functions is included as an example of a compensatory model.

2.1. Conditional Distribution of the Item Response Vector

In the context of cognitive diagnosis, many models have been proposed for relating the distribution of $\mathbf{Y}$ to the attribute vector $\boldsymbol{\alpha}$. Although the parametric forms differ, the universal

simplifying assumption is conditional independence. For a response pattern $\mathbf{y}$, the conditional distribution is given by

$$P(\mathbf{y} \mid \boldsymbol{\alpha}) = \prod_{j=1}^{J} P(y_i \mid \boldsymbol{\alpha}).$$

All of the models for the item response functions (IRFs) $P(y_j \mid \boldsymbol{\alpha})$ that we consider require construction of a $\mathbf{Q}$-matrix (Tatsuoka, 1985), which is a matrix that indicates which attributes are needed for each item. $\mathbf{Q}$ is a $J \times K$ matrix with j, k entry $q_{jk} = 1$ if the correct application of attribute k influences the probability of correctly answering the jth item, and equals 0 otherwise. Several useful models of this sort are discussed below.

2.1.1. DINA Model

The deterministic inputs, noisy "and" gate (DINA) model is an example of a stochastic conjunctive model. It is conjunctive in the sense that all attributes specified in $\mathbf{Q}$ for an item are required, and having only a fraction of them results in a success probability equal to that of a subject possessing none of the attributes. The stochastic element of the model is that having all of these attributes does not guarantee a correct response, and lacking all of them does not guarantee an incorrect response. The deterministic aspect of the model pertains to the generation of a latent response η_{ij}, which is precisely determined by $\boldsymbol{\alpha}_i$, the attribute vector for the ith subject, and $\mathbf{q}_j$, the row of $\mathbf{Q}$ that corresponds to the jth item through the equation $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$.

The deterministic latent response η_{ij} indicates whether or not subject i possesses all of the attributes required for item j. The parameters for a correct response to item j are denoted by s_j and g_j. The parameter s_j refers to the probability of slipping and incorrectly answering the item when $\eta_{ij} = 1$, and g_j is the probability of correctly guessing the answer when $\eta_{ij} = 0$. Maris (1999) alternatively describes g_j as the probability of successfully relying on other mental resources. The parameters s_j and g_j are formally defined by

$$s_j = P(Y_{ij} = 0 \mid \eta_{ij} = 1) \quad \text{and} \quad g_j = P(Y_{ij} = 1 \mid \eta_{ij} = 0).$$

The IRF is then

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

Assuming conditional independence as well as independence among subjects, the joint likelihood function of the DINA model is

$$L(\mathbf{s}, \mathbf{g}; \boldsymbol{\alpha}) = \prod_{i=1}^{N} \prod_{j=1}^{J} \left[s_j^{1-y_{ij}} (1 - s_j)^{y_{ij}} \right]^{\eta_{ij}} \left[g_j^{y_{ij}} (1 - g_j)^{1-y_{ij}} \right]^{1-\eta_{ij}}.$$

The conditional distribution of Y_{ij} depends on $\boldsymbol{\alpha}_i$ only through η_{ij}. Because of this reduction, many different attribute patterns may result in the same latent response. Thus, as discussed in Tatsuoka (1995) and Tatsuoka (2002), the conditional distribution of an item response variable generates equivalence classes of attribute vectors. A test design issue is constructing items so that the equivalence classes generated by the distribution of an item response vector $\mathbf{Y}$ are small, and consist of similar attribute patterns. It is often unrealistic to require equivalence classes of single attribute patterns.

The parsimonious DINA model requires only two parameters for the conditional distribution of each item, and serves as a simple and interpretable model that is appropriate when the conjunction of several equally important attributes is required, and lacking one required attribute

is the same as lacking all the required attributes. Applications of the DINA model along with MCMC algorithms for estimation are given in Junker and Sijstma (2001) and Tatsuoka (2002). The DINA model is also discussed in Macready and Dayton (1977), Haertel (1989), and Doignon and Falmagne (1999).

2.1.2. *NIDA Model*

The noisy inputs, deterministic, "and" gate (NIDA) model was introduced in Maris (1999). The NIDA model, like the DINA model, involves latent response variables determined in a conjunctive manner, but "noisy inputs" refers to the stochastic nature under which these latent responses are determined from $\boldsymbol{\alpha}$. A fundamental difference between the DINA model and the NIDA model is that DINA has item-level parameters whereas NIDA has attribute-level parameters. However, the stochastic element of the latent responses in the NIDA model may be closer to the underlying cognitive process. A thorough and more psychologically oriented discussion of multicomponent latent response models is given in Embretson (1997).

Let η_{ijk} indicate whether the ith subject correctly applied the kth attribute in completing the jth item. Again, we define probabilities of "slips" and "guesses." However, for the NIDA model they are defined at the level of the latent response variables,

$$s_k = P(\eta_{ijk} = 0 \mid \alpha_{ik} = 1, q_{jk} = 1) \quad \text{and} \quad g_k = P(\eta_{ijk} = 1 \mid \alpha_{ik} = 0, q_{jk} = 1).$$

As a technical matter $P(\eta_{ijk} = 1 \mid q_{jk} = 0)$ is set equal to 1, regardless of the value of α_{ik}. According to the model an item Y_{ij} will be correct if all of the latent responses are successful. This can be expressed by $Y_{ij} = \prod_{k=1}^{K} \eta_{ijk}$. By assuming the latent responses are independent conditional on $\boldsymbol{\alpha}_i$, the IRF has the form

$$P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i, \mathbf{s}, \mathbf{g}) = \prod_{k=1}^{K} P(\eta_{ijk} = 1 \mid \alpha_{ik}, s_k, g_k) = \prod_{k=1}^{K} \left[(1 - s_k)^{\alpha_{ik}} g_k^{1-\alpha_{ik}}\right]^{q_{jk}}.$$

By assuming conditional independence of item responses given $\boldsymbol{\alpha}$ as well as independence among subjects, the likelihood function is given by

$$L(\mathbf{s}, \mathbf{g}; \boldsymbol{\alpha}) = \prod_{i=1}^{N} \prod_{j=1}^{J} \left\{ \prod_{k=1}^{K} \left[(1 - s_k)^{\alpha_{ik}} g_k^{1-\alpha_{ik}}\right]^{q_{jk}} \right\}^{y_{ij}} \left\{ 1 - \prod_{k=1}^{K} \left[(1 - s_k)^{\alpha_{ik}} g_k^{1-\alpha_{ik}}\right]^{q_{jk}} \right\}^{1-y_{ij}}.$$

The NIDA model as presented here and in Junker and Sijstma (2001) is actually a simplification of the conjuctive model given in Maris (1999), in which $\mathbf{s}$ and $\mathbf{g}$ are allowed to vary across the items. DiBello, Stout, and Roussos (1995) developed the Unified Model, which is yet another extension of the NIDA model. In the Unified Model $\mathbf{s}$ and $\mathbf{g}$ are allowed to vary across the items, and a single continuous latent trait is incorporated into the conditional distribution as a way to account for attributes which were either purposely or accidentally omitted from the $\mathbf{Q}$-matrix. The parameters of the Unified Model as presented in DiBello et al. (1995) are not identifiable. Hartz (2002) rectified this by reparametrizing the model, which made possible an MCMC approach for parameter estimation. Another related approach is Embretson's (1997) noncompensatory multidimensional item response model for multicomponent latent responses. In this model the latent trait vector is comprised of continuous latent traits rather than binary attributes.

2.1.3. *Logistic Model*

The models given above are examples of conjunctive models, in which attaining the highest probability of a correct response for an item requires all of the attributes specified by $\mathbf{Q}$ for that

item. A *disjunctive* model differs in that possession of a subset of the attributes can completely compensate for the lack of the others. Such models can be useful and theoretically justified in instances when there are several possible strategies for solving a problem. A discussion of disjunctive models is given in Maris (1999).

Disjunctive models are closely related to *compensatory* models, in which lacking certain attributes can be compensated for by possessing other attributes. In fact, a disjunctive model is a special case of a compensatory model in which it is possible to completely compensate. The linear logistic model (LLM) is a simple compensatory model that is considered in Maris (1999) as well as in Hagenaars (1990, 1993). It is quite similar to the item-factor analysis and multidimensional item response models of Muthén (1978), Bock and Aitken (1981), and Reckase (1997). The only noteworthy distinction is that the latent variables in this model are binary rather than continuous. The IRFs for the LLM have the form

$$P(Y_{ij} = 1 \mid \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) = \frac{\exp[\beta_{0j} + \sum_{k=1}^{K} \beta_{kj}\alpha_{ik}]}{1 + \exp[\beta_{0j} + \sum_{k=1}^{K} \beta_{kj}\alpha_{ik}]},$$

where α_{ij} is the indicator of the kth attribute for the ith subject as before, and β_{kj} is the log-odds ratio for attribute k and item j.

A related unidimensional IRT model is the linear logistic test model (LLTM) (Draney, Pirolli, & Wilson, 1995; Fischer, 1995). The LLTM is a Rasch item response model that utilizes a **Q**-matrix to model how separate cognitive operations combine to influence the difficulty parameter of the model. In this manner, differences between difficulty parameters are completely due to the set of cognitive operations required by the items.

Most cognitive theories utilizing single strategies for item responses naturally lead to conjunctive and noncompensatory models rather than compensatory models. Nevertheless, the LLM is quite similar to the most common multidimensional item response models, and we consider this model in subsequent sections to investigate if it can adequately model data and classify subjects even when responses are generated from a conjunctive process.

2.2. *Joint Distribution of Attributes*

After specifying the conditional distribution of $\mathbf{Y}$ given $\boldsymbol{\alpha}$, a final step in the model is to consider the probability distribution of $\boldsymbol{\alpha}$. The saturated model for the 2^K possible values that $\boldsymbol{\alpha}$ can take requires $2^K - 1$ parameters, so some simplification might be desired when the dimension of $\boldsymbol{\alpha}$ is larger than $K = 3$ or perhaps $K = 4$.

Maris (1999) discusses several possible models for latent class membership. One of the simplest is the independence model, in which the components of $\boldsymbol{\alpha}$ are assumed to be statistically independent. This requires estimating K parameters for the joint distribution, which are the population proportions for each attribute. However, this model would not generally be plausible in the context of cognitive diagnosis when the components of $\boldsymbol{\alpha}$ can be viewed as knowledge states that may be associated with some notion of general intelligence. In a later section we will demonstrate how this assumption can result in a poorly fitting model. An alternative would be to construct a loglinear model for the distribution of $\boldsymbol{\alpha}$, which can range from a relatively parsimonious main effects model to more complicated models with any order of interactions. Yet another method is to assume that $\boldsymbol{\alpha}$ arises by dichotomizing each component of a multivariate normal variable $\boldsymbol{\alpha}^*$ (Hartz, 2002). If we assume known variances of the α_k^*'s for $k = 1, 2, \ldots, K$, there remain $K(K+1)/2$ unknown parameters to estimate, including K threshold parameters and $K(K-1)/2$ unknown terms of the tetrachoric correlation matrix.

The models we consider stem from the observation that, despite the aim of obtaining specific cognitive diagnostic information, many of the examinations used for this purpose could also be

seen as primarily measuring a small number of general abilities. Whether a diagnostic model or an IRT model is used reflects the desire for formative or summative assessment, respectively. Our approach is to combine these points of view by assuming conditional independence of $\mathbf{Y}$ given $\boldsymbol{\alpha}$, and also assuming that the components of $\boldsymbol{\alpha}$ are independent conditional on $\boldsymbol{\theta}$, a latent vector representing general ability in the studied domain.

In the context of cognitive diagnosis, Tatsuoka (1995) refers to $\boldsymbol{\alpha}$ as a *knowledge state*. Specifically, each element of $\boldsymbol{\alpha}$ is an indicator for knowledge or mastery of a very specific rule or piece of information. A model for attainment of these attributes would be to assume that their acquisition is related to one or more broadly-defined constructs of general intelligence or aptitude. Those with greater aptitude more readily acquire the specific attributes that are required for the test items. This notion essentially induces an item response model at a higher order in which the latent attributes play the role of the items and they are locally independent given the general aptitude for acquiring knowledge in this domain, which is represented by $\boldsymbol{\theta}$.

In an example of fraction subtraction given in a later section, specific rules for manipulating fraction and whole numbers and subtracting them are identified, and are used to define the attribute vector $\boldsymbol{\alpha}$. In this somewhat narrow domain it is reasonable to assume that mastery of such rules is related to a unidimensional trait θ, which might be interpreted as general arithmetic ability. In more complex settings, a multidimensional $\boldsymbol{\theta}$ might be required. In either case, the probability model for $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\theta}$ is

$$P(\boldsymbol{\alpha} \mid \boldsymbol{\theta}) = \prod_{k=1}^{K} P(\alpha_k \mid \boldsymbol{\theta}). \tag{1}$$

The particular model that we consider is a logistic regression model with latent covariate $\boldsymbol{\theta}$

$$P(\alpha_k \mid \boldsymbol{\theta}) = \frac{\exp(\lambda_{0k} + \boldsymbol{\lambda}_k' \boldsymbol{\theta})}{1 + \exp(\lambda_{0k} + \boldsymbol{\lambda}_k' \boldsymbol{\theta})}. \tag{2}$$

In many applications, such as the one given in this paper, $\boldsymbol{\theta}$ will be unidimensional and normally distributed with mean 0 and variance 1. This implies that $2K$ parameters will be required. If D is multidimensional, a structured factor loading matrix would be used, where $\boldsymbol{\lambda}_k$ denotes the factor loading vector corresponding to α_k. Just as expert opinion is used to construct $\mathbf{Q}$, expert opinion is also used in deciding for each α_k which components of $\boldsymbol{\lambda}_k$ are nonzero. Because the number of attributes would generally need to be much less than the number of items and much greater than the dimension of $\boldsymbol{\theta}$, the cases $D = 1$ and $D = 2$ would encompass the majority of applications. For this reason we have primarily focused on the case where $D = 1$, and recognize that the two-dimensional case may also be of practical value. In that case, one needs to fit the K intercepts λ_{0k}, as well as the nonzero factor loadings. With a factor loading matrix sufficiently structured to ensure identifiability, the remaining parameter to fit is ρ, which is the correlation between θ_1 and θ_2.

This hierarchy where the item responses are independent given $\boldsymbol{\alpha}$, and the components of $\boldsymbol{\alpha}$ are independent given θ, is natural in conjunctive models for cognitive diagnosis. In such models, we consider knowledge of $\boldsymbol{\alpha}$ as sufficient for determining item responses apart from random slips and guesses in a latent response model. However, the joint distribution of $\boldsymbol{\alpha}$ must be modeled, and in many cases, it may be reasonable to think of the components of $\boldsymbol{\alpha}$ as independent given a more broadly-defined ability θ.

Modeling the joint distribution of $\boldsymbol{\alpha}$ using higher-order latent traits has several advantages. It greatly reduces the complexity of the saturated model in cases where it is reasonable to view the examination as measuring one or perhaps two general abilities in addition to the specific knowledge states that comprise $\boldsymbol{\alpha}$. The linear logistic model that we have proposed is rather easy

to fit using Markov chain Monte Carlo. Finally, it enables one to classify each α_k and obtain an estimate $\hat{\boldsymbol{\theta}}$ in the same analysis. In an example to follow, we will demonstrate how this estimator correlates with ability estimates obtained from a two-parameter logistic (2PL) item response model fitted with the same data.

3. Parameter Estimation

In estimating the parameters of the models, a fully Bayesian formulation was adopted. The complexity of the joint posterior distribution (see (9) below) precluded sampling directly from the posterior distribution; hence, sampling was carried out using Markov chain Monte Carlo (MCMC) simulation. In addition, because the full conditional distributions cannot also be sampled directly, samples were iteratively drawn from these distributions using the Metropolis-Hastings algorithm (Casella & George, 1992; Chib & Greenberg, 1995; Geman & Geman, 1984; Patz & Junker, 1999a, 1999b). Parameter estimates were based on the mean of the draws of the remaining iterations after the burn-in.

3.1. The Higher-Order DINA Model

3.1.1. Prior, Joint, and Conditional Distributions

The following prior distributions for $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$, $\mathbf{g}$, and $\mathbf{s}$ are used in conjunction with the higher-order DINA model.

$$\lambda_{0k} \sim \text{Normal}(\mu_{\lambda_0}, \sigma^2_{\lambda_0}) \tag{3}$$

$$\lambda_{1k} \sim \text{Lognormal}(\mu_{\lambda_1}, \sigma^2_{\lambda_1}) \tag{4}$$

$$\theta_i \sim \text{Normal}(\mu_\theta, \sigma^2_\theta) \tag{5}$$

$$\alpha_{ik}|\theta_i, \boldsymbol{\lambda}_k \sim \text{Bernoulli}\left(\{1 + \exp(-1.7\lambda_{1k}(\theta_i - \lambda_{0k})\}^{-1}\right) \tag{6}$$

$$g_j \sim \text{4-Beta}(\upsilon_g, \omega_g, a_g, b_g) \tag{7}$$

$$1 - s_j \sim \text{4-Beta}(\upsilon_s, \omega_s, a_s, b_s) \tag{8}$$

4-Beta(υ, ω, a, b) is the four-parameter beta distribution, and for $a < x < b$ its density function is given by

$$f(x) = \frac{(x-a)^{\upsilon-1}(b-x)^{\omega-1}}{\beta(\upsilon,\omega)(b-a)^{\upsilon+\omega-1}},$$

where $\beta(\upsilon, \omega) = \int_0^1 u^{\upsilon-1}(1-u)^{\omega-1}\,du$. The functional forms of the prior distributions were chosen out of convenience, and the associated hyperparameters were selected to be reasonably vague within the range of realistic item parameters. With the large sample sizes worked with in this paper, the prior distributions have little influence.

Using the conditional independence of $\mathbf{Y}$ given $\boldsymbol{\alpha}$, and $\boldsymbol{\alpha}$ given $\boldsymbol{\theta}$, the joint posterior distribution of $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$, $\mathbf{g}$, and $\mathbf{s}$ given $\mathbf{Y}$ is

$$P(\boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{s}, \mathbf{g}|\mathbf{Y}) \propto L(\mathbf{s}, \mathbf{g}; \boldsymbol{\alpha})P(\boldsymbol{\alpha}|\boldsymbol{\lambda}, \boldsymbol{\theta})P(\boldsymbol{\lambda})P(\boldsymbol{\theta})P(\mathbf{g})P(\mathbf{s}). \tag{9}$$

Finally, the full conditional distributions of the parameters given the data and the rest of the parameters are as follows:

$$P(\boldsymbol{\lambda}|\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{s}, \mathbf{g}) \propto P(\boldsymbol{\alpha}|\boldsymbol{\lambda}, \boldsymbol{\theta})P(\boldsymbol{\lambda}) \tag{10}$$

$$P(\boldsymbol{\theta}|\mathbf{Y}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \mathbf{s}, \mathbf{g}) \propto P(\boldsymbol{\alpha}|\boldsymbol{\lambda}, \boldsymbol{\theta})P(\boldsymbol{\theta}) \tag{11}$$

$$P(\boldsymbol{\alpha}|\mathbf{Y}, \boldsymbol{\lambda}, \boldsymbol{\theta}, \mathbf{s}, \mathbf{g}) \propto L(\mathbf{s}, \mathbf{g}; \boldsymbol{\alpha})P(\boldsymbol{\alpha}|\boldsymbol{\lambda}, \boldsymbol{\theta}) \tag{12}$$

$$P(\mathbf{s}, \mathbf{g}|\mathbf{Y}, \boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\alpha}) \propto L(\mathbf{s}, \mathbf{g}; \boldsymbol{\alpha})P(\mathbf{s})P(\mathbf{g}) \tag{13}$$

3.1.2. MCMC Algorithm

Below is an outline of the MCMC algorithm used in the parameter estimation. At iteration t:

1. For $\boldsymbol{\lambda}$, draw the candidate values $\lambda_{0k}^{(*)}$ from Uniform$(\lambda_{0k}^{(t-1)} - \delta_{\lambda_0}, \lambda_{0k}^{(t-1)} + \delta_{\lambda_0})$ and $\lambda_{1k}^{(*)}$ from Uniform$(\lambda_{1k}^{(t-1)} - \delta_{\lambda_1}, \lambda_{1k}^{(t-1)} + \delta_{\lambda_1})$, and accept $\boldsymbol{\lambda}^{(*)}$ with probability

$$p(\boldsymbol{\lambda}^{(t-1)}, \boldsymbol{\lambda}^{(*)}) = \min\left\{\frac{P(\boldsymbol{\alpha}^{(t-1)}|\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\lambda}^{(*)})P(\boldsymbol{\lambda}^{(*)})}{P(\boldsymbol{\alpha}^{(t-1)}|\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\lambda}^{(t-1)})P(\boldsymbol{\lambda}^{(t-1)})}, 1\right\} \tag{14}$$

2. For $\boldsymbol{\theta}$, draw the candidate value $\theta_i^{(*)}$ from Normal$(\theta_i^{(t-1)}, \sigma^2_{C\theta})$, and accept $\boldsymbol{\theta}^{(*)}$ with probability

$$p\left(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\theta}^{(*)}\right) = \min\left\{\frac{P(\boldsymbol{\alpha}^{(t-1)}|\boldsymbol{\theta}^{(*)}, \boldsymbol{\lambda}^{(t)})P(\boldsymbol{\theta}^{(*)})}{P(\boldsymbol{\alpha}^{(t-1)}|\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\lambda}^{(t)})P(\boldsymbol{\theta}^{(t-1)})}, 1\right\} \tag{15}$$

3. For $\boldsymbol{\alpha}$, draw the candidate value $\alpha_{ik}^{(*)}$ from Bernoulli(.5), and accept $\boldsymbol{\alpha}^{(*)}$ with probability

$$p(\boldsymbol{\alpha}^{(t-1)}, \boldsymbol{\alpha}^{(*)}) = \min\left\{\frac{L(\mathbf{s}^{(t-1)}, \mathbf{g}^{(t-1)}; \boldsymbol{\alpha}^{(*)})P(\boldsymbol{\alpha}^{(*)}|\boldsymbol{\theta}^{(t)}, \boldsymbol{\lambda}^{(t)})}{L(\mathbf{s}^{(t-1)}, \mathbf{g}^{(t-1)}; \boldsymbol{\alpha}^{(t-1)})P(\boldsymbol{\alpha}^{(t-1)}|\boldsymbol{\theta}^{(t)}, \boldsymbol{\lambda}^{(t)})}, 1\right\} \tag{16}$$

4. For $\{\mathbf{g}, \mathbf{s}\}$, draw the candidate values $g_{jk}^{(*)}$ from Uniform$(g_{jk}^{(t-1)} - \delta_g, g_{jk}^{(t-1)} + \delta_g)$ and $s_{jk}^{(*)}$ from Uniform$(s_{jk}^{(t-1)} - \delta_s, g_{jk}^{(t-1)} + \delta_s)$, and accept $\{\mathbf{g}^{(*)}, \mathbf{s}^{(*)}\}$) with probability

$$p\left(\{\mathbf{g}^{(t-1)}, \mathbf{s}^{(t-1)}\}, \{\mathbf{g}^{(*)}, \mathbf{s}^{(*)}\}\right) = \min\left\{\frac{L(\mathbf{s}^{(*)}, \mathbf{g}^{(*)}; \boldsymbol{\alpha}^{(t)})P(\mathbf{s}^{(*)})P(\mathbf{g}^{(*)})}{L(\mathbf{s}^{(t-1)}, \mathbf{g}^{(t-1)}; \boldsymbol{\alpha}^{(t)})P(\mathbf{s}^{(t-1)})P(\mathbf{g}^{(t-1)})}, 1\right\} \tag{17}$$

3.2. The Higher-Order Linear Logistic Model

The same prior distributions for $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$, and $\boldsymbol{\alpha}$ as in the DINA model were used for the higher-order LLM. The prior distributions for $\boldsymbol{\beta}$ were

$$\beta_{j0} \sim 4\text{-Beta}(\upsilon_{\beta_0}, \omega_{\beta_0}, a_{\beta_0}, b_{\beta_0}) \tag{18}$$

$$\beta_{jk} \sim 4\text{-Beta}(\upsilon_{\beta_j}, \omega_{\beta_j}, a_{\beta_j}, b_{\beta_j}) \tag{19}$$

By replacing the item parameters $\{\mathbf{s}, \mathbf{g}\}$ with $\boldsymbol{\beta}$, the joint posterior and full conditional distributions of the parameters of the higher-order LLM can be expressed in the same way as (9) through (13).

Lastly, the MCMC algorithm for this model is also the same as that of the previous model except for step 4, which was carried out as

4. For $\boldsymbol{\beta}$, draw the candidate values $\beta_{j0}^{(*)}$ from Normal$(\beta_{j0}^{(t-1)}, \sigma^2_{C\beta_0})$ and β_{jk}^{*} from Normal $(\beta_{jk}^{(t-1)}, \sigma^2_{C\beta_j})$, and accept $\boldsymbol{\beta}^{(*)}$ with probability

$$p(\boldsymbol{\beta}^{(t-1)}, \boldsymbol{\beta}^{(*)}) = \min\left\{\frac{L(\boldsymbol{\beta}^{(*)}; \boldsymbol{\alpha}^{(t)})P(\boldsymbol{\beta}^{(*)})}{L(\boldsymbol{\beta}^{(t-1)}; \boldsymbol{\alpha}^{(t)})P(\boldsymbol{\beta}^{(t-1)})}, 1\right\} \tag{20}$$

4. Simulation Study

4.1. Method

To investigate how accurately the parameters of the models can be recovered using the estimation method described above, 25 data sets with five attributes, 30 items and 1000 examinees for each model were simulated. The structural parameters, namely $\boldsymbol{\lambda}$ and $\mathbf{s}$, $\mathbf{g}$, and $\boldsymbol{\beta}$, were fixed across the 25 replications. For each replication, θ_i was generated from Normal(0, 1), and α_{ik} was generated from Bernoulli($\{1 + \exp(-1.7\lambda_{1k}(\theta_i - \lambda_{0k}))\}^{-1}$). The $\mathbf{Q}$-matrix used in the simulation study can be found in Table 1. This Q-matrix was constructed such that each attribute appears alone, in a pair, or in a triple the same number of times as other attributes.

For the prior distributions of $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$, the parameters μ and σ^2 were set 0 and 1. The distributions 4-Beta(.4, 1, 2, 1), 4-Beta(0, .6, 1, 2), 4-Beta(−2.5, 0, 2, 2.5), and 4-Beta($.5/K_j$, $5/K_j$, 1, 2) (where K_j is the number of relevant attributes for item j) were used as priors of $1-\mathbf{s}$, $\mathbf{g}$, $\boldsymbol{\beta}_0$, and β_{jk}, respectively. Each chain in the simulation study was of length 5000. The draws from the first 1000 iterations were discarded, and parameter estimates were based on the draws from the last 4000 iterations. Using the convergence criterion of Gelman and Rubin (1992) as implemented in the software CODA (Best, Cowles, & Vines, 1995) we verified that a burn-in of 1000 iterations, followed by 4000 iterations, is more than sufficient. Gelman and Rubin (1992) defined an index $\hat{R}$, which uses multiple parallel Markov chains to estimate the portion of the posterior mean estimator that is due to Monte Carlo error. A rule of thumb is that $\hat{R}$ should be less than 1.2. By running five parallel chains for the first simulated data set, this criterion was satisfied for all structural parameters.

For each simulation, parameters that are common to the two models (i.e., $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$, and $\boldsymbol{\alpha}$) were estimated using both the DINA model and the LLM; the item parameters, $\mathbf{s}$, $\mathbf{g}$, and $\boldsymbol{\beta}$, were estimated using the appropriate models.

TABLE 1.
The Transposed $\mathbf{Q}$-matrix for the Simulation Study

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

	Item														
Attribute	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0
2	1	1	0	0	0	1	1	1	0	0	0	1	1	1	0
3	0	0	1	1	0	1	0	0	1	1	0	1	1	0	1
4	1	0	1	0	1	0	1	0	1	0	1	1	0	1	1
5	0	1	0	1	1	0	0	1	0	1	1	0	1	1	1

TABLE 2.
Mean and SD of $\boldsymbol{\lambda}$ Estimates over 25 Independent Replications for Data Generated Using the DINA Model

	Parameter		DINA				LLM			
Attribute	λ_0	λ_1	$\hat{\lambda}_0$	SD($\hat{\lambda}_0$)	$\hat{\lambda}_1$	SD($\hat{\lambda}_1$)	$\hat{\lambda}_0$	SD($\hat{\lambda}_0$)	$\hat{\lambda}_1$	SD($\hat{\lambda}_1$)
1	−0.95	1.34	−1.00	0.14	1.34	0.32	0.51	0.42	2.41	0.69
2	−1.42	1.22	−1.49	0.14	1.29	0.22	−0.78	0.35	1.27	0.37
3	−0.66	1.08	−0.67	0.07	1.11	0.13	−0.41	0.08	1.21	0.19
4	0.50	1.11	0.49	0.08	1.11	0.16	0.73	0.09	1.55	0.24
5	−0.05	0.97	−0.04	0.08	0.94	0.17	0.22	0.10	1.00	0.19

4.2. Results

The mean estimates and standard deviations of the estimates for $\boldsymbol{\lambda}$, the parameters which determine the population proportions for each attribute, are given in Tables 2 and 3. The results indicate that using the correct model impacts both the accuracy and the stability of the estimates. For example, in Table 2 where the data were generated using the DINA model, the estimates based on the DINA model are closer to the true values compared to estimates based on the LLM. At the same time, the standard deviations of the estimates obtained using the DINA model are smaller compared to the corresponding standard deviations using the LLM. The corresponding results were seen for data generated using the LLM given in Table 3.

The posterior mean of examinee i on attribute k, $\hat{\alpha}_{ik}$, was used in determining whether or not the examinee possesses this attribute. Only when $\hat{\alpha}_{ik} > .5$, was examinee i considered to possess attribute k. The mean proportion of attributes correctly classified by each model for all the simulated data are given in Table 4. The importance of using the correct model is again evident from these results. Although a high number of attributes were still correctly classified

TABLE 3.
Mean and SD of $\boldsymbol{\lambda}$ Estimates over 25 Independent Replications for Data Generated Using the LLM

	Parameter		DINA				LLM			
Attribute	λ_0	λ_1	$\hat{\lambda}_0$	SD($\hat{\lambda}_0$)	$\hat{\lambda}_1$	SD($\hat{\lambda}_1$)	$\hat{\lambda}_0$	SD($\hat{\lambda}_0$)	$\hat{\lambda}_1$	SD($\hat{\lambda}_1$)
1	−0.95	1.34	−1.43	0.20	1.58	0.31	−1.01	0.18	1.39	0.20
2	−1.42	1.22	−1.70	0.23	1.47	0.29	−1.36	0.21	1.19	0.22
3	−0.66	1.08	−1.15	0.15	1.42	0.25	−0.70	0.12	1.18	0.22
4	0.50	1.11	0.28	0.12	1.59	0.35	0.56	0.12	1.29	0.23
5	−0.05	0.97	−0.74	0.15	1.34	0.36	−0.07	0.12	0.98	0.19

TABLE 4.
Mean of Proportion of Correct $\boldsymbol{\alpha}$ Classification and Agreement over 25 Replications

	Generating Model									
	DINA					LLM				
Fitted Model	α_1	α_2	α_3	α_4	α_5	α_1	α_2	α_3	α_4	α_5
DINA	0.88	0.90	0.93	0.97	0.92	0.84	0.89	0.87	0.91	0.80
LLM	0.70	0.83	0.90	0.96	0.88	0.89	0.91	0.90	0.95	0.87
κ	0.44	0.66	0.85	0.93	0.83	0.70	0.76	0.81	0.86	0.69

TABLE 5.
Mean Correlation and RMSE Between θ and $\hat{\theta}$ over 25 Replications

Fitted Model	Generating Model			
	DINA		LLM	
	Correlation	RMSE	Correlation	RMSE
DINA	0.78	0.63	0.76	0.67
LLM	0.76	0.65	0.76	0.65

even with misspecified models, the proportions of attributes correctly classified using the right models were consistently higher. The κ statistic is a chance-corrected index of agreement, which is the ratio of the number of agreements minus what would be expected by chance and the expected number of disagreements due to chance (Everitt, 1998). The κ statistics in this table indicate that the agreements between the two fitted models in classifying the examinees are very high.

Two measures of fit were used to evaluate how accurately θ can be estimated: (1) the correlation between the true and estimated θ, and (2) the root mean squared error (RMSE) of the estimates from the true θ. These measures were computed for each replication, and the results in Table 5 are averages over the 25 replications. Results show that specifying the correct model resulted in better estimates (i.e., higher correlation and lower RMSE). Nevertheless, the differences are small, which may indicate that specifying the **Q**-matrix correctly is of greater importance than identifying the correct response model. However, this needs to be investigated further.

Table 6 shows that the item parameters of the DINA model can be accurately estimated using MCMC simulation. For almost all the items, the mean estimates do not deviate from the true value by more than 0.02. At the same time, the estimates across the 25 replications have small variabilities.

Similarly, the parameters of the LLM were accurately estimated by the MCMC algorithm (see Table 7). However, it can be noted that the estimates for this model were not as accurate and as stable compared to the estimates for the DINA model. This may be due to the greater number of item parameters in the LLM, and the wider range of values the parameters can assume.

5. Fraction Subtraction Test

5.1. Data

The data consist of responses to 20 items involving subtraction of fractions by 2144 examinees. They were originally used and described by Tatsuoka (1990), and were recently analyzed in Tatsuoka (2002). The eight attributes required to answer these items are: (1) Convert a whole number to a fraction, (2) Separate a whole number from a fraction, (3) Simplify before subtracting, (4) Find a common denominator, (5) Borrow from whole number part, (6) Column borrow to subtract the second numerator from the first, (7) Subtract numerators, and (8) Reduce answers to simplest form. Based on these definitions, the **Q**-matrix of attributes necessary to correctly answer each item were constructed (refer to Table 8). Because all the required attributes must be present before an examinee can correctly answer an item, a conjunctive model was appropriate for this problem. Our analysis of the data indicated that the DINA model provided a better fit compared to the NIDA model. Hence, the DINA model with a common discrimination parameter for the higher-order latent trait was used. Also, because of the complexity of the compensatory model, which requires many more parameters than the DINA model, all attempts to estimate

TABLE 6.
Mean Estimates of the Parameters of the DINA Model over 25 Replications

Item	g	$1-s$	$\hat{g}$	SD($\hat{g}$)	$1-\hat{s}$	SD($1-\hat{s}$)
1	0.35	0.67	0.33	0.04	0.67	0.02
2	0.40	0.66	0.37	0.05	0.66	0.02
3	0.13	0.67	0.12	0.03	0.67	0.02
4	0.15	0.90	0.14	0.01	0.89	0.02
5	0.29	0.65	0.29	0.02	0.65	0.02
6	0.39	0.60	0.38	0.04	0.60	0.02
7	0.10	0.61	0.11	0.05	0.61	0.02
8	0.40	0.81	0.39	0.03	0.80	0.02
9	0.15	0.74	0.15	0.02	0.74	0.02
10	0.16	0.76	0.17	0.02	0.75	0.03
11	0.38	0.73	0.38	0.04	0.73	0.02
12	0.11	0.83	0.11	0.02	0.83	0.02
13	0.26	0.89	0.26	0.02	0.89	0.01
14	0.35	0.85	0.35	0.02	0.83	0.02
15	0.13	0.87	0.15	0.02	0.87	0.02
16	0.18	0.69	0.18	0.02	0.69	0.02
17	0.26	0.75	0.26	0.02	0.75	0.02
18	0.11	0.70	0.11	0.01	0.70	0.03
19	0.37	0.80	0.37	0.02	0.80	0.02
20	0.23	0.84	0.23	0.01	0.83	0.02
21	0.38	0.85	0.37	0.02	0.85	0.02
22	0.20	0.89	0.20	0.02	0.89	0.02
23	0.23	0.63	0.24	0.02	0.64	0.03
24	0.10	0.73	0.10	0.01	0.72	0.03
25	0.30	0.72	0.29	0.02	0.72	0.03
26	0.11	0.82	0.12	0.01	0.82	0.03
27	0.11	0.73	0.11	0.01	0.73	0.03
28	0.22	0.73	0.22	0.01	0.73	0.02
29	0.12	0.79	0.13	0.01	0.79	0.02
30	0.23	0.69	0.22	0.01	0.68	0.03

parameters resulted in Markov chains for which convergence could not be obtained. This is in contrast to the simulation study, in which MCMC could be used to fit the compensatory model, even in cases where the data were generated using the DINA model. Two versions of the DINA model were used to analyze the data: the higher-order DINA model that posits a higher-order structure among the attributes, and the independence DINA model that disregards any higher-order structure. The prior distributions described in the simulation section were used for the relevant parameters of the two models.

Parameter estimates were based on averaging the estimates from 10 parallel chains with randomly chosen starting values. The squared standard errors were obtained by averaging the sample variances of the parameters from the separate chains. Each of these parallel chains was run for 20000 iterations with the first 10000 iterations as burn-in. This choice of chain length was conservative, but easily satisfied Gelman's and Rubin's rule that $\hat{R}$ be less than 1.2, with an exception of one parameter. This was a location parameter for the higher-order DINA, and had a $\hat{R}$ of approximately 1.58.

TABLE 7.
Mean Estimates of the Parameters of the LLM over 25 Replications

Item	β_0	β_1	β_2	$\hat{\beta}_0$	SD($\hat{\beta}_0$)	$\hat{\beta}_1$	SD($\hat{\beta}_1$)	$\hat{\beta}_2$	SD($\hat{\beta}_2$)
1	−0.62	1.32	–	−0.68	0.15	1.40	0.19	–	–
2	−0.43	–	1.11	−0.53	0.19	–	–	1.24	0.23
3	−1.93	–	–	−1.86	0.19	–	–	–	–
4	−1.74	–	–	−1.68	0.13	–	–	–	–
5	−0.89	–	–	−0.94	0.13	–	–	–	–
6	−0.45	0.86	–	−0.57	0.12	1.02	0.14	–	–
7	−2.15	–	2.59	−1.94	0.13	–	–	2.38	0.15
8	−0.39	–	–	−0.45	0.10	–	–	–	–
9	−1.76	–	–	−1.76	0.08	–	–	–	–
10	−1.64	–	–	−1.65	0.13	–	–	–	–
11	−0.48	1.00	0.49	−0.71	0.14	0.99	0.16	0.75	0.20
12	−2.05	1.70	–	−1.90	0.15	1.59	0.21	–	–
13	−1.06	1.67	–	−1.07	0.16	1.67	0.18	–	–
14	−0.62	1.32	–	−0.66	0.13	1.33	0.14	–	–
15	−1.87	–	1.88	−1.67	0.17	–	–	1.69	0.17
16	−1.52	–	0.92	−1.54	0.20	–	–	0.97	0.24
17	−1.04	–	0.99	−1.13	0.25	–	–	1.08	0.24
18	−2.05	–	–	−1.94	0.11	–	–	–	–
19	−0.53	–	–	−0.62	0.14	–	–	–	–
20	−1.19	–	–	−1.21	0.12	–	–	–	–
21	−0.49	0.65	0.81	−0.58	0.17	0.74	0.14	0.85	0.17
22	−1.36	1.08	1.17	−1.30	0.18	1.07	0.19	1.12	0.15
23	−1.18	0.74	0.52	−1.23	0.16	0.69	0.15	0.60	0.12
24	−2.19	1.25	–	−1.98	0.11	1.00	0.14	–	–
25	−0.86	0.77	–	−0.93	0.11	0.73	0.16	–	–
26	−2.07	1.40	–	−1.87	0.13	1.14	0.11	–	–
27	−2.06	–	1.17	−1.82	0.14	–	–	0.95	0.17
28	−1.24	–	0.77	−1.28	0.15	–	–	0.75	0.14
29	−1.95	–	1.33	−1.73	0.12	–	–	1.11	0.12
30	−1.23	–	–	−1.28	0.13	–	–	–	–

5.2. Results

The estimated posterior means and the posterior standard deviations for both the higher-order and independence model are shown in Table 9. Slip and guessing parameters of 0 represent ideal conditions under which all attributes have been identified, the **Q**-matrix has been corrected specified, and responses are deterministic. By allowing small but nonzero slip and guessing parameters, the model allows random responses. However, if the slip and guessing parameters become too large, one might suspect either the attributes are not correctly identified, or the **Q**-matrix is not correctly specified, though no formal rules exist to test these assumptions. In our analysis using the higher-order model, guessing parameters ranged from 0.00 to 0.44 with 17 out of 20 less than 0.20. Slip parameters ranged from 0.04 to 0.33 with 16 out of 20 less than or equal to 0.20. Similar results have been observed by Tatsuoka (2002).

High estimates for guesses or slips are indications of poor fit. In particular, they suggest that the posited attributes may not be sufficient in explaining the responses of the examinees, that is, a different strategy for answering the problem may exist. For example, the estimate for the guessing parameter of item 8 is 0.44 using the higher-order model and 0.47 using the independence

TABLE 7.
(*continued*)

Item	β_3	β_4	β_5	$\hat{\beta}_3$	SD($\hat{\beta}_3$)	$\hat{\beta}_4$	SD($\hat{\beta}_4$)	$\hat{\beta}_5$	SD($\hat{\beta}_5$)
1	–	–	–	–	–	–	–	–	–
2	–	–	–	–	–	–	–	–	–
3	2.62	–	–	2.53	0.21	–	–	–	–
4	–	3.95	–	–	–	3.94	0.26	–	–
5	–	–	1.52	–	–	–	–	1.60	0.19
6	–	–	–	–	–	–	–	–	–
7	–	–	–	–	–	–	–	–	–
8	1.81	–	–	1.90	0.16	–	–	–	–
9	–	2.80	–	–	–	2.81	0.15	–	–
10	–	–	2.79	–	–	–	–	2.76	0.16
11	–	–	–	–	–	–	–	–	–
12	1.95	–	–	1.85	0.15	–	–	–	–
13	–	1.43	–	–	–	1.46	0.20	–	–
14	–	–	1.01	–	–	–	–	1.05	0.17
15	1.89	–	–	1.81	0.13	–	–	–	–
16	–	1.38	–	–	–	1.36	0.16	–	–
17	–	–	1.13	–	–	–	–	1.19	0.18
18	1.15	1.76	–	1.04	0.12	1.73	0.15	–	–
19	0.97	–	0.96	1.03	0.17	–	–	1.05	0.19
20	–	1.63	1.18	–	–	1.59	0.13	1.19	0.14
21	0.75	–	–	0.74	0.12	–	–	–	–
22	–	1.18	–	–	–	1.18	0.16	–	–
23	–	–	0.46	–	–	–	–	0.53	0.11
24	0.85	1.05	–	0.82	0.14	1.03	0.18	–	–
25	0.48	–	0.55	0.56	0.11	–	–	0.60	0.12
26	–	1.23	0.96	–	–	1.19	0.12	0.98	0.17
27	0.84	1.07	–	0.80	0.14	1.04	0.14	–	–
28	0.76	–	0.71	0.81	0.15	–	–	0.77	0.15
29	–	1.10	0.85	–	–	1.05	0.13	0.88	0.11
30	0.61	0.76	0.66	0.67	0.13	0.77	0.11	0.64	0.10

model. Based on the **Q**-matrix, this item requires attribute 7, "subtract numerators," to be correctly answered. However, the high value of the estimate indicates that even examinees who do not possess this attribute have a good chance of answering the item correctly. A closer inspection of the item reveals that examinees who are familiar with the inverse property of addition but do not know fraction subtraction can still answer the problem "$\frac{2}{3} - \frac{2}{3} =$" correctly.

Although the item parameter estimates for both models are quite similar, the estimated proportion of examinees with the specific attributes were very discrepant. Table 10 shows that the independence model has higher estimates for all attributes except for attribute 8. In addition, the classification agreements of the two models for most of the attributes are low. This discrepancy will become apparent when computing the Bayes factor for the two models.

5.3. Model Fit

A method of investigating the model fit is to use the estimated parameters to predict the pairwise relationship of the items, specifically, the observed log-odds-ratio of the item-pairs. The log-odds ratio is a common measure of association for binary random variables, and is useful in this context for diagnosing the correctness of the **Q**-matrix as well as evaluating the fit of the

TABLE 8.
The Transposed **Q**-matrix for the Fraction Subtraction Data

	Item									
Attribute	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	1	0	0	0
2	0	0	0	1	1	0	1	0	1	1
3	0	0	0	1	0	0	0	0	0	0
4	1	1	1	0	1	0	0	0	0	0
5	0	0	0	1	0	0	0	0	0	1
6	1	0	0	0	0	0	0	0	0	0
7	1	1	1	1	1	1	1	1	0	1
8	0	0	0	0	1	0	0	0	0	1

	Item									
Attribute	11	12	13	14	15	16	17	18	19	20
1	0	0	0	0	1	0	0	0	1	0
2	1	0	1	1	0	1	1	1	1	1
3	0	0	0	0	0	0	0	0	1	1
4	0	0	1	0	0	0	0	0	0	0
5	1	0	1	0	0	0	1	1	1	1
6	0	0	0	0	0	0	0	1	0	0
7	1	1	1	1	1	1	1	1	1	1
8	0	1	0	0	0	0	0	0	0	0

TABLE 9.
Parameter Estimation Using the DINA Model

	Higher-Order Model				Independence Model			
Item	$\hat{g}$	PSD($\hat{g}$)	$1-\hat{s}$	PSD(1-$\hat{s}$)	$\hat{g}$	PSD($\hat{g}$)	$1-\hat{s}$	PSD(1-$\hat{s}$)
1	0.04	0.01	0.90	0.01	0.06	0.01	0.90	0.01
2	0.03	0.01	0.96	0.01	0.05	0.01	0.97	0.01
3	0.00	0.00	0.88	0.01	0.01	0.00	0.89	0.01
4	0.22	0.01	0.89	0.01	0.22	0.01	0.89	0.01
5	0.30	0.02	0.82	0.01	0.31	0.01	0.84	0.01
6	0.03	0.02	0.96	0.01	0.45	0.02	0.97	0.00
7	0.03	0.01	0.81	0.01	0.02	0.01	0.81	0.01
8	0.44	0.03	0.81	0.01	0.47	0.02	0.89	0.01
9	0.18	0.03	0.75	0.01	0.26	0.06	0.67	0.01
10	0.03	0.01	0.79	0.01	0.03	0.01	0.81	0.01
11	0.07	0.01	0.93	0.01	0.07	0.01	0.93	0.01
12	0.13	0.02	0.96	0.01	0.37	0.02	0.95	0.01
13	0.02	0.00	0.67	0.02	0.02	0.00	0.67	0.02
14	0.05	0.01	0.93	0.01	0.32	0.02	0.96	0.01
15	0.04	0.01	0.90	0.01	0.03	0.01	0.87	0.01
16	0.10	0.02	0.88	0.01	0.33	0.02	0.92	0.01
17	0.04	0.01	0.86	0.01	0.05	0.01	0.86	0.01
18	0.13	0.01	0.85	0.01	0.13	0.01	0.86	0.01
19	0.02	0.00	0.76	0.02	0.02	0.00	0.77	0.02
20	0.01	0.00	0.84	0.01	0.02	0.00	0.84	0.01

TABLE 10.
Estimated Proportion of Examinees Possessing $\boldsymbol{\alpha}$ and Agreement of Classification

	Attribute							
Model	1	2	3	4	5	6	7	8
Higher-Order	0.49	0.82	0.70	0.60	0.49	0.85	0.77	0.82
Independence	0.76	0.94	0.97	0.87	0.76	0.99	0.94	0.66
κ	0.33	0.12	0.05	0.22	0.31	0.00	0.09	0.59

parametric model. Under the estimated model parameters, the joint distribution for pairs of items can be computed, and the log-odds ratio for items j and j' is

$$\log\left[\frac{P(Y_j=1,Y_{j'}=1)P(Y_j=0,Y_{j'}=0)}{P(Y_j=1,Y_{j'}=0)P(Y_j=0,Y_{j'}=1)}\right]. \tag{21}$$

The mean absolute difference between the observed and expected log-odds-ratios of an item averaged over the rest of the items were computed for each item. The higher-order model produced a smaller mean absolute difference compared to the independence model for all item except item 8. Incidentally for the higher-order model, the mean absolute deviation of item 8 (1.05) is the only value that exceeds 0.54, an indication of poor fit for this item. In addition to having better fit for almost all of the items, an overall fit, computed as the mean absolute difference across the 190 pairs, is also smaller for the higher-order model, 0.43, compared to the independence model, 0.55.

A more global measure of model fit (i.e., test-level measure as opposed to item-level measure) was obtained using the Bayes factor. This is analogous to the likelihood ratio, but is used in a Bayesian context, and can be used as measure of evidence for a model with respect to another model even when the models are not nested. The Bayes factor, which is the ratio of the marginal likelihoods (i.e., the likelihoods after integrating over the model parameters), is computed as

$$B_{HI}=\frac{P(\mathbf{Y}|M_H)}{P(\mathbf{Y}|M_I)}. \tag{22}$$

In (22),

$$P(\mathbf{Y}|M_m)=\int P(\mathbf{Y}|\boldsymbol{\lambda}_m,\mathbf{s}_m,\mathbf{g}_m,M_m)P(\boldsymbol{\lambda}_m,\mathbf{s}_m,\mathbf{g}_m|M_m)d\boldsymbol{\lambda}_m d\mathbf{s}_m d\mathbf{g}_m \tag{23}$$

where $\boldsymbol{\lambda}_m$, $\mathbf{s}_m$, and $\mathbf{g}_m$ are the parameters under Model m, $P(\boldsymbol{\lambda}_m,\mathbf{s}_m,\mathbf{g}_m|M_m)$ is the prior density, and $m=\{H,I\}$.

Raftery (1996) proposed the Laplace-Metropolis estimator of the marginal likelihood. Dropping the index M_m, an approximation of the marginal likelihood is given by

$$P(\mathbf{Y})\approx(2\pi)^{d/2}|\Psi|^{1/2}P(\mathbf{Y}|\tilde{\boldsymbol{\lambda}},\tilde{\mathbf{s}},\tilde{\mathbf{g}})P(\tilde{\boldsymbol{\lambda}},\tilde{\mathbf{s}},\tilde{\mathbf{g}}) \tag{24}$$

where $\tilde{\boldsymbol{\lambda}}$, $\tilde{\mathbf{s}}$, and $\tilde{\mathbf{g}}$ are the posterior modes, Ψ is asymptotically equal to the posterior variance matrix of the parameters as sample size approaches infinity, and d is the number of parameters.

Because our interest is in the structural parameters (i.e., $\boldsymbol{\lambda}$, $\mathbf{s}$, $\mathbf{g}$), the incidental parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, needed to be integrated out in computing the $P(\mathbf{Y}|\tilde{\boldsymbol{\lambda}},\tilde{\mathbf{s}},\tilde{\mathbf{g}})$. For the higher-order DINA model, the conditional likelihood given the structural parameters was

$$P(\mathbf{Y}|\boldsymbol{\lambda}, \mathbf{s}, \mathbf{g}) = \int P(\mathbf{Y}|\boldsymbol{\alpha}, \mathbf{s}, \mathbf{g}) \left(\int P(\boldsymbol{\alpha}|\boldsymbol{\lambda}, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta} \right) d\boldsymbol{\alpha}, \tag{25}$$

and for the independence model, this conditional likelihood was

$$P(\mathbf{Y}|\boldsymbol{\lambda}, \boldsymbol{\beta}) = \int P(\mathbf{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta}) P(\boldsymbol{\alpha}|\boldsymbol{\lambda}) \, d\boldsymbol{\alpha}. \tag{26}$$

Computationally, because of the relatively large sample size, we substituted the posterior mean $\hat{\boldsymbol{\lambda}}$, $\hat{\mathbf{s}}$, and $\hat{\mathbf{g}}$ for $\tilde{\boldsymbol{\lambda}}$, $\tilde{\mathbf{s}}$, and $\tilde{\mathbf{g}}$, respectively; Ψ was estimated by computing the variance matrix of the simulation output. In addition, the integration of $\boldsymbol{\theta}$ was approximated using quadrature nodes. Hence, the numerator of the Bayes factor in (22) was computed as

$$P(\mathbf{Y}) \approx (2\pi)^{d/2} |\hat{\Psi}|^{1/2} \sum_{\forall \boldsymbol{\alpha} \in \mathbf{A}} \sum_{\forall n} P(\mathbf{Y}|\boldsymbol{\alpha}, \hat{\mathbf{s}}, \hat{\mathbf{g}}) P(\boldsymbol{\alpha}|n, \hat{\boldsymbol{\lambda}}) w(n) P(\hat{\boldsymbol{\lambda}}, \hat{\mathbf{s}}, \hat{\mathbf{g}}), \tag{27}$$

where $\mathbf{A}$ is the collection of all the possible attribute patterns, n is the quadrature node, and $w(n)$ is the weight of the node n. The denominator of (22) was computed in the same manner.

The $\log(B_{HI})$ is equal to 46.00, which indicates strong evidence for the higher-order model over the independence model (Raftery, 1995; 1996). As mentioned earlier, the difference between the log-marginal likelihoods can be traced mainly to the discrepancy in estimating the prevalence of the attributes: The higher-order model allows for better estimation of the proportion of the population possessing the required attributes.

5.4. The Higher-Order DINA Model and the 2PL Model

In addition to a better model fit afforded by the higher-order DINA model, it also allowed for the estimation of a broader latent trait within the same procedure. As previously discussed, this latent trait can be thought of as the analog of the latent trait in traditional IRT models (e.g., the logistic models). Because the test format was not multiple choice, an appropriate IRT model for these data is the 2PL model. It should be noted that the guessing parameter for the DINA model is more general and refers to any strategy by which one correctly answers an item without possessing the required attributes. Thus, it may be used for multiple choice exams, as well as other dichotomously scored formats. Two sets of ability estimates, $\boldsymbol{\theta}$, were computed, one using the 2PL model and another using the higher-order DINA model. The scatter plot of the these estimates is shown in Figure 1. Notwithstanding the shrinkage in the DINA model estimates that were based on only 8 attributes, the high correspondence between the two estimates, which has a correlation of 0.96, is evident. Therefore, using the higher-order DINA model, inference regarding more specific knowledge states and a more general ability trait can be obtained from the same data set in a single analysis.

5.5. Simulation Study Using Parameters Recovered from Actual Data

To verify whether the results in estimating the parameters of the DINA model obtained in the simulation section can be generalized to real data analysis, a simulation study with the same characteristics as the fraction subtraction data was conducted. Twenty-five data sets with 20 items, 8 attributes, and 2144 examinees using the DINA model were simulated. The estimates of $\boldsymbol{\lambda}$, $\mathbf{s}$, and $\mathbf{g}$ from the fraction subtraction analysis were used as the structural parameters. Compared to the simulation study above, this study involves fewer items, more attributes, and a more complex $\mathbf{Q}$-matrix, but at the same time, a larger number of examinees.

The mean estimates and standard deviations of the estimates for $\boldsymbol{\lambda}$, and the proportion of correct attribute classification, are given in Table 11; the mean estimates and standard deviation of

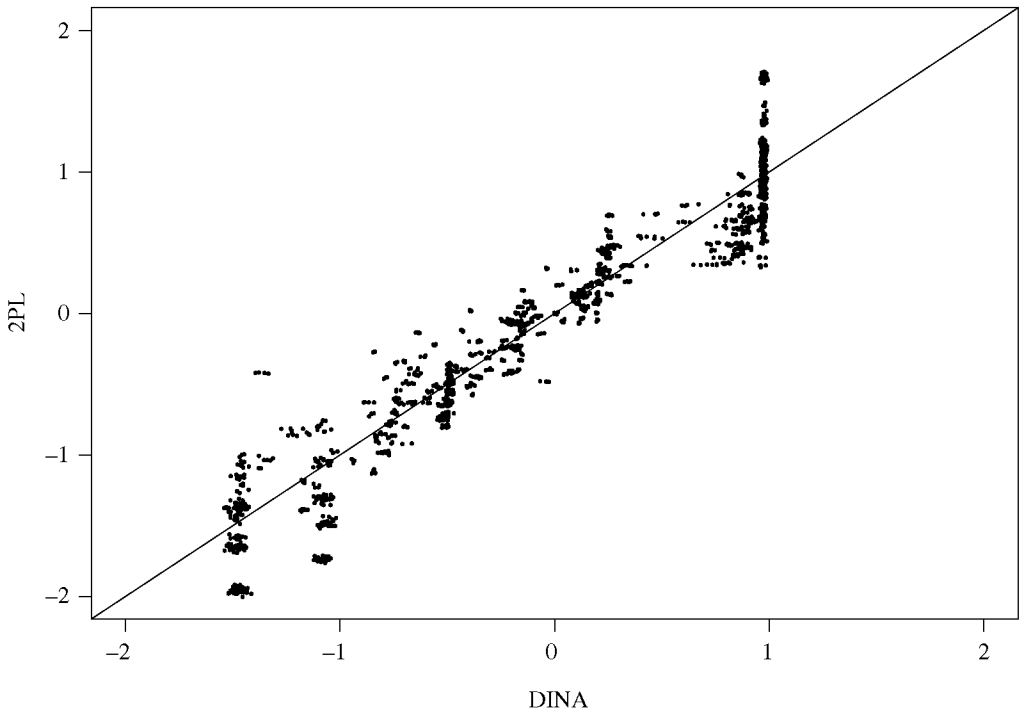

FIGURE 1.
Scatter Plot of the $\hat{\theta}$ Estimated Using the Higher-Order DINA Model and the 2PL Model

the estimates for $\boldsymbol{\lambda}$ and $\mathbf{s}$ are given in Table 12. In terms of accuracy, these results are comparable to the results obtained in the previous simulation study despite the differences in the $\mathbf{Q}$-matrix, and the number of items, attributes, and examinees. It can be noted that given the same level of accuracy in estimating $\boldsymbol{\alpha}$, $\boldsymbol{\lambda}$, $\mathbf{s}$, and $\mathbf{g}$, the larger number of attributes allowed for more accurate estimation of $\boldsymbol{\theta}$ as indicated by a mean correlation between the true and estimated θ of 0.83, and a RMSE of 0.56.

TABLE 11.
$\boldsymbol{\lambda}$ Estimates and Proportion of $\boldsymbol{\alpha}$ Correctly Classified (over 25 Replications)

Attribute	λ_0	λ_1	$\hat{\lambda}_0$	$SD(\hat{\lambda}_0)$	$\hat{\lambda}_1$	$SD(\hat{\lambda}_1)$	Proportion correct
1	0.05	4.94	−0.03	0.18	4.02	0.45	0.95
2	−1.05	1.82	−1.06	0.08	1.80	0.13	0.93
3	−1.53	0.96	−1.40	0.16	1.30	0.30	0.82
4	−0.28	1.96	−0.30	0.08	1.99	0.15	0.94
5	0.03	2.08	0.01	0.09	2.10	0.19	0.95
6	−1.45	1.78	−1.38	0.18	1.25	0.31	0.78
7	−0.98	3.12	−0.97	0.13	2.91	0.32	0.90
8	−0.89	1.60	−0.86	0.13	1.56	0.37	0.98

TABLE 12.
Estimation of Guess and Slip Parameters (over 25 Replications)

Item	g	$1-s$	$\hat{g}$	SD($\hat{g}$)	$1-\hat{s}$	SD($1-\hat{s}$)
1	0.04	0.90	0.04	0.01	0.90	0.01
2	0.03	0.96	0.03	0.01	0.96	0.01
3	0.00	0.88	0.01	0.00	0.87	0.01
4	0.23	0.89	0.23	0.02	0.89	0.01
5	0.30	0.83	0.31	0.01	0.83	0.01
6	0.05	0.96	0.05	0.02	0.96	0.01
7	0.03	0.80	0.03	0.01	0.80	0.01
8	0.44	0.81	0.43	0.02	0.81	0.01
9	0.17	0.75	0.17	0.03	0.75	0.02
10	0.03	0.79	0.03	0.01	0.79	0.02
11	0.07	0.92	0.07	0.01	0.92	0.01
12	0.13	0.96	0.14	0.01	0.96	0.01
13	0.01	0.67	0.02	0.00	0.67	0.02
14	0.05	0.94	0.05	0.01	0.93	0.01
15	0.04	0.89	0.04	0.01	0.89	0.01
16	0.11	0.88	0.11	0.01	0.88	0.01
17	0.04	0.86	0.04	0.01	0.86	0.01
18	0.12	0.85	0.13	0.01	0.86	0.01
19	0.02	0.76	0.03	0.00	0.76	0.02
20	0.01	0.84	0.01	0.00	0.84	0.01

6. Discussion

When fitting multiple classification latent class models for cognitive diagnosis, difficult model selection and model fitting decisions arise. We have investigated the importance of correct model specification, and have introduced higher-order latent traits. These higher-order traits simultaneously afford a parsimonious model and express the concept of more general abilities affecting the acquisition of specific knowledge. Models for cognitive diagnosis are often constructed to recognize the steps that are required in problem solving, and lead to conjunctive models in which each step must be correctly executed. Conjunctive models contrast with the formulation of most familiar models in item response theory and item factor analysis, in which compensatory models are most common. Simulation reveals that using an incorrect model can lead to poor attribute classification. Fraction subtraction is an apparent case of a conjunctive process, and we have seen that the DINA model with a single higher-order trait provides a good fit and can be used to diagnose the eight attributes that were listed. In this case where a conjunctive model is the better choice, the linear logistic model fitted with the same $\mathbf{Q}$ gave completely unsatisfactory results that were not reported. A DINA model with statistically independent attributes gave more reasonable results than the compensatory model, but the clearly incorrect assumption of independence resulted in a fit inferior to that of the higher-order model. However, in some situations the use of the compensatory model will be more appropriate. For example, in psychiatric or medical diagnosis, a particular symptom is an indication of the presence of at least one of the disorders.

The simulation and real data examples considered here utilized a unidimensional higher-order trait. In some applications a two-dimensional higher-order trait will be preferred and minor modifications to the MCMC procedure will be required, which is the addition of a step to estimate the correlation parameter for the traits. If an estimate of θ is desired for scoring in the sense

of IRT modeling, a requirement is that there should be many more attributes than higher-order traits, and this issue becomes increasingly important as the number of traits increases due to the quadratically increasing number of correlation parameters. For this reason we see unidimensional and two-dimensional traits as the most feasible and practically useful cases. However, this restriction may not be as critical if the higher-order traits are merely nuisance parameters used to model the joint distribution of the attributes correctly.

Higher-order traits can be used with virtually any multiple classification latent class model. Algorithms for using MCMC were given for the particular models that we considered, DINA and LLM. Simulation results show that estimation with MCMC is quite effective for both of these models using higher-order traits. The MCMC code written in Ox (Doornik, 2002), an object-oriented mathematical programming language, run on a 2.5 GHz processor, was capable of performing approximately 500 iterations per minute with the fraction subtraction data, which included 2144 examinees. The code for doing this research can be made available by contacting the first author.

References

Best, N.G., Cowles, M.K., & Vines, S.K. (1995). CODA: Convergence diagnosis and output analysis software for Gibbs sampling output (Version 0.30). [Computer software]. Cambridge: MRC Biostatistics Unit.

Bock, R.D., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443–459.

Casella, G., & George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician, 46*, 167–174.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm. *The American Statistician, 49*, 327–335.

DiBello, L.V., Stout, W.F., & Roussos, L.A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P.D. Nichols, S.F. Chipman, & R.L. Brennan, *Cognitively diagnostic assessment* (pp. 361–389). Hillsdale, NJ: Erlbaum.

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.

Draney, K.L., Pirolli, P., & Wilson, M. (1995). A measurement model for complex cognitive skill. In P.D. Nichols, S.F. Chipman, R.L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 103–125). Hillsdale, NJ: Erlbaum.

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

Embretson, S. (1997). Multicomponent response models. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 305–321). New York: Springer–Verlag.

Everitt, B. S. (1998). *The Cambridge dictionary of statistics*. Cambridge, UK: Cambridge University Press.

Fischer, G. H. (1995). The linear logistic test model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 131–155) New York: Springer–Verlag.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721–741.

Gelman, A., & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science, 7*, 457–511.

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

Hagenaars, J.A. (1990). *Categorical longitudinal data: Loglinear panel, trend, and cohort analysis*. Newbury Park, CA: Sage.

Hagenaars, J.A. (1993). *Loglinear models with latent variables*. Newbury Park, CA: Sage.

Hartz, S. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practicality*. Unpublished doctoral dissertation.

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.

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.

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

Muthén, B. (1978). Contribution to factor analysis of binary variables. *Psychometrika, 43*, 551–560.

Patz, R.J., & Junker, B.W. (1999a). A straightforward approach to Markov chain Monte Carlo methods for item response theory. *Journal of Educational and Behavioral Statistics, 24*, 146–178.

Patz, R.J., & Junker, B.W. (1999b). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics, 24*, 342–366.

Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163.

Raftery, A.E. (1996). Hypothesis testing and model selection. In R. W. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 163–187). London: Chapman & Hall.

Reckase, M. (1997). A linear logistic multidimensional model for dichotomous item response data. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 271–286). New York: Springer–Verlag.

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

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

Manuscript received 27 AUG 2002
Final version accepted 14 JUL 2003