Article

Applied Psychological Measurement
35(1) 8–26

DOI: 10.1177/0146621610377081
http://apm.sagepub.com
SAGE

On the Analysis of Fraction Subtraction Data: The DINA Model, Classification, Latent Class Sizes, and the Q-Matrix

Lawrence T. DeCarlo[1]

Abstract

Cognitive diagnostic models (CDMs) attempt to uncover latent skills or attributes that examinees must possess in order to answer test items correctly. The DINA (deterministic input, noisy "and") model is a popular CDM that has been widely used. It is shown here that a logistic version of the model can easily be fit with standard software for latent class analysis. A partly Bayesian approach to estimation, posterior mode estimation, is used as a simple alternative to a fully Bayesian approach via Markov chain Monte Carlo methods. A latent-class analysis of a widely analyzed data set, the fraction subtraction data of K. K. Tatsuoka, reveals some neglected problems with respect to the classification of examinees; for example, examinees who get all of the items incorrect are classified as having most of the skills. It is also noted that obtaining large estimates of the latent class sizes can indicate misspecification of the Q-matrix, such as the inclusion of an irrelevant skill. It is shown, analytically and via simulations, that the problems are largely associated with the structure of the Q-matrix.

Keywords

cognitive diagnosis, DINA, fraction subtraction data, latent class models, posterior mode estimation, MCMC

Cognitive diagnostic models (CDMs) attempt to uncover latent skills or attributes that examinees must possess in order to answer test items correctly (for recent reviews, see DiBello, Roussos, & Stout, 2007; Fu & Li, 2007; Rupp & Templin, 2008a). The DINA model is a popular CDM that has been widely used. It is shown here that a logistic version of the model can easily be fit using standard software for latent class analysis. A partly Bayesian approach to estimation, posterior mode estimation, is used as a simple alternative to a fully Bayesian approach via Markov chain Monte Carlo methods; the approach offers a way to deal with boundary problems. A latent class analysis of a widely analyzed data set, the fraction subtraction data of K. K. Tatsuoka (1990), reveals some neglected problems with respect to the classification of examinees; for example,

[1]Columbia University, New York, New York, USA

Corresponding Author:
Lawrence T. DeCarlo, Department of Human Development, Box 118, Teachers College, Columbia University, 525 West 120th Street, New York, NY 10027, USA
Email: decarlo@tc.edu

Table 1. Fraction Subtraction Data, 20 Items, 8 Hypothesized Skills

Item no.	Item	Skills	Item no.	Item	Skills
1	5/3 – 3/4	$\alpha_4, \alpha_6, \alpha_7$	11	4 1/3 – 2 4/3	$\alpha_2, \alpha_5, \alpha_7$
2	3/4 – 3/8	α_4, α_7	12	1 1/8 – 1/8	α_7, α_8
3	5/6 – 1/9	α_4, α_7	13	3 3/8 – 2 5/6	$\alpha_2, \alpha_4, \alpha_5, \alpha_7$
4	3 1/2 – 2 3/2	$\alpha_2, \alpha_3, \alpha_5, \alpha_7$	14	3 4/5 – 3 2/5	α_2, α_7
5	4 3/5 – 3 4/10	$\alpha_2, \alpha_4, \alpha_7, \alpha_8$	15	2 – 1/3	α_1, α_7
6	6/7 – 4/7	α_7	16	4 5/7 – 1 4/7	α_2, α_7
7	3 – 2 1/5	$\alpha_1, \alpha_2, \alpha_7$	17	7 3/5 – 4/5	$\alpha_2, \alpha_5, \alpha_7$
8	2/3 – 2/3	α_7	18	4 1/10 – 2 8/10	$\alpha_2, \alpha_5, \alpha_6, \alpha_7$
9	3 7/8 – 2	α_2	19	4 – 1 4/3	$\alpha_1, \alpha_2, \alpha_3, \alpha_5, \alpha_7$
10	4 4/12 – 2 7/12	$\alpha_2, \alpha_5, \alpha_7, \alpha_8$	20	4 1/3 – 1 5/3	$\alpha_2, \alpha_3, \alpha_5, \alpha_7$

Note: α_1 = convert a whole number to a fraction; α_2 = separate a whole number from a fraction; α_3 = simplify before subtracting; α_4 = find a common denominator; α_5 = borrow from whole number part; α_6 = column borrow to subtract the second numerator from the first; α_7 = subtract numerators; and α_8 = reduce answers to simplest form. The Q-matrix is from de la Torre and Douglas (2004).

examinees who get all of the items incorrect are classified as having most of the skills. It is also noted that obtaining large estimates of the latent class sizes can indicate misspecification of the Q-matrix, such as the inclusion of an irrelevant skill. It is shown, analytically and via simulations, that the problems are largely associated with the structure of the Q-matrix. Some issues with respect to the use of higher order DINA models are also noted.

The DINA Model

Although a number of CDMs have been proposed, the DINA (deterministic input, noisy ''and'') model (see Haertel, 1989; Junker & Sijtsma, 2001; Macready & Dayton, 1977) has been used to analyze the fraction subtraction data in many articles (e.g., de la Torre, 2009; de la Torre & Douglas, 2004, 2008; Henson, Templin, & Willse, 2009; Templin, Henson, & Douglas, 2006) and is of focus here. To make things concrete, Table 1 shows 20 items from the fraction subtraction test; the data are available at the website for the *Journal of the Royal Statistical Society* www.blackwellpublishing.com/rss/Volumes/Cv51p3.htm. The data are for 536 middle school children.[1]

The Q-matrix specifies the skills that are believed to be involved in solving the tasks; the matrix used here is the same as that used by de la Torre and Douglas (2004; given in their Table 5). The skills are as follows: (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. The skills assumed necessary for each item of the fraction subtraction data are shown in Table 1.

The ideas underlying the DINA model are relatively simple. The basic idea is that in order to answer an item correctly, the examinee must have certain skills or attributes, typically denoted as α_{ik}, for examinee i and skill k. The examinee is viewed as either having or not having a particular skill, and so the α_{ik} are dichotomous latent variables, with values of zero or one indicating absence or presence of a skill, respectively. The deterministic part of the DINA model is that in order for an examinee to answer an item correctly, they must have all of the skills required by the item, and so DINA is a *conjunctive* model (see Maris, 1999). For example, it is assumed that in order to answer the item 6/7–4/7 correctly (Item 6), the examinee must have a skill of "subtract numerators" (Skill 7), whereas to answer 3/4–3/8 (Item 2) correctly, the examinee must have both the "subtract numerators" skill and an additional skill of "find a common denominator" (Skill 4), as shown in Table 1.

Based on these ideas, the DINA model can be developed as follows. Let Y_{ij} be a binary variable that indicates whether the response of the ith examinee to the jth item is correct or incorrect (1 or 0). Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_K)'$ denote the vector of K skills that are needed to solve the items. The particular skills needed to solve the jth item are given by the Q-matrix (K. K. Tatsuoka, 1990), which consists of j rows and k columns. Each row of the Q-matrix consists of zeroes and ones, with a value of zero indicating that the kth skill is not needed for the jth item, and a value of one indicating that the skill is needed.

Let η_{ij} be a binary variable that indicates whether the ith examinee has the set of skills (attributes) needed to solve the jth item, with η_{ij} determined by the skills as follows:

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}} = \mathbf{h}(\boldsymbol{\alpha}_i, \mathbf{q}_j), \tag{1}$$

where the function $\mathbf{h}(\boldsymbol{\alpha}_i, \mathbf{q}_j)$ denotes the operation of multiplying each element of $\boldsymbol{\alpha}$ raised to a power of q_{jk}, where q_{jk} is either zero or one, as specified by the Q-matrix (cf. Henson et al., 2009; von Davier, 2005). The above represents the *deterministic input* part of the model, in that the presence or absence of the necessary set of skills determines whether η_{ij} is zero or one. It is a conjunctive model in that the examinee must have all of the skills necessary to solve the problem. To illustrate, suppose that out of a set of four skills, Skills 1 and 3 are necessary to solve the first item. Then, Equation 1 together with the Q-matrix gives, for the first item, $\eta_{i1} = {\alpha_{i1}}^1 \times {\alpha_{i2}}^0 \times {\alpha_{i3}}^1 \times {\alpha_{i4}}^0 = \alpha_{i1} \times \alpha_{i3}$. Thus, because of the multiplicative term, η_{ij} is zero if either α_{i1} or α_{i3}, or both, are zero, and η_{ij} is equal to 1 if and only if both skills are present, which is the conjunctive aspect of the model.

Next, let g_j denote the probability that the examinee gets the jth item correct if they do not have the necessary skills:

$$p(Y_{ij} = 1 | \eta_{ij} = 0) = g_j, \tag{2}$$

which is the *guess* rate. Note that g_j is written on the right side of the equal sign (it is often written on the left side as a defined quantity) to reflect that Equation 2 is a model of the response probabilities with parameter g_j. Let s_j denote the probability that the examinee gets the jth item incorrect when they have the necessary skills:

$$p(Y_{ij} = 0 | \eta_{ij} = 1) = s_j, \tag{3}$$

which is the *slip* rate. Using terminology from signal detection theory (e.g., Macmillan & Creelman, 2005), g_j is the probability of a *false alarm*, $1 - s_j$ is the probability of a *hit*, and s_j is the probability of a *miss*. Note that the notion of guessing is simply an interpretation of the false alarm rate. It should also be recognized that the guessing and slip parameters are conditional on the specific skills specified by the Q-matrix, and so changing the Q-matrix can change the slip and guess parameters.

It follows from Equations 2 and 3 that the probability that an examinee gets an item correct is

$$p(Y_{ij} = 1 \, | \eta) = (1 - s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}},$$

which is the DINA model (e.g., see de la Torre, 2009; de la Torre & Douglas, 2004; Henson et al., 2009; Junker & Sijtsma, 2001). The DINA model has been fit within a fully Bayesian framework using Markov chain Monte Carlo (MCMC) methods (e.g., de la Torre & Douglas, 2004, 2008; Henson et al., 2009; Junker & Sijtsma, 2001). Templin and Hoffman (2010) have shown how to incorporate the model into software for second-generation structural equation modeling, namely Mplus (Muthén & Muthén, 1998-2007).

The model can also be fit, using maximum-likelihood estimation (MLE), with standard software for latent class analysis. This can most easily be done by reparameterizing the model as a logistic regression model with latent classes. In particular, note that g_j can be replaced by a function that gives a positive value in the range of zero to one. For example,

$$p(Y_{ij} = 1|\eta_{ij} = 0) = \exp(f_j)/[1 + \exp(f_j)],$$

where exp is the exponential function. The above gives the probability that an examinee gets an item correct given that they do not have the requisite skills (a false alarm) and can be written more simply by using a logit function, $\log[p/(1 - p)]$:

$$\text{logit } p(Y_{ij} = 1|\eta_{ij} = 0) = f_j.$$

The parameter f_j has a simple interpretation, in that it is the log odds of a false alarm. Similarly, the log odds of a hit is

$$\text{logit } p(Y_{ij} = 1|\eta_{ij} = 1) = f_j + d_j.$$

The parameter d_j indicates how well item j detects the presence versus absence of the requisite skill set. Subtracting the first equation from the second equation shows that d_j is the log odds ratio of a hit versus false alarm.

The above two equations can be written as a single model as follows,

$$\text{logit } p(Y_{ij} = 1|\eta_{ij}) = f_j + d_j\eta_{ij}. \tag{4}$$

Equation 4 is a logistic model with latent classes, with the items serving as detectors of the skill sets. The parameter f_j gives the log odds of a false alarm, where a false alarm is the probability that an examinee gets item j correct given that they do not have the requisite skills. The parameter d_j reflects how well the item discriminates between examinees with and without the requisite skills; for a different (yet related) discrimination index, see Henson, Roussos, Douglas, and He (2008). The model of Equation 4 is a type of latent class signal detection model (DeCarlo, 2002, 2005); indeed, Junker and Sijtsma (2001) noted in their discussion of the DINA model that it was basically a ''simple signal detection model for detecting'' the skill set from ''noisy observations'' (p. 263), which is exactly the interpretation of the model as formulated here.[2] The model is within a general class of latent structure models that have been used for CDMs (see Henson et al., 2009; von Davier, 2005, 2009).

Given that Equation 4 is simply a reparameterization of the DINA model, it will be referred to here as the *reparameterized* DINA (RDINA) model. Note that if one prefers the original probability formulation of the model, then it is simple to recover the DINA model parameters from the RDINA parameters. In particular, the guessing parameter is

$$g_j = \exp(f_j)/[1 + \exp(f_j)],$$

and one minus the slip parameter is

$$(1-s_j) = \exp(f_j + d_j)/[1 + \exp(f_j + d_j)].$$

Including the vector of latent skills $\boldsymbol{\alpha}$ into Equation 4 gives

$$\text{logit } p(Y_{ij} = 1|\boldsymbol{\alpha}) = f_j + d_j\mathbf{h}(\boldsymbol{\alpha}_i, \mathbf{q}_j), \tag{5}$$

where $\mathbf{h}(\boldsymbol{\alpha}_i, \mathbf{q}_j)$ is given by Equation 1.

The RDINA model can be viewed as the first part of a restricted latent class model (see Clogg, 1995; Dayton, 1998). The complete model also consists of a second part, which is a model for the skills. Note that the basic data are observed response patterns for the J items (i.e., correct or incorrect for each item), that is, $p(Y_{i1}, Y_{i2}, \ldots, Y_{iJ})$. A latent class model relates the response pattern probabilities to the conditional response probabilities of Equation 5 as follows:

$$p(Y_{i1}, Y_{i2}, \ldots, Y_{iJ}) = \sum\nolimits_{\boldsymbol{\alpha}} p(\boldsymbol{\alpha})\ p(Y_{i1}, Y_{i2}, \ldots, Y_{iJ}|\boldsymbol{\alpha}),$$

where the sum is over the values of $\boldsymbol{\alpha}$ (i.e., the 2^K patterns for the K skills). The above simply reflects a basic premise of latent class analysis—that the unconditional probability of the response vector is a weighted sum of the conditional probabilities over the latent classes. An assumption of local independence of the conditional response probabilities is made:

$$p(Y_{i1}, Y_{i2}, \ldots, Y_{iJ}|\boldsymbol{\alpha}) = \prod\nolimits_j p(Y_{ij}|\boldsymbol{\alpha}),$$

where the $p(Y_{ij}|\boldsymbol{\alpha})$ are obtained from Equation 5 by using the inverse of the logit link. Finally, a structure for $p(\boldsymbol{\alpha})$ must also be assumed. The simplest assumption is independence:

$$p(\boldsymbol{\alpha}) = p(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_K) = \prod\nolimits_k p(\boldsymbol{\alpha}_k).$$

The model used here for $p(\alpha_k)$ is

$$p(\boldsymbol{\alpha}_k) = \exp(b_k)/[1 + \exp(b_k)], \tag{6}$$

where b_k is an item difficulty parameter.

The Higher Order DINA Model

Although the independence model might be appropriate in some situations, and provides a useful baseline for comparison for more complex models, there are many situations where it is likely that there is a higher order structure among the skills. One way to recognize this is through the use of a *higher order* DINA model (de la Torre & Douglas, 2004; also see Templin, Henson, Templin, & Roussos, 2008), which will be denoted here as the HO-RDINA model. The model replaces the independence assumption for $p(\boldsymbol{\alpha})$ shown above with a higher order structure. Specifically, one can assume that the probability of having the skills depends on a higher order continuous factor, such as an overall ability θ:

$$p(\boldsymbol{\alpha}) = \int_{\theta} p(\boldsymbol{\alpha}|\theta)\ p(\theta)\ d(\theta),$$

where $p(\boldsymbol{\alpha}|\theta)$ is given by the higher order model. The higher order model again assumes conditional independence, specifically that the skills are independent conditional on θ:

$$p(\boldsymbol{\alpha}|\theta) = p(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_K|\theta) = \prod\nolimits_k p(\boldsymbol{\alpha}_k|\theta).$$

The higher order model for the skills uses a simple extension of Equation 6, as also used by de la Torre and Douglas (2004):

$$p(\boldsymbol{\alpha}_k|\theta) = \exp(b_k + a_k\theta)/[1 + \exp(b_k + a_k\theta)], \tag{7}$$

with item difficulty parameter b_k and item discrimination parameter a_k. The model includes a latent continuous variable θ to account for associations among the K skills. Note that de la Torre and Douglas (2004) fit a restricted version of the model where the discrimination parameters a_k were restricted to be equal across the eight skills, which is referred to here as the *restricted* higher order RDINA model, or RHO-RDINA.

Implementation of RDINA and HO-RDINA as Latent Class Models

It is informative to write out the set of equations implied by Equation 5 for each of the items, in that it shows exactly how the model can be specified in software for latent class analysis (it also illustrates a problem discussed below). For example, using the attributes shown in Table 1 for the fraction subtraction data, obtained from the Q-matrix, the equations for the first four items are

$$\begin{aligned}
\operatorname{logit} p(Y_{i1} = 1|\boldsymbol{\alpha}) &= f_1 + d_1\ \alpha_{i4}\ \alpha_{i6}\ \alpha_{i7}, \\
\operatorname{logit} p(Y_{i2} = 1|\boldsymbol{\alpha}) &= f_2 + d_2\ \alpha_{i4}\ \alpha_{i7}, \\
\operatorname{logit} p(Y_{i3} = 1|\boldsymbol{\alpha}) &= f_3 + d_3\ \alpha_{i4}\ \alpha_{i7}, \\
\operatorname{logit} p(Y_{i4} = 1|\boldsymbol{\alpha}) &= f_4 + d_4\ \alpha_{i2}\ \alpha_{i3}\ \alpha_{i5}\ \alpha_{i7},
\end{aligned} \tag{8}$$

and similarly for the remaining items.

Equation 8 is informative in several ways. First, it explicitly shows the assumed skills for each item, which is the theory as specified by the Q-matrix. Second, it clarifies that the f_j and d_j parameters are conditional on the specific skills that are included for each item. Third, it shows that d_j indicates how well an item discriminates between the presence and absence of the requisite set of skills, given that the α_{ik} are simply zero/one latent variables that are multiplied. Fourth, Equation 8 shows that the conjunctive aspect of the DINA model is reflected by multiplicative interaction terms for the latent binary skills. This is important because it shows that the model can be fit by simply defining latent binary variables α_{ik} and including them as multiplicative interaction terms in a logistic model. The model will be illustrated via an analysis of the fraction subtraction data of K. K. Tatsuoka (1990) as given on the *Journal of the Royal Statistical Society* website.

The syntax required to fit the model follows directly from Equation 8 and is shown in Appendix A for Latent Gold (Vermunt & Magidson, 2007) and in Appendix B for LEM (Vermunt, 1997); some comments on the syntax for Latent Gold (LG) follow. The *variables* section defines the model variables. The dependent variables i1 to i20 in Appendix A are the item scores for the 20 items, coded as zero for incorrect and one for correct; the cumlogit function (which is $p(Y > m)$ for $m = 0, 1$, for incorrect/correct responses) is used so that f_j is as defined above (the default in LG of adjacent category logits can also be used). The variables a1 to a8 are the eight latent binary attributes measured by the items; they could be also declared as ''nominal'' in LG, but they are declared as ''ordinal'' here with scores of zero or one (which simply gives one more control over the parameterization; the model is in fact the same if the latent variables are set as nominal). For the higher order model, a latent continuous variable ''theta'' is included in the variables statement (the default number of 10 nodes is used). Next, the model is specified in the *equations* section of the LG syntax. The ''<-'' symbol indicates that the equation is a regression equation, with a dependent variable on the left hand side and predictor variables on the right. The interaction terms shown in Equation 8 are specified directly in the model, as shown in Appendix A, and follow directly from the Q-matrix. For a similar example of including a latent continuous variable as a predictor of latent categorical variables in a higher order model, see DeCarlo (2010).

It is well known in latent class analysis that *boundary* problems often occur (e.g., Clogg & Eliason, 1987; Maris, 1999). In the present context, boundary problems occur when one or more of the latent class size estimates are zero (or one), with large or indeterminate standard errors, and/or when the DINA model parameter estimates (and *SE*s) are large or indeterminate. To address this problem, Maris (1999) discussed an approach based on maximum a posteriori estimation, also referred to as *posterior mode estimation* (PME; also see Fahrmeir & Tutz, 2001; Galindo-Garre & Vermunt, 2006; Gelman, Carlin, Stern, & Rubin, 1995; Schafer, 1997; Vermunt & Magidson, 2005). In PME, the maximum of the log posterior function is found, where the log posterior function is the log likelihood plus a log prior; the log prior in essence serves as a penalty for solutions that are close to the boundary, so that parameter estimates are smoothed away from the boundary. PME is less computationally intense than a fully Bayesian approach (via MCMC) in that only the mode of the posterior distribution needs to be determined, and so PME is simple to implement. For example, Schafer (1997) noted that any algorithm used for MLE, such as the iterative proportional fitting algorithm commonly used for log-linear models, can easily be modified to find posterior modes when Dirichlet priors are used (p. 307; also see Gelman et al., 1995); in the present case, the Dirichlet priors are for the conditional response probabilities and for the latent class probabilities.

Galindo-Garre and Vermunt (2006) presented a simulation that suggested that posterior mode estimation gave more reliable parameter estimates and standard errors than either MLE or parametric bootstrapping; Galindo-Garre, Vermunt, and Bergsma (2004) presented a simulation that suggested advantages of posterior mode estimation over *expected a posteriori* estimation. In the context of latent class SDT, simulations presented in DeCarlo (2008, 2010) indicated that posterior mode estimation (with Bayes constants of unity) ameliorated boundary problems and led to good parameter recovery. PME was used here with values of unity for the Bayes constants; the Bayes constants are hyperparameters of the Dirichlet priors that are used for the conditional response probabilities (i.e., Equation 5) and for the latent class probabilities (i.e., Equations 6 or 7), see Vermunt and Magidson (2005, 2007) for details.

Analysis of the Fraction Subtraction Data

The analysis presented here replicates and extends that presented by de la Torre and Douglas (2004), using PME in lieu of MCMC.

Parameter Estimates

For both the RDINA and RHO-RDINA models, the parameter estimates, transformed as shown above to get g_j and s_j, were virtually identical to the guess and slip estimates shown in Table 9 of de la Torre and Douglas (2004), and so they are not shown. However, parameter estimates for the higher order part of the HO-RDINA model (i.e., Equation 7) have not previously been presented for the fraction subtraction data and so are shown in Table 2. Estimates of b_k range from −0.62 to 4.10; the discrimination estimates, a_k, are all close to 3, except for Skill 3, where $\hat{a}_3$ is small (0.70, $SE = 1.09$) and is not significantly different from zero. For the RHO-RDINA model, the common estimate of a_k is 3.43 with an *SE* of 0.23. Given that the RHO-RDINA model, which restricts a_k to be equal across the items, is nested within the HO-RDINA model, a likelihood ratio (LR) test of the parameter restriction can be performed by subtracting the −2 log likelihoods. For

Table 2. Parameter Estimates for the Second Level of the HO-RDINA Model

Skill	b_k	a_k	Skill	b_k	a_k
1	−0.62 (0.50)	5.26 (1.01)	5	−0.18 (0.20)	3.11 (0.41)
2	2.97 (0.45)	3.16 (0.78)	6	3.00 (0.63)	2.25 (1.14)
3	2.65 (0.78)	0.70 (1.09)	7	4.10 (0.99)	4.60 (1.51)
4	0.89 (0.21)	3.25 (0.44)	8	2.25 (0.34)	2.47 (0.65)

Note: *SE* values are in parentheses.

Table 3. Estimates of the Latent Class Sizes for the Eight Hypothesized Skills

Model	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
RDINA	.75 (.03)	.93 (.02)	.96 (.02)	.87 (.02)	.76 (.03)	.99 (.01)	.66 (.02)	.93 (.02)
HO-RDINA	.47 (.02)	.80 (.03)	.92 (.09)	.59 (.02)	.48 (.02)	.85 (.07)	.81 (.02)	.77 (.03)
RHO-RDINA	.48 (.02)	.80 (.02)	.74 (.07)	.59 (.02)	.48 (.02)	.82 (.04)	.81 (.02)	.76 (.02)

Note: RDINA = reparameterized DINA (deterministic input, noisy "and"); HO = higher order; RHO = restricted higher order.

the fraction subtraction data, this gives an LR statistic of 14.70 on 7 degrees of freedom with $p = .04$, which indicates rejection at the .05 level of the restriction of equal a_k.

Latent Class Sizes

The first line of Table 3 shows, for the RDINA model, estimates of the latent class sizes for the eight attributes, α_k; note that the estimates are nearly identical to those shown in Table 10 of de la Torre and Douglas (2004).[3] The estimated latent class sizes for Skills 3 and 6 (α_3 and α_6) are large (.96 and .99) and within 2 standard errors of 1.0. As discussed below, this result could indicate misspecification of the Q-matrix. The latent class size estimates for Skills 2 and 8 are also large (.93, both with $SE = 0.02$). Also note that Skill 7 has the smallest class size, .66, whereas all of the other skill class sizes are .75 or larger. If only 66% of the examinees have Skill 7 (subtract numerators), which can be viewed as being a lower level skill compared to the other skills, then how can more than 76% of the examinees have the other (higher level) skills?

The second line of Table 3 shows estimates of the latent class sizes obtained for the higher order model, HO-RDINA. The table shows that the estimated class sizes are smaller than those for the independence model, except for Skill 7, which has a larger estimated class size. Thus, the latent class size estimates are generally smaller for the higher order model; Skill 3, however, still has a large class size estimate of .92 ($SE = 0.09$). The third line of Table 3 shows the latent class size estimates for a fit of the RHO-RDINA model (with a_k restricted to be equal across the eight skills). The estimates are virtually identical to those found for the unrestricted HO-DINA model, except that the class size estimate for Skill 3 is now considerably smaller, as also found by de la Torre and Douglas (2004; Table 10). Note, however, that it is questionable as to whether the smaller class size found for Skill 3 with RHO-RDINA is an improvement, as compared to the results found with HO-RDINA, or is a distortion. For example, simulations presented below suggest that the finding of large latent class size estimates for Skill 3 for both the RDINA and HO-RDINA models might indicate that the Q-matrix is misspecified (in that Skill 3 should not be included).

Table 4. Posterior Probabilities of $\alpha = 1$ for Each Skill: 13 Examinees With a Total Score of Zero

	Skill							
Model	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
RDINA	.752	.862	.963	.868	.760	.985	.000	.931
HO-RDINA	.007	.118	.828	.061	.028	.441	.002	.252
RHO-RDINA	.029	.131	.159	.069	.028	.267	.003	.181

Note: RDINA = reparameterized DINA (deterministic input, noisy "and"); HO = higher order; RHO = restricted higher order.

Classification

The goal in latent class analysis is typically classification, which can be done using the modal posterior probabilities (see Clogg, 1995; Dayton, 1998). An examination of classifications obtained for the fraction subtraction data clearly suggests problems. For example, for the RDINA model, all 536 examinees are classified as having Skill 6 (''column borrow''), which can be viewed as a higher level skill, yet many of the examinees are classified as not having Skill 7 (''subtract numerators''), a lower level skill. Furthermore, examinees who got every item incorrect are classified as having every skill except Skill 7. This is shown in Table 4, which presents the posterior probabilities for 13 examinees whose total score was zero. Note that all of the estimated posterior probabilities, except that for Skill 7, are greater than .50, and so the examinees are classified as having $\alpha = 1$ for all of the skills except Skill 7, for which $\alpha = 0$ (at the other extreme, examinees with perfect scores were classified as having all of the skills).

The second row of Table 4 shows results for classifications obtained from the HO-RDINA model. In this case, examinees with a total score of zero are classified as not having most of the skills, and so the higher order model ameliorates the classification problems to some extent. However, the posterior probabilities in Table 4 show that examinees with total scores of zero are classified as having Skill 3, and so classification problems still appear in the higher order DINA model. The basic problem is that Skill 3 does not load significantly on the latent continuous variable, as shown in Table 2, and so it is not constrained by the latent variable as the other skills are (i.e., it is not highly associated with the other skills; note that this problem will also arise if a simple correlation structure for the skills is used). The third row shows classifications for the restricted high-order model, as used by de la Torre and Douglas (2004). In this case, examinees with scores of zero are now classified by the posterior probabilities as not having Skill 3.

In sum, the classification of examinees and the estimated latent class sizes obtained from an analysis of the fraction subtraction data suggest some problems, each of which is considered in the next sections.

Classification and the Q-Matrix

Given that the main selling point of CDMs is that they can be used to classify examinees into skill sets, which is what provides ''diagnostic feedback,'' the performance of the models in terms of classification is of central interest. As shown above, classifications obtained for both the RDINA and HO-RDINA models suggest problems. For example, although it seems reasonable to assume that examinees with a total score of zero should be classified as not having the skills, the results in Table 4 show that for the basic RDINA model, examinees with total scores of zero are classified as having every skill except the basic skill of ''subtract numerators,'' Skill 7. In fact, the problem is even deeper, in that 34% of the 536 examinees are classified as not having Skill 7, yet virtually all of them are classified as having the other (higher level) skills. As shown above, the higher order model ameliorates this problem; however, Skill 3 still presents a problem, in that examinees with scores of zero are classified as having the skill. The restricted higher order model helps with this problem, but there are questions as to whether the restriction is appropriate or if the restriction is simply masking misspecification of the Q-matrix (see below).

Here it is shown that the classification problems arise largely because of the Q-matrix specification. In particular, note that Skill 3 only enters the model as an interaction term. As a result, the posterior probabilities for Skill 3 are determined largely by the prior probabilities. This can be shown in several ways. One interesting and informative way is to generate data using the design used for the fraction subtraction data. One can then vary the latent class sizes and examine the effect on classification. Note that this can easily be done using the LG program used for the fraction subtraction data (Appendix A) by simply adding *outfile* and *simulation* commands, as well as values for the population parameters (see the LG manual). The reader can then verify that for the fraction subtraction design, if values greater than .5 (say .7) are used for the latent class sizes, then examinees with a total score of zero tend to be classified as having the skills, whereas if values less than .5 (say .3) are used, then the examinees tend to be classified as not having the skills. This shows that the posterior probabilities are largely determined by the prior probabilities (i.e., the latent class sizes).

To show this analytically, consider a simple example with two items that require two skills, with the first item requiring the first skill, α_1, and the second item requiring both skills, which enter as the product $\alpha_1 \times \alpha_2$ (cf. Equation 8); note that α_1 appears in both items, similar to Skill 7 in the fraction subtraction Q-matrix, whereas α_2 only enters in the interaction, as does Skill 3 in the fraction subtraction Q-matrix. It is then simple to show that for examinees who get both items incorrect, the posterior probability of the skill vector (0, 1) for Skills 1 and 2 will have higher probability than the vector (0, 0) if the prior probability of the second skill is greater than 0.5. In other words, examinees who get both items incorrect will be classified as having Skill 2 (given a prior greater than .5), as found for Skill 3 above. To see this, note that the posterior probability that both $\alpha_1 = 0$ and $\alpha_2 = 0$ for examinees who got both items incorrect is

$$p(\alpha_1 = 0, \alpha_2 = 0 | Y_1 = 0, Y_2 = 0) = \frac{p(\alpha_1 = 0)p(\alpha_2 = 0) \prod_j p(Y_j = 0 | \alpha_1 = 0, \alpha_2 = 0)}{\sum_\alpha p(\alpha_1)p(\alpha_2) \prod_j p(Y_j = 0 | \alpha_1, \alpha_2)},$$

where the sum is over the values of α (i.e., 0 and 1) and $j = 1, 2$ for the two responses (i.e., Y_1 and Y_2). The use of the product of the latent class sizes in the numerator follows from the independence assumption for skills, and similarly, the use of the product of conditional response probabilities follows from the assumption of independence of the responses given α. The posterior probability of (0, 1) for examinees who got both items incorrect is

$$p(\alpha_1 = 0, \alpha_2 = 1 | Y_1 = 0, Y_2 = 0) = \frac{p(\alpha_1 = 0)p(\alpha_2 = 1) \prod_j p(Y_j = 0 | \alpha_1 = 0, \alpha_2 = 1)}{\sum_\alpha p(\alpha_1)p(\alpha_2) \prod_j p(Y_j = 0 | \alpha_1, \alpha_2)}.$$

Note that the denominators of the above two equations are equal, and so only the numerators differ. Next, it is important to recognize that because of the design of the Q-matrix, the product terms in the numerators are also equal, that is, $\prod_j p(Y_j = 0|\alpha_1 = 0, \alpha_2 = 0) = \prod_j p(Y_j = 0|\alpha_1 = 0, \alpha_2 = 1)$. This occurs because (a) $p(Y_1 = 0|\alpha_1 = 0, \alpha_2 = 0) = p(Y_1 = 0|\alpha_1 = 0, \alpha_2 = 1)$, because $p(Y_1)$ only depends on α_1, and because (b) $p(Y_2 = 0|\alpha_1 = 0, \alpha_2 = 0) = p(Y_2 = 0|\alpha_1 =$

0, $\alpha_2 = 1$), because $p(Y_2)$ depends on α only through the product $\alpha_1 \times \alpha_2$, and so if $\alpha_1 = 0$, then there is no difference in the conditional response probabilities across $\alpha_2 = 0$ and $\alpha_2 = 1$. Thus, given that the denominators of the above probabilities are equal, and the product terms in the numerators are also equal, it follows that the two posterior probabilities will differ solely depending on the product $p(\alpha_1)\,p(\alpha_2)$ in the numerator. If $p(\alpha_2 > .5)$, then $p(\alpha_1 = 0)\,p(\alpha_2 = 1) > p(\alpha_1 = 0)\,p(\alpha_2 = 0)$, and so the posterior probability of $p(0, 1|0, 0)$ will be larger than $p(0, 0|0, 0)$. It can further be shown that $p(\alpha_2 = 1|Y_1 = 0, Y_2 = 0)$ will tend to be greater than $p(\alpha_2 = 0|Y_1 = 0, Y_2 = 0)$ for typical values of the slip and guess parameters. Thus, examinees who get both items incorrect will be classified as having Skill 2 if its prior probability is greater than .5, and so the test in essence offers no information beyond the prior.

The above results account for the findings shown in Table 4. In particular, for the RDINA model, examinees with total scores of zero are classified as having all of the other skills, except Skill 7, simply because the prior probabilities of the other skills are greater than .5, as shown in Table 3. Table 4 shows that the higher order RDINA model ameliorates the classification problem to some extent; however, it does so only for skills that load significantly on the higher order factor, with problems appearing for those that do not, with Skill 3 being a case in point. Restricting the a_k parameters to be equal across the skills helps (with respect to Skill 3); however, the simulation presented in the next section suggests that the restriction could be masking a misspecification of the Q-matrix.

Note that problems such as the above can be avoided if, for each of the skills, there are items that load solely on that skill (given that CDMs are a type of discrete factor model); this type of Q-matrix has been used in some simulations (e.g., Rupp & Templin, 2008b). However, it may frequently not be possible to arrive at this type of Q-matrix with items such as those used in the fraction subtraction test, which ''has been established as a realistic example for the implementation of cognitive diagnosis'' (Henson et al., 2009, p. 204), or in other real-world tests, because the items tend to require multiple skills. Thus, although a design where each skill is indicated by a single item, at least for some items, is desirable, it may not be possible to implement in many situations. In those cases, one must be concerned with the classification problem discussed here.

In sum, there are potential problems with classifications obtained from the DINA model and higher order models, depending on the Q-matrix specification. In particular, the way in which the skills enter the model, as interaction terms as shown in Equation 8, can affect the utility of classifications obtained for some examinees. The test might, for example, offer little or no information about some of the skills for some examinees, beyond the prior probabilities; the fraction subtraction data present a case in point. Although the fraction subtraction data have been analyzed many times, this issue has never been noted in prior research, to my knowledge. The results suggest that classification problems might also arise for other CDMs, depending on the Q-matrix specification, which is an issue that needs to be examined in future research. Other limitations of classifications obtained from CDMs have been noted by Gierl and Cui (2008) and Leighton (2008).

The Q-matrix and Latent Class Sizes

This section notes that there is a potential ambiguity with respect to the meaning of $\alpha = 1$ in the DINA (and HO-DINA) model if the Q-matrix is misspecified. In particular, note that for a correctly specified Q-matrix, $\alpha = 1$ indicates the presence of a skill. However, for an incorrectly specified Q-matrix, $\alpha = 1$ could instead indicate that the skill should not have been included, in that setting $\alpha = 1$ for every examinee simply removes the skill from the model, just as the Q-matrix removes a skill by raising it to the power $q = 0$, so that $\alpha^0 = 1$. This suggests that the finding of a latent class size estimate near unity, as found above for the fraction subtraction data, could simply indicate that the skill should not be included.

The above conjecture can be examined in several ways. For example, for the fraction subtraction data, suppose a ninth skill (α_9) is arbitrarily included in some items, say Items 2, 9, and 14. The reader can verify that for a fit of the model with an extra skill, the estimate of the latent class size for Skill 9 for the RDINA model is .99 ($SE = 0.01$, using PME), for HO-RDINA is .86 ($SE = 0.05$), and for RHO-RDINA is .86 ($SE = 0.03$). Note that the results for the RDINA model clearly suggest that the skill is not necessary, in that the class size estimate for Skill 9 is close to unity, whereas the higher order models do not capture this; the simulation presented next shows a similar result.

The effects of adding a nonnecessary skill can also be examined with simulations. For example, data for four skills and 15 items were generated using the Q-matrix given in Rupp and Templin (2008b; see their Table 2), using values of f_j, d_j, and b_k similar to those found here for the fraction subtraction data. The data were generated according to the RHO-RDINA model with a common value of a_k of 3.5, similar to that found for the fraction subtraction data (data generated according to the RDINA and HO-RDINA models were also examined and gave similar results). Note that the purpose here is not to examine parameter recovery (which the simulation showed was quite good), given that a number of prior simulation studies have already shown that parameter recovery is good for the DINA and (restricted) higher order DINA models (e.g., de la Torre & Douglas, 2004, 2008; Templin et al., 2008) but rather is to examine the effects of including an irrelevant skill on estimates of the latent class sizes.

For a fit of the RHO-RDINA model, with the correct Q-matrix, the estimates of b_k (averaged over the 100 replications) were −0.38 (0.25), −0.07 (0.28), 0.24 (0.31), and 0.57 (0.31), which are quite close to the population values of −0.405, −0.080, 0.241, and 0.575; the estimate of a_k was 3.49 (0.30), which is close to the population value of 3.5; estimates of the latent class sizes (derived from a_k and b_k) were .46, (0.02), .49 (0.02), .52 (0.02), and .55 (0.02). This shows that for the RHO-RDINA model, the parameters for the higher order part of the model are well recovered when the Q-matrix is correctly specified (the other parameters were also well recovered).

Next, the RHO-RINA model was fit with a misspecified Q-matrix, and in particular a fifth skill was arbitrarily added to four of the items (Items 3, 7, 12, and 13). Estimates of the five latent class sizes (and *SE*s) were now .47 (0.02), .49 (0.02), .52 (0.02), .56 (0.02), and .85 (0.04). Note that the estimates for the first four skills are close to the values found for the correctly specified model, given above, which shows that adding an unnecessary skill has little effect on the estimates of the other class sizes. It is interesting to note that the same result was found in a study of the rule space model by Im (2007), in that there was zero to small positive bias for the other class size estimates when an irrelevant skill was included (Im did not examine the estimated class size for the nonnecessary skill). The main result of interest here is that the estimate of the latent class size for the nonnecessary skill, .85, is large, but taken together with the *SE* (0.04), it appears to be less than unity.

In contrast, fitting the RDINA model (with independent skills) to the RHO-RDINA-generated data, again with a fifth skill, gave estimates of the latent class sizes of .51 (0.02), .60 (0.02), .66 (0.02), .76 (0.02), and .99 (0.01). This shows that incorrectly fitting the RDINA model to RHO-RDINA generated data gives estimates of the latent class sizes that are biased upwards, in that the latent class size estimates are all larger; it is important to note that this is consistent with the larger class sizes for the RDINA model as compared to the RHO-RDINA found above for the analysis of the fraction subtraction data, as also found by de la Torre and Douglas (2004). It is also interesting to note that the results for the RDINA model clearly suggest that Skill 5 is not necessary (in light of the observation given above), in that the estimate of the class size is close to unity.

Table 5. Information Criteria for Various Models: Fraction Subtraction Data, $n = 536$

Model	No. of parameters	BIC	AIC
RDINA	48	9414.2	9208.5
HO-RDINA	56	9195.1	8955.2
HO-RDINA (no a3)	54	9190.8	8959.5
RHO-RDINA	49	9165.8	8955.9
RHO-RDINA (no a3)	48	9162.2	8956.5

Note: BIC = Bayesian information criterion; AIC = Akaike information criterion; RDINA = reparameterized DINA (deterministic input, noisy "and"); HO = higher order; RHO = restricted higher order.

Finally, a fit of the HO-RDINA model, again with a fifth skill, gave latent class size estimates of .47 (0.02), .49 (0.02), .53 (0.03), .56 (0.04), and .86 (0.07). Once again, the class size estimates of the first four skills are still accurate when an irrelevant skill is included. The class size estimate for the fifth skill is large (with a large standard error). It is interesting to note that the RDINA model (and to some extent the HO-RDINA model) clearly picks up that the fifth skill is not necessary, in that its latent class size estimate is close to unity, whereas the RHO-RDINA model does not, even though the data were generated according to the RHO-RDINA model (with four, not five, skills). This suggests that the restricted higher order model of de la Torre and Douglas (2004) might be ''over-smoothing'' the latent class size estimates, and so it misses evidence of the inclusion of an irrelevant skill (note that the restriction was also rejected by the LR test given above). The results also show that, in addition to the higher order models (e.g., HO-RDINA and RHO-RDINA), it is informative to fit the simple RDINA model (i.e., with independence).

The above simulations have implications for the results shown above for the analysis of the fraction subtraction data. In particular, note that for fits of both the RDINA and HO-RDINA models, the estimated latent class size for Skill 3 was close to unity. Second, with respect to classification, every examinee was classified as having Skill 3 for both models. Furthermore, the estimate of a_3 for the HO-RDINA model was small and not significant, and so the skill appears to have little or no association with the other skills. Together, these results suggest that the Q-matrix might be misspecified, in that Skill 3 should not be included (or the test provides no information about the skill).

The best way to study misspecification of the Q-matrix is to attempt to design items that better measure the skill in dispute and then test the items on a new sample (if one believes the skill is actually needed). Barring that, it is informative to examine relative fit statistics for models with and without the skill, an approach that has also been used in exploratory studies (e.g., Cen, Koedinger, & Junker, 2005; also see Barnes, Bitzer, & Vouk, 2005). Table 5 shows information criteria, Bayesian information criterion (BIC) and Akaike information criterion (AIC) (see Agresti, 2002; Dayton, 1998), for higher order RDINA models with and without Skill 3. The table shows that BIC is smallest for the RHO-RDINA model without Skill 3, whereas AIC is smallest for the HO-RDINA model, but is close in magnitude to the values found for the RHO-RDINA models with and without Skill 3. Based on these results, one could argue for the RHO-RDINA model without Skill 3 as being the most parsimonious model, as indicated by the BIC (cf. Kang & Cohen, 2007; Li, Cohen, Kim, & Cho, 2009); however, the view here is that decisions about the Q-matrix should not be based on fit statistics alone, in that validity studies and other evidence are needed. The take-home message is simply that the finding of a class size estimate for a skill (or skills) near unity should raise questions about the skill.

Conclusion

The present article offers a logistic latent class version of the DINA model and shows that it is simple to implement in standard software for latent class analysis. Posterior mode estimation is used as a simple approach to boundary problems, which are often encountered in latent class analysis. An application of the DINA model to the widely analyzed fraction subtraction data shows that there are some neglected issues with respect to the classification of examinees and the latent class sizes. It is shown that the problems are largely associated with the specification of the Q-matrix; of course, misspecification of other aspects of the model could also play a role (cf. Henson et al., 2009).

The present results have implications about the diagnostic utility of the DINA model and higher order extensions, as well as other CDMs, in that they show that classifications obtained from the models can be heavily affected by the Q-matrix specification. This raises serious issues, in that correct specification of the Q-matrix is not an easy task, given that one has to specify the nature of the skills that are involved for a particular set of items, the number of skills involved, the combination of skills that affect each item (i.e., the Q-matrix), and an appropriate higher order structure for the skills (and there are many possibilities). One also has to keep the labeling fallacy in mind, in that the ''true skills'' may not correspond to the labels given by the researcher, and changing one or more skill labels can completely change the Q-matrix. The fraction subtraction data present a case in point about difficulties with respect to specifying a Q-matrix, in that even for these 20 simple items, it is still debatable (after 20 years) as to the correct specification of the Q-matrix; in fact, virtually every researcher who has considered the fraction subtraction data has suggested different modifications of the Q-matrix (e.g., de la Torre, 2009; de la Torre & Douglas, 2004, 2008; Henson et al., 2009; C. Tatsuoka, 2002; K. K. Tatsuoka, 1990). This might limit the utility of CDMs in practice, in that new items are needed on a regular basis in real-world tests.

In sum, it is shown that if one finds latent class size estimates for one or more skills that are close to unity, as for the fraction subtraction data, then the possibility that the Q-matrix has been misspecified (with respect to the inclusion of an irrelevant skill) should be considered. The present results also show that one needs to pay close attention to the way the skills enter the Q-matrix. For example, if they only enter as interaction terms and do not have direct indicators, then classifications obtained for those skills might only reflect the prior probabilities, in which case the test offers no information beyond the priors. A simple and useful way to examine this possibility, suggested here, is to perform simulations with the particular Q-matrix under question, which provides information about the sensitivity of the classifications to the priors, as shown above. It also appears to be useful to fit the independence DINA model, in addition to higher order models, in that the independence model in some cases reveals when a skill is irrelevant, as shown in the above simulations (via the finding of estimates of latent class sizes near unity). The problems noted here, such as effects of the Q-matrix specification on the classification of examinees and estimation of the latent class sizes, require closer attention in future research.

Appendix A

Latent Gold 4.5 Syntax for the HO-RDINA Model

```
Variables
 dependent i1 cumlogit, i2 cumlogit, i3 cumlogit, i4 cumlogit, i5 cumlogit,
 i6 cumlogit, i7 cumlogit, i8 cumlogit, i9 cumlogit, i10 cumlogit, i11 cumlogit,
 i12 cumlogit, i13 cumlogit, i14 cumlogit, i15 cumlogit, i16 cumlogit,
 i17 cumlogit, i18 cumlogit, i19 cumlogit, i20 cumlogit;
Latent
 a1 ordinal 2 score = (0 1), a2 ordinal 2 score = (0 1),
 a3 ordinal 2 score = (0 1), a4 ordinal 2 score = (0 1),
 a5 ordinal 2 score = (0 1), a6 ordinal 2 score = (0 1),
 a7 ordinal 2 score = (0 1), a8 ordinal 2 score = (0 1),
 theta, continuous;
Equations
 a1-a8 <- 1 + theta;
Note: to fit the RHO-RDINA model, replace the above with a1-a8 <- 1 + (a)theta
 i1 <- 1 + a4 a6 a7;
 i2 <- 1 + a4 a7;
 i3 <- 1 + a4 a7;
 i4 <- 1 + a2 a3 a5 a7;
 i5 <- 1 + a2 a4 a7 a8;
 i6 <- 1 + a7;
 i7 <- 1 + a1 a2 a7;
 i8 <- 1 + a7;
 i9 <- 1 + a2;
 i10 <- 1 + a2 a5 a7 a8;
 i11 <- 1 + a2 a5 a7;
 i12 <- 1 + a7 a8;
 i13 <- 1 + a2 a4 a5 a7;
 i14 <- 1 + a2 a7;
 i15 <- 1 + a1 a7;
 i16 <- 1 + a2 a7;
 i17 <- 1 + a2 a5 a7;
 i18 <- 1 + a2 a5 a6 a7;
 i19 <- 1 + a1 a2 a3 a5 a7;
 i20 <- 1 + a2 a3 a5 a7;
Note: it is useful to provide starting values for the latent class sizes; see the LG manual.
```

Note: RDINA = reparameterized DINA (deterministic input, noisy “and”); HO = higher order; RHO = restricted higher order.

Appendix B

LEM Syntax for the HO-RDINA Model

lat 9	There are 9 latent variables.
man 20	There are 20 manifest vars.
dim 10 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	Dimensions of the variables.
lab X a b c d e f g h A B C D E F G H I J K L M N O P Q S T U	Lower case = latent class vars
mod X {wei(X)}	X is continuous (theta).
a|X cum(a) {spe(X,1b)}	Model for the latent classes
b|X cum(a) {spe(X,1b)}	"spe" uses a design matrix.
c|X cum(a) {spe(X,1b)}	See the LEM manual.
d|X cum(a) {spe(X,1b)}	
e|X cum(a) {spe(X,1b)}	
f|X cum(a) {spe(X,1b)}	
g|X cum(a) {spe(X,1b)}	
h|X cum(a) {spe(X,1b)}	
A|dfg cum(a) {cov(dfg,1)}	Model for the observed vars
B|dg cum(a) {cov(dg,1)}	Q-matrix shown in Table 1
C|dg cum(a) {cov(dg,1)}	cum(a) for logistic model
D|bceg cum(a) {cov(bceg,1)}	"cov" requires the user to
E|bdgh cum(a) {cov(bdgh,1)}	input the design matrix.
F|g cum(a) {cov(g,1)}	It is given below as "des."
G|abg cum(a) {cov(abg,1)}	
H|g cum(a) {cov(g,1)}	
I|b cum(a) {cov(b,1)}	
J|begh cum(a) {cov(begh,1)}	
K|beg cum(a) {cov(beg,1)}	
L|gh cum(a) {cov(gh,1)}	
M|bdeg cum(a) {cov(bdeg,1)}	
N|bg cum(a) {cov(bg,1)}	
O|ag cum(a) {cov(ag,1)}	
P|bg cum(a) {cov(bg,1)}	
Q|beg cum(a) {cov(beg,1)}	
S|befg cum(a) {cov(befg,1)}	
T|abceg cum(a) {cov(abceg,1)}	
U|bceg cum(a) {cov(bceg,1)}	
rec 536	Next is the design matrix.
des [0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 1	
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1	
0 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	
0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1	
0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1]	
wla fract_out.txt	Write out the latent classes.
sta wei(X) nor(1,8)	X is normally distributed.
dat fraction.txt	

Note: In the data file, recode 0, 1 indicators as 1, 2 for LEM. RDINA = reparameterized DINA (deterministic input, noisy "and"); HO = higher order; RHO = restricted higher order.

Author's Note

Links to software and sample programs are available at the author's website: http://www.columbia.edu/~ld208.

Declaration of Conflicting Interests

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

Funding

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

Notes

1. The data consist of 536 observations; however, a number of articles have reported a sample size of 2144. This arose from a mistake where the original data was simply repeated and stacked four times to give 2144 observations (K. K. Tatsuoka, personal communication, April 2009).
2. Latent class SDT also assumes that there is a continuous underlying variable that is associated with the latent class variable η, which offers some interesting possibilities, but an underlying latent variable need not be assumed. For this reason, the notation f_j is used here in lieu of c_j, because the criterion c_j is usually associated with an interpretation in terms of a continuous underlying distribution and a cut point.
3. Note that Attributes 7 and 8 in Table 10 of de la Torre and Douglas (2004) are switched from what is shown in their Q-matrix (J. de la Torre, personal communication, May 2009).

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

Barnes, T., Bitzer, D., & Vouk, M. (2005, May). *Experimental analysis of the Q-matrix method in knowledge discovery*. Paper presented at the 15th International Symposium on Methodologies for Intelligent Systems 2005, Saratoga Springs, NY.

Cen, H., Koedinger, K., Junker, B. (2005). Learning factors analysis: A general method for cognitive model evaluation and improvement. In M. Ikeda, K. Ashley, & T. Chan (Eds.), *Intelligent Tutoring Systems: 8th International Conference* (pp. 164-175). Berlin, Germany: Springer.

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311-359). New York, NY: Plenum.

Clogg, C. C., & Eliason, S. R. (1987). Some common problems in log-linear analysis. *Sociological Methods and Research, 16*, 8-44.

Dayton, C. M. (1998). *Latent class scaling analysis*. Thousand Oaks, CA: Sage.

DeCarlo, L. T. (2002). A latent class extension of signal detection theory, with applications. *Multivariate Behavioral Research, 37*, 423-451.

DeCarlo, L. T. (2005). A model of rater behavior in essay grading based on signal detection theory. *Journal of Educational Measurement, 42*, 53-76.

DeCarlo, L. T. (2008). *Studies of a latent class signal detection model for constructed response scoring* (ETS Technical Report No. RR-08-63). Princeton NJ: Educational Testing Service.

DeCarlo, L. T. (2010). *Studies of a latent class signal detection model for constructed response scoring II: Incomplete and hierarchical designs* (ETS Technical Report No. RR-10-08). Princeton NJ: Educational Testing Service.

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

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

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

DiBello, L. V., Roussos, L. A., & Stout, W. F. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26. Psychometrics* (pp. 979-1030). Amsterdam, Netherlands: Elsevier.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models* (2nd ed.). New York, NY: Springer-Verlag.

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, IL.

Galindo-Garre, F., & Vermunt, J. K. (2006). Avoiding boundary estimates in latent class analysis by Bayesian posterior mode estimation. *Behaviormetrika, 33*, 43-59.

Galindo-Garre, F., Vermunt, J. K., & Bergsma, W. P. (2004). Bayesian posterior estimation of logit parameters with small samples. *Sociological Methods & Research, 33*, 88-117.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. New York, NY: Chapman & Hall.

Gierl, M. J., & Cui, Y. (2008). Defining characteristics of diagnostic classification models and the problem of retrofitting in cognitive diagnostic assessment. *Measurement, 6*, 263-268.

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

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

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

Im, S. (2007). *Statistical consequences of attribute misspecification in the rule space model* (Unpublished doctoral dissertation). Teachers College, Columbia University, New York, NY.

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.

Kang, T., & Cohen, A. S. (2007). Model selection methods for dichotomous items. *Applied Psychological Measurement, 31*, 331-358.

Leighton, J. P. (2008). Where's the psychology? A commentary on ''Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art.'' *Measurement, 6*, 272-275.

Li, F., Cohen, A. S., Kim, S.-H., & Cho, S.-J. (2009). Model selection methods for dichotomous IRT models. *Applied Psychological Measurement, 33*, 353-373.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). New York, NY: Cambridge University Press.

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

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

Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

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

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

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York, NY: Chapman & Hall.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society, Series C (Applied Statistics), 51*, 337-350.

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

Templin, J., Henson, R., & Douglas, J. (2006, April). *General theory and estimation of cognitive diagnosis models: Using Mplus to derive model estimates*. Paper presented at the 2007 National Council on Measurement in Education training session, Chicago, IL.

Templin, J. L., Henson, R. A., Templin, S. E., & Roussos, L. (2008). Robustness of hierarchical modeling of skill association in cognitive diagnosis models. *Applied Psychological Measurement, 32*, 559-574.

Templin, J. L., & Hoffman, L. (2010). *Obtaining diagnostic classification model estimates using Mplus*. Manuscript submitted for publication.

Vermunt, J. K. (1997). *LEM: A general program for the analysis of categorical data*. Tilburg, Netherlands: Tilburg University.

Vermunt, J. K., & Magidson, J. (2005). *Technical guide for Latent Gold 4.0: Basic and advanced*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K., & Magidson, J. (2007, February). *LG-SyntaxTM user's guide: Manual for Latent Gold 4.5 Syntax Module*. Belmont, MA: Statistical Innovations.

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

von Davier, M. (2009). Some notes on the reinvention of latent structure models as diagnostic classification models. *Measurement, 7*, 67-74.