Article

Applied Psychological Measurement
2014, Vol. 38(6) 464–485

DOI: 10.1177/0146621614533986
apm.sagepub.com

Estimating a Cognitive Diagnostic Model for Multiple Strategies via the EM Algorithm

Yan Huo[1] and Jimmy de la Torre[1]

Abstract

The single-strategy deterministic, inputs, noisy "and" gate (SS-DINA) model has previously been extended to a model called the multiple-strategy deterministic, inputs, noisy "and" gate (MS-DINA) model to address more complex situations where examinees can use multiple problem-solving strategies during the test. The main purpose of this article is to adapt an efficient estimation algorithm, the Expectation–Maximization algorithm, that can be used to fit the MS-DINA model when the joint attribute distribution is most general (i.e., saturated). The article also examines through a simulation study the impact of sample size and test length on the fit of the SS-DINA and MS-DINA models, and the implications of misfit on item parameter recovery and attribute classification accuracy. In addition, an analysis of fraction subtraction data is presented to illustrate the use of the algorithm with real data. Finally, the article concludes by discussing several important issues associated with multiple-strategies models for cognitive diagnosis.

Keywords

cognitive diagnostic models, multiple strategies, DINA model, EM algorithm, estimation

Traditional (i.e., unidimensional) item response models (IRMs) primarily use overall scores to compare and rank examinees along a proficiency continuum. In more recent applications of traditional IRMs to large-scale assessments (e.g., the National Assessment of Education Progress; Lee, Grigg, & Dion, 2007), these models have been used to identify what examinees with varying proficiencies can do differentially by providing exemplar problems along the different points of the proficiency continuum. However, because items in such tests are not purposely designed to be diagnostic in nature, they may not provide a sufficiently informative diagnosis of students' strengths and weaknesses. In contrast, cognitive diagnosis models (CDMs), a family of psychometric models, which can provide score profiles in place of, or possibly in addition to, the overall test scores, have been developed in recent years to assist educational practitioners in evaluating students' mastery or nonmastery of finer grained skills or attributes required for solving problems in a test. Among various CDMs, the deterministic, inputs, noisy ''and'' gate (DINA; Haertel, 1984; Junker & Sijtsma, 2001) model is one of the most popular CDMs that

[1]Rutgers, The State University of New Jersey, New Brunswick, USA

Corresponding Author:
Yan Huo, Department of Educational Psychology, Rutgers, The State University of New Jersey, 10 Seminary Place, New Brunswick, NJ 08901, USA.
Email: yan.huo@gmail.com

have been widely studied (e.g., de la Torre, 2009a; de la Torre & Douglas, 2004; C. Tatsuoka, 2002).

Implicitly, the DINA model, as most CDMs, is a single-strategy model. For the purposes of this article, the DINA model will be referred to as the single-strategy DINA (SS-DINA) model. Recently, the SS-DINA model has been extended to a model called the multiple-strategy DINA (MS-DINA) model to address a more complex situation where examinees can use multiple problem-solving strategies employed in the test (de la Torre & Douglas, 2008). Such situations are not uncommon in education. For example, Fuson et al. (1997) identified three strategies, which were invented by children at elementary schools rather than taught by teachers in class, to solve problems in the domain of multidigit addition and subtraction. These strategies are named as the sequential, the combining-units-separately, and the compensating strategies. The detailed illustrations of these three strategies and their associated four attributes are presented in Appendix A.

In cases where examinees can apply more than one strategy to solve a problem, a single-strategy model such as the SS-DINA model may not adequately capture the complex nature of multiple-strategy phenomena. In these situations, CDMs that can accommodate multiple strategies such as the MS-DINA model are more appropriate. Specifically, the MS-DINA model evaluates whether an examinee fulfills at least one of the possible strategies for solving a problem. Previous research (i.e., de la Torre & Douglas, 2008) has shown that this model can correct the size of the guessing parameter estimate when an alternative strategy is being used.

Objectives

One serious limitation of the current implementation of the MS-DINA model is that it relies on Markov chain Monte Carlo (MCMC; e.g., Carlin & Louis, 2000; Gamerman, 1997; Gelman, Carlin, Stern, & Rubin, 2003) algorithm to estimate its model parameters. Although flexible, MCMC is a computer-intensive and time-consuming estimation algorithm and, thus, can be impractical when dealing with larger data sets. Another limitation of the MS-DINA model as described and estimated by de la Torre and Douglas (2008) is that the joint distribution of the attributes is expressed as a function of a higher order ability. This is a very specific form of the joint distribution of the attributes and, therefore, may not apply to all settings.

To address these two issues, the main purpose of this article is to develop a more efficient estimation procedure, namely, Expectation–Maximization (EM; Dempster, Laird, & Rubin, 1977) algorithm that can estimate the parameters of the MS-DINA model when the joint attribute distribution is allowed to be as general as possible (i.e., saturated). This article also examines the impact of sample size and test length on the SS-DINA and MS-DINA model fit, and the implications of misfit on item parameter recovery and attribute classification accuracy. The remaining sections of the article are laid out as follows. The ''Overview and Background'' section introduces the notations and formulations for the SS-DINA and the MS-DINA models, as well as the joint distributions of the attributes. The EM algorithm for the MS-DINA model is described in the ''Estimation'' section. A simulation study and a real data analysis are given in the ''Simulation Study'' and ''Fraction Subtraction Data Illustration'' sections, respectively. Last, a ''Discussion and Conclusion'' section presents the conclusions of this article.

Overview and Background

Let Y_{ij} be the response of examinee i $(i = 1, \ldots, N)$ to item j $(j = 1, \ldots, J)$, and let $\boldsymbol{\alpha}_i = \{\alpha_{ik}\}$ denote the ith examinee's attribute vector $k = 1, \ldots, K$. The kth attribute (i.e., α_k) can only take binary values, where $\alpha_{ik} = 1$ indicates the presence or mastery of attribute k by examinee i, whereas $\alpha_{ik} = 0$ indicates that examinee i lacks or has not mastered the attribute. Like most CDMs, the DINA model utilizes a $J \times K$ Q-matrix to summarize the required attributes for each item (K. K. Tatsuoka, 1983). The element q_{jk} of the Q-matrix specifies whether attribute k is required to correctly answer item j. If a correct answer to item $SE(1-\hat{s})$ requires attribute k, then $q_{jk} = 1$; if it does not, $q_{jk} = 0$.

SS-DINA Model

In the SS-DINA model, the entire examinee population is partitioned into two latent groups at the level of each item. In one group, examinees possess all the attributes required for the item, and in the other group, examinees lack at least one required attribute. Let η_{ij} represent whether the ith examinee has all the required attributes for the jth item. This can be computed as $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$, where $\eta_{ij} = 1$ indicates that examinee i has mastered all the required attributes for item j, and $\eta_{ij} = 0$ that examinee i lacks at least one of the required attributes for the item.

Using the DINA model, the conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{\alpha}_j$ can be specified. The probability that examinee i with the attribute vector $\boldsymbol{\alpha}_i$ will answer the jth item correctly is given by,

$$P(Y_{ij} = 1|\alpha_{ik}) = P(Y_{ij} = 1|\eta_{ij}) = (1 - s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}}. \tag{1}$$

In this model, parameter $s_j = P(Y_{ij} = 0|\eta_{ij} = 1)$ denotes the probability that examinees who have mastered all the required attributes will slip and incorrectly answer the jth item; $g_j = P(Y_{ij} = 1|\eta_{ij} = 0)$ denotes the probability that examinees who lack at least one required attribute will correctly guess the answer to the jth item. For an item to be diagnostically informative, $1 - s_j$ should be greater than g_j to indicate that more capable examinees have a higher probability of answering the item correctly relative to their less capable counterparts. There are many factors that can cause examinees who have all the required attributes to slip. Examples of these factors include fatigue or momentary distraction, and unfamiliarity with the testing situations or an item format. Similarly, although labeled as the guessing parameter, g_j carries a broader meaning. In addition to cases where examinees who do not have all the required attributes may luckily guess the answer to the problem, g_j can also refer to situations where examinees use a different strategy not included in the Q-matrix. The SS-DINA model does not distinguish between these two types of ''guessing'' (de la Torre, 2009b; de la Torre & Douglas, 2004).

MS-DINA Model

As an extension of the SS-DINA model, the MS-DINA model proposed by de la Torre and Douglas (2008) offers the possibility that problems can be solved in multiple ways, and those alternative strategies can be decoupled from lucky guessing. To do so, the MS-DINA model requires constructing M distinct Q-matrices, $Q_1, Q_2, \ldots, Q_M$, corresponding to the M different strategies. The definition of η_{ij} is accordingly adapted to incorporate the M strategies. Let $\eta_{ijm} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jkm}}$, for $m = 1, 2, \ldots, M$, denote whether examinee i has mastered all the required attributes required for item j as specified by strategy m. Given the $\eta_{ij1}, \eta_{ij2}, \ldots, \eta_{ijM}$, η_{ij} is determined as

$$\eta_{ij} = \max\{\eta_{ij1}, \eta_{ij2}, \ldots, \eta_{ijM}\}. \tag{2}$$

In other words, η_{ij} is 1 if examinee i satisfies the attribute requirements of at least one of the M strategies.

The item response function of the MS-DINA model is the same as Equation 1 once η_{ij} is established through Equation 2. It has been noted by de la Torre and Douglas (2008) that this model assumes that g_j and s_j are constant across the different strategies, and the application of each strategy is equally difficult.

The item response function for the DINA models assumes that examinees belonging to the same group with respect to item j have the same probability of answering the item correctly. In particular, examinees in the group where $\eta_{ij} = 0$, regardless of their number or type of deficiencies are assumed to have the same probability (i.e., g_j) of answering item j correctly.

The formulation of the item response function demonstrates two advantages associated with the SS-DINA and MS-DINA models. First, the DINA models are parsimonious because only two parameters (i.e., g_j and s_j) are needed to be estimated for each item. Second, the DINA models offer a conjunctive interpretation of cognitive diagnosis in that successfully solving a problem requires all the required attributes, and students lacking any, even a single attribute, are not expected to successfully fulfill the task. This conjunctive interpretation is more appropriate to explain the educational testing data compared with other CDMs (Huebner, Wang, & Lee, 2009).

Joint Distribution of Attributes

In addition to the item response function given in the previous section, the DINA models also need the specification of the joint distribution of latent attribute patterns, each of which represents a unique latent class. The most general and fundamental formulation of the joint distribution of attributes allows for all the possible latent classes. Assuming the total number of attributes is K, there are 2^K possible latent attribute patterns available. A model for the joint distribution of the attributes with this complete specification is called a saturated model (de la Torre, 2009b) and is considered to be unstructured (de la Torre, Hong, & Deng, 2010). Although very general, in some cases where the number of attributes is fairly large, the number of attribute patterns can be enormous resulting in a saturated model that may be computationally inefficient. Therefore, a special case of the saturated form, named as the higher order formulation specifying a relationship between a higher order continuous latent trait θ and the discrete attributes α_k was proposed (de la Torre & Douglas, 2004). The DINA model based on the higher order formulation is called the higher order DINA (HO-DINA) model.

Estimation

To date, the estimation of the DINA model typically uses two types of model parameter estimation algorithms. One is the EM algorithm for the DINA model with the saturated latent class specification (e.g., de la Torre, 2009b), and the other is the MCMC algorithm for the HO-DINA model (e.g., de la Torre & Douglas, 2004). Although both algorithms were originally developed for the SS-DINA model, de la Torre and Douglas (2008) have extended the MCMC algorithm for the MS-DINA model with a higher order specification of the joint distribution. This article aimed to adapt the EM algorithm for the MS-DINA model with the saturated joint attribute distribution. The estimation algorithm for the MS-DINA model differs from that of the SS-DINA model in that it needs to be able to handle multiple strategies and incorporate the M different Q-matrices. To accomplish this, the EM algorithm for the SS-DINA model by de la Torre (2009b) was modified to meet the multiple-strategy requirement. The necessary steps of the algorithm that were adapted for the multiple-strategy purpose have been highlighted in Appendix B. Readers interested in additional technical details are referred to the article by de la Torre.

The MS-DINA EM algorithm is a straightforward extension of the SS-DINA EM algorithm. The MS-DINA EM algorithm is a two-stage process. The first stage determines the appropriate η_{ij} among multiple strategies. Multiple η_{ij}s are calculated based on the attribute pattern and the Q-matrix specification. Once η_{ij} is determined, the second stage involving the EM estimation procedure is essentially the same as the SS-DINA EM algorithm documented by de la Torre (2009b). For this reason, although the uniqueness of the attribute classification cannot always be guaranteed due to the more complex Q-matrix specifications, the EM algorithm for the MS-DINA model is expected to perform similarly as the EM algorithm for the SS-DINA model with respect to the convergence of the item parameter estimates. For more technical details of the EM algorithm for estimating the DINA model parameters and their corresponding standard errors, refer to de la Torre (2009b). The estimation code for the MS-DINA model was written in Ox (Doornik, 2009).

Simulation Study

Equipped with the newly adapted EM algorithm, a simulation study was designed to examine the performance of the algorithm on the MS-DINA model calibration. The simulation study also aimed to investigate the impact of sample size and the test length on the SS-DINA and MS-DINA model fit, and the implications of misfit on item parameter recovery and attribute classification accuracy. In addition, given that previous research (e.g., de la Torre & Douglas, 2008) has been conducted through the MCMC algorithm for the SS-DINA and MS-DINA models, the article also compared the current simulation results obtained through the EM algorithms for the SS-DINA and MS-DINA models with the previous simulation results using the MCMC algorithms for both models in de la Torre and Douglas.

Design

The simulation study examined four factors: generating or true model (SS-DINA and MS-DINA models), fitted model (SS-DINA and MS-DINA models), sample size (N = 500, 1,000, and 2,000), and test length ($J = 10$ and 20). For each of the 12 possible conditions across the generating models, sample sizes and test lengths, 100 data sets were generated. All the generated data sets were fitted using the SS-DINA and MS-DINA models.

To generate the item response data, the examinee's parameters θ and $\boldsymbol{\alpha}_k$ were sampled from $N(0,1)$ and Bernoulli $\left(\{1+\exp[-1.7\lambda_{1k}(\theta-\lambda_{0k})]\}^{-1}\right)$, respectively, where $\lambda_{0k}=-1$ and $\lambda_{1k}=1$ for all the attributes. The parameters of both DINA models were $g=0.2$ and $s=0.2$ across all the items. Table C1 shown in Appendix C gives the Q-matrix corresponding to Strategies A and B in the conditions of $J=20$. A subset of the Q-matrix for $J=20$ (i.e., denoted by *) was used as the Q-matrix in the condition $J=10$. Strategy A was used with the SS-DINA model; Strategies A and B were used jointly with the MS-DINA model. As a result, the number of attributes K is equal to five for the SS-DINA model, and seven for the MS-DINA model. Among the seven attributes, Attributes 3, 4, and 5 are commonly used in Strategies A and B. The preceding parameters for data generation and Q-matrices for $J=20$ are identical to those used in de la Torre and Douglas (2008).

Table 1. Summary of MADs of Estimates of g and $1 - s$. Across the 24 Simulation Conditions.

True model		SS-DINA				MS-DINA			
Fitted model		SS-DINA		MS-DINA		SS-DINA		MS-DINA	
N	J	g	$1 - s$	g	$1 - s$	g	$1 - s$	g	$1 - s$
500	10	0.010	0.003	0.052	0.006	0.062	0.012	0.036	0.003
	20	0.006	0.002	0.024	0.012	0.073	0.018	0.015	0.002
1,000	10	0.005	0.002	0.039	0.004	0.061	0.013	0.028	0.002
	20	0.003	0.001	0.019	0.011	0.071	0.018	0.010	0.002
2,000	10	0.002	0.001	0.036	0.004	0.056	0.013	0.014	0.002
	20	0.002	0.001	0.016	0.012	0.071	0.018	0.006	0.001

Note. MAD = mean absolute deviation;
SS-DINA = single-strategy deterministic, inputs, noisy "and" gate; MS-DINA = multiple-strategy deterministic, inputs, noisy "and" gate.

Results

Item parameter estimates. The item parameter (i.e., g and $1 - s$) estimates for each item averaged across 100 replications are available for all the 24 simulation conditions. However, due to space constraints and to highlight the estimates of g and $1 - s$ that can be easily compared across different simulation conditions, an aggregate measure at the test level—mean absolute deviation (MAD),[1] the average absolute difference between the true values, and estimates across the J items for g and $1 - s$ have been introduced.

Table 1 shows the results of MAD_g and MAD_{1-s} on all possible combinations of the four simulation factors. The MAD values of the correctly fitted model conditions were consistently smaller than the values of the incorrectly fitted model conditions. These findings held across all possible combinations of N and J. It can also be noted that using the SS-DINA model to fit the SS-DINA data always produced more accurate estimates than using the MS-DINA model to fit the MS-DINA data. One possible reason for this finding is that data generated using the SS-DINA model were less complex than the data generated using the MS-DINA model. Consequently, fitting the SS-DINA model to the SS-DINA data was a relatively straightforward task compared with fitting the MS-DINA model to the MS-DINA data.

Substantial differences in terms of estimation accuracy were observed between the two types of misfit as shown in columns 3 through 7. Using the SS-DINA model to fit the data generated by the MS-DINA model produced less accurate estimates than using the MS-DINA model to fit the SS-DINA data. As will be shown later, the overestimation of the guessing parameter was clearly observed at the item level. Because the SS-DINA model cannot detect the usage of alternative strategies, and treat unexpected correct responses as the product of random guessing, the guessing parameters were always overestimated. The discrepancies between these two models, which were incorrectly fitted, were largely caused by overestimation of the guessing parameter by the SS-DINA model.

Moreover, in examining the results from all possible conditions, it can be seen that test length and sample size were also important factors that affected the parameter estimation accuracy. In general, as expected, larger sample sizes and longer test length yielded more accurate item parameter estimates. However, the impact of test length and sample sizes was more pronounced and complex when incorrectly fitted models were involved, and the magnitudes of the difference varied depending on the nature of the fitted models. For instance, when the data were incorrectly fitted with the MS-DINA model, the shorter test length produced less accurate item

Table 2. Average of Computed $SE(SE_c)$ and Empirical $SE(SE_e)$ When the Fitted Model Is Correct.

		SS-DINA				MS-DINA			
		g		$1-s$		g		$1-s$	
N	J	SE_c	SE_e	SE_c	SE_e	SE_c	SE_e	SE_c	SE_e
500	10	0.04	0.05	0.03	0.03	0.05	0.08	0.03	0.03
	20	0.03	0.04	0.02	0.02	0.04	0.05	0.02	0.02
1,000	10	0.03	0.04	0.02	0.02	0.03	0.06	0.02	0.02
	20	0.02	0.02	0.02	0.02	0.03	0.03	0.02	0.02
2,000	10	0.02	0.03	0.01	0.02	0.02	0.04	0.01	0.01
	20	0.02	0.02	0.01	0.01	0.02	0.03	0.01	0.01

Note. SS-DINA = single-strategy deterministic, inputs, noisy "and" gate; MS-DINA = multiple-strategy deterministic, inputs, noisy "and" gate.

parameter estimates of g than the longer test length. As the sample size increased, the accuracy of estimation of g improved in both test length conditions. In other words, either larger samples or longer test length, or both, can compensate for the estimation inaccuracy caused by misspecification using the MS-DINA model. However, this trend was not observed when the SS-DINA model was incorrectly fitted to the MS-DINA data. Larger sample size and longer test had very limited impact on item parameter estimation when the SS-DINA model was incorrectly fitted.

In addition to the point estimates of g and $1-s$, two types of *SE*s, the computed (i.e., theoretical) *SE* and empirical *SE*, were used to quantify the dispersion of the estimates. The former, the theoretical *SE*, was the average of the standard error of the g and $1-s$ over 100 replications; the latter, the empirical *SE*, was the standard deviation of the g and s parameter estimates across the 100 replications. The means of the two types *SE*s over J items in the conditions where the fitted model was the correct model are displayed in Table 2. Fitting the data with the SS-DINA model generally produced smaller *SE*s, particularly the empirical *SE*s of g, than using the MS-DINA model. This is because fitting the multiple-strategy data with the MS-DINA model is relatively a more complex process than fitting the single-strategy data with the SS-DINA model. The values of the computed *SE*s and the empirical *SE*s showed some differences. In the case of s, they were reasonably close to each other; however, the values of the two *SE*s were more different for g. Across all simulation conditions, the computed *SE*s of g were consistently smaller than the empirical *SE*s. These findings can be explained in part by a well-established theorem, the Cramer–Rao inequality (Kendall & Stuart, 1973), which states that $\mathrm{Var}(\hat{\phi}) \geq \frac{1}{I(\phi)}$, where ϕ refers to any unknown parameter. According to this theorem, the variance of the estimator is at least as high as the inverse of its Fisher information. In this setup, the empirical *SE* and computed *SE* correspond to $\mathrm{Var}(\hat{\phi})$ and $\frac{1}{I(\phi)}$, respectively. This help explains why the empirical *SE*s were larger than the computed *SE*s in the case of the guessing parameter. The sample size and the test length also affected the magnitudes of *SE*s and discrepancies between two types of *SE*s. Shorter test length and smaller sample size yielded relatively larger *SE*s for both *SE* types. As the test length, or sample size, or both increased, the magnitudes of the two types of *SE*s became smaller and their discrepancy became less apparent.

Chosen from the results on all possible simulation conditions, the estimates of g and $1-s$, as well as the two types of *SE*s at the item level in the condition of $N = 2,000$ and $J = 20$ are presented in Tables C2 and C3. Consistent with the aggregated MAD results at the test level, the item parameter estimates obtained in this specific simulation condition showed that fitting the MS-DINA data with the SS-DINA model produced less accurate estimates than did fitting

Table 3. Percent of Correct Attribute Classification of α_{1-5} (100 Replications).

Generating model	Fitted model	$J = 10$			$J = 20$		
		N = 500	*N* = 1,000	*N* = 2,000	*N* = 500	*N* = 1,000	*N* = 2,000
Single	Single	59.27	60.11	60.51	73.74	73.89	74.08
	Multiple	53.23	53.97	54.62	69.07	69.53	70.26
Multiple	Single	51.15	51.84	52.02	57.81	57.88	57.92
	Multiple	51.36	52.62	52.84	64.06	64.37	64.48

the SS-DINA data with the MS-DINA model. More specifically, the results clearly showed that the misspecified SS-DINA model tended to overestimate the g parameter compared with the misspecified MS-DINA model. The same simulation condition was also used by de la Torre and Douglas (2008) in comparing the SS-DINA and MS-DINA model parameter estimation using the MCMC approach. Both studies generated the attributes based on a higher order formulation. Although the EM and MCMC algorithms have different estimation mechanisms and underlying distribution assumptions, as the simulation results shown, both algorithms produced essentially comparable parameter estimates of g and $1 - s$, and the *SE*s were almost of the same magnitudes. Both studies produced similar results in the estimation of the SS-DINA and MS-DINA models.

Attribute classification. Estimated attribute patterns for all examinees were compared with the true attribute patterns to see how accurately the examinees' attribute patterns can be recovered by the different DINA models. Table 3 shows the percent of correct classification for the attribute vector $(\boldsymbol{\alpha}_{1-5})$. Based on Table 3, correct model specification, either the SS-DINA or MS-DINA model, always resulted in higher correct classification rates of the attribute vector than incorrect model specification. In the correct model specification conditions, the SS-DINA model produced higher correct classification rates than did the MS-DINA model. In contrast, the MS-DINA model produced higher correct classification rates than the SS-DINA model when misspecified models were involved. When test length was longer, the classification rates substantially increased, especially when the model is correctly specified. However, sample size had little positive impact on increasing the correct classification rates; the classification rates were quite comparable across different sample sizes.

The individual attribute classification rates shown in Tables C4 and C5 were generally consistent with the attribute vector rates. When models were correctly specified, the SS-DINA model always yielded higher correct classification rates for α_1 and α_2 than did the MS-DINA model. In the incorrectly fitted model condition, the percentages of correct attribute classification for α_3, α_4, and α_5 were lower using the MS-DINA model than those using the SS-DINA model when $J = 10$. Their discrepancies became negligible when $J = 20$. In the correctly fitted model condition, using the MS-DINA model generally produced higher correct attribute classification rates for α_3 to α_5 compared with those based on the SS-DINA model with the only exception of α_4 in the condition $J = 10$. Such results can be contributed to the fact that Attributes 3 to 5 are commonly shared by both strategies, and thus are measured more times in the multiple-strategy than in the single-strategy case. Test length was an important factor that affected the correct attribute classification rates. When $J = 10$, the SS-DINA model was always superior to the MS-DINA model when the data were generated using the SS-DINA model. Multiple-strategy data were more complex, and a relatively short test may not provide sufficient information for the MS-DINA model to be more useful. As the test length increased

$(J = 20)$, the MS-DINA model can estimate the attributes with sufficient accuracy, resulting in correct classification rates that were comparable.

MS-DINA model can provide additional information about attribute classification of α_6, α_7, and $\boldsymbol{\alpha}_{1-7}$.Table C6 clearly shows that when the MS-DINA model matches the data generating mechanism, the correct classification rates on the α_6, α_7, and $\boldsymbol{\alpha}_{1-7}$ were substantially better. Longer test length facilitated the improvement of correct classification when the MS-DINA model was correctly specified, but it had essentially no effect on correct classification rate when the MS-DINA model was misspecified. The percent of correct attribute classification remained relatively constant across different sample sizes in both the correctly specified and misspecified conditions.

Fraction Subtraction Data Illustration

Data

To further compare the SS-DINA and MS-DINA models, real data involving responses of 536 middle school students to 15 fraction items that can be solved by two strategies were analyzed. After the original data with responses to 20 items were introduced by K. K. Tatsuoka (1987, 1990), a subset of the data with 15 items were analyzed in some studies (de la Torre, 2009b; de la Torre & Douglas, 2004, 2008; Mislevy, 1996; C. Tatsuoka, 2002). For these data, students solved the mixed number subtraction problems using either Strategy A or B. The major difference between Strategies A and B lies in the different ways students used the strategies to deal with mixed numbers. Students who used Strategy A separated mixed numbers into the whole number and fractional parts before performing subtraction on each part. In contrast, students who used Strategy B converted mixed numbers to improper fractions prior to the subsequent subtraction operation. A total of seven attributes were shared by the two strategies. These seven attributes originally described by Mislevy (1996) are summarized in Table C7. The attributes for Strategy A were numbered as 1, 2, 3, 4, and 5, and for Strategy B, 1, 2, 5, 6, and 7. The Q-matrices for both strategies are given in Table C8.

Results

Item parameter estimation. The SS-DINA (assuming the use of Strategy A only) and MS-DINA (assuming the use of both Strategies A and B) models have been used to fit the fraction subtraction data, and the model parameter estimates are summarized in Table 4. As expected, the results of parameter estimates in the two DINA models, as well as their corresponding theoretical *SE*s, were very similar, but not necessarily identical to the results reported by de la Torre and Douglas (2008) using the MCMC algorithm. de la Torre (2009b) enumerated two major distinctions between the EM for the DINA model with the saturated attribute distribution and the MCMC for the HO-DINA model. First, the two models employ different joint distributions of the attributes. The joint distribution of the attributes in the DINA model with the saturated form is based on a multinomial distribution, whereas the HO-DINA model defines the distribution of the attributes to conditionally independent given a higher order latent proficiency. The former subsumes the latter. Second, their estimates are based on different measures of central tendency, specifically, the mode for the EM algorithm and the mean for the MCMC algorithm. As a result, the two algorithms may not yield identical results. For example, the guessing parameter estimates for Items 3 and 8 based on the MS-DINA model were slightly higher than those based on the SS-DINA model through the current EM algorithm. Such small discrepancies were not observed between the two models using MCMC. Despite the minor differences between the

Table 4. Parameter Estimates for the Fraction Subtraction Data Based on the SS-DINA and MS-DINA Model.

	Fitted DINA model							
	SS-DINA				MS-DINA			
Item	$\hat{g}$	$SE(\hat{g})$	$\hat{s}$	$SE(\hat{s})$	$\hat{g}$	$SE(\hat{g})$	$\hat{s}$	$SE(\hat{s})$
1	0.00	0.09	0.28	0.02	0.00	0.09	0.27	0.02
2	0.21	0.02	0.12	0.02	0.20	0.02	0.12	0.02
3	0.13	0.04	0.04	0.01	0.18	0.04	0.04	0.01
4	0.12	0.02	0.13	0.03	0.11	0.02	0.17	0.03
5	0.30	0.05	0.25	0.02	0.33	0.04	0.20	0.02
6	0.03	0.01	0.23	0.03	0.03	0.01	0.22	0.03
7	0.07	0.02	0.08	0.02	0.07	0.02	0.07	0.02
8	0.15	0.04	0.05	0.01	0.19	0.04	0.05	0.01
9	0.08	0.03	0.06	0.01	0.10	0.03	0.06	0.01
10	0.17	0.02	0.07	0.02	0.04	0.02	0.07	0.02
11	0.10	0.03	0.11	0.02	0.11	0.03	0.10	0.02
12	0.03	0.01	0.13	0.02	0.03	0.01	0.14	0.02
13	0.13	0.02	0.16	0.02	0.14	0.02	0.15	0.02
14	0.02	0.01	0.20	0.03	0.03	0.01	0.24	0.03
15	0.01	0.01	0.18	0.02	0.01	0.01	0.17	0.02

Note. DINA = deterministic, inputs, noisy "and" gate; SS-DINA = single-strategy deterministic, inputs, noisy "and" gate; MS-DINA = multiple-strategy deterministic, inputs, noisy "and" gate.

results produced by the two algorithms, the guessing parameter estimate of Item 10 based on the SS-DINA model was substantially larger than that based on the MS-DINA model. This result suggests that the SS-DINA model might have failed to identify the application of an alternative strategy by some examinees and attributed their correct responses to this item to guessing. However, overall the single- and multiple-strategy models produced largely similar parameter estimates.

In addition, to compare the relative model fit between these two models at the test level, the statistics of the $-2\ln(L)$, Akaike information criterion (AIC; Akaike, 1974) and Bayesian information criterion (BIC; Schwarz, 1978) were computed.[2] Based on the comparable values of the $-2\ln(L)$, that is, 6,911.53 for the SS-DINA model versus 6,857.53 for the MS-DINA model, the two models have roughly similar model fit. Taking the model complexity into account, the SS-DINA model (AIC = 7,033.53 and BIC = 7,294.86) provided much better fit than the MS-DINA model (AIC = 7,171.53 and BIC = 7,844.14). For this reason, the SS-DINA model may be preferred for this data set because its relatively simpler formulation did not lead to worse fit compared with the MS-DINA model.

Attribute classification. The SS-DINA model can provide diagnostic information on Attributes 1 to 5; the MS-DINA model can offer examinees with additional diagnostic information on Attributes 6 and 7. The common classification of Attributes 1 to 5 for both the models has been compared. The SS-DINA and MS-DINA models had a high degree of attribute classification agreement in most cases. The common classification rates for the attributes were around 95% or higher, with the exception of Attribute 3, which is 89.74%. In investigating further, it was found that the discrepancy in the Attribute 3 classifications was due to the fact that the proportion of examinees estimated to have mastered Attribute 3 was higher using the SS-DINA model than the MS-DINA model. In addition, the SS-DINA and MS-DINA models had common attribute vector classification for 81% of the students.

Discussion and Conclusion

de la Torre (2009b) discussed the EM and MCMC algorithms for estimating the SS-DINA model parameters. A previous study (de la Torre & Douglas, 2008) implemented the MCMC algorithm for the SS-DINA and MS-DINA models. This article shows that the EM algorithm for the SS-DINA model can be extended to the multiple-strategy situation. The EM algorithm is relatively more efficient. On a desktop computer with 3.0 GHz processor and 1 GB of memory, fitting the MS-DINA model took about 20 seconds when the convergence criterion was set at 0.0001 for the fraction subtraction data set used in this article. For the comparison purposes, the higher order multiple-strategy deterministic, inputs, noisy ''and'' gate (HO-MS-DINA) model MCMC algorithm (de la Torre & Douglas, 2008), which was run on the same computer, took almost 2 hours to analyze the same fraction subtraction data using a single chain of 250,000 iterations. Additional time would be needed to run multiple chains (e.g., four) to examine convergence. In addition to its efficiency, both the simulation and real data analyses showed that the EM algorithm produced the results that were comparable with those obtained using the MCMC algorithm.

Another contribution of this article is the systematic investigation of the impact of model misspecification, sample size, and test length on model parameter estimation and attribute classification for both the SS-DINA and MS-DINA models. A misspecified MS-DINA model always yields more accurate item parameter estimates and has generally better attribute classification accuracy than the SS-DINA model when both of them are misspecified. These findings may have important practical implications because with real data the true underlying model and number of strategies employed are seldom, if at all, known. For this reason, it might be ''safer'' to fit the DINA model using the MS-DINA approach because the MS-DINA model is more robust against model misspecification than the SS-DINA model. However, the authors would like to note that such potential gain from the MS-DINA model is closely contingent on some prerequisites, such as, the availability of correct multiple-strategy Q-matrices, and constructing and validating Q-matrices for multiple strategies can be more challenging than for a single strategy. As such, the advantages of the MS-DINA model may not be realized in situations where additional effort cannot be expended to guarantee that the Q-matrices of the MS-DINA model are correct. In addition, a more comprehensive approach to model-selection should also take into account the complexity of the model vis-à-vis its fit to the data.

The fraction subtraction data example illustrates that both the SS-DINA and MS-DINA models produce comparable item parameters in most cases and that the SS-DINA model provides better model fit than does the MS-DINA model when model complexity is taken into account. Given these findings, the necessity of the MS-DINA model might be called into question. Although it is true that there is no guarantee that the MS-DINA model can outperform the SS-DINA model all the time, the authors would like to argue that statistical criteria (e.g., model fit indices) are not the only standards to evaluate the necessity and utility of more complex models, such as the MS-DINA model. In addition to statistical evaluation, it would be helpful to also pay attention to non-statistical rationales for considering the MS-DINA model.

Multiple-strategies CDMs were developed in response to existing practical needs and their potential applications in the near future. Single-strategy problem solving may be too strict an assumption and may not always be appropriate for all kinds of problems encountered in practical educational settings. School teachers typically teach students more than one way to solve problems; at the same time, students may spontaneously invent multiple strategies on their own. It is therefore reasonable to expect that more mature learners to have greater likelihood of applying multiple strategies to solve problems in more complex and advanced teaching and learning environments. Moreover, the unique feature of CDMs that distinguishes them from traditional item response theory (IRT) lies in their ability to provide refined attribute profiles to diagnose what skills students have or have not mastered. As indicated in the real data illustration, although the MS-DINA model is not necessarily better than the SS-DINA model in the statistical sense, the MS-DINA model can provide attribute profiles that contain two additional attributes than can the SS-DINA model. Such additional information is afforded by using a multiple-strategy model, and represents the type of information one would expect of the cognitively diagnostic assessments.

For future research, it is worthwhile to further investigate whether the substantial difference in theoretical and empirical *SE*s is due to the complexity of the Q-matrix in the simulation study. As indicated by the simulation results shown in a previous study by de la Torre et al. (2010), the structure of the Q-matrix can affect the accuracy of item parameter estimates. In the future, a simpler Q-matrix where items are measured by fewer attributes can be considered to examine its impact on the *SE*s. The current MS-DINA model assumes that the probabilities of the guessing and slip across different strategies are identical. This assumption could be too strict and may not be suitable in many practical situations. A straightforward extension of this article is to develop a CDM that allows for the guessing and slip parameters of the same item to change across strategies. The current MS-DINA model also assumes that the application of each strategy is equally difficult (de la Torre & Douglas, 2008). Therefore, students can apply different strategies from item to item throughout the entire test according to the MS-DINA model. However, this flexibility does not always reflect the truth for all real data applications. In the fraction subtraction data set, it is more reasonable to assume that students have tendency to use one particular strategy over the other. Already noted by de la Torre and Douglas, another extension of the MS-DINA model can incorporate a latent variable to associate individual students with each strategy. Such latent variable can define which strategy is more likely to be chosen than others by each student. It would also be interesting to explore additional variations of the MS-DINA model. For example, a more general version of the MS-DINA model would allow for different numbers of strategies for the different items. This would subsume the current MS-DINA model, and the model where, in the same test, some items can be solved using only a single strategy, and the remaining items using multiple strategies. Developing efficient algorithms to accompany those more extensions of the MS-DINA model would also be necessary.

The main focus of this article is the applicability of the CDMs in multiple-strategies settings. The two Q-matrices used in the fraction subtraction illustration in this article were theoretically derived (Mislevy, 1996), and have been used in the previous and current studies for the purpose of examining multiple-strategy cognitive modeling in fraction subtraction. Although the authors assume that the two Q-matrices are correct for the purpose of this article, they do not necessarily assume that the Q-matrices, in the single or multiple-strategy context, are always correct. However, examining the correctness of Q-matrices is beyond the scope of the current work. As well known, Q-matrix validation is another key component in cognitive diagnostic assessment in addition to diagnostic modeling. For example, as noted by DeCarlo (2011), misspecification of the Q-matrix can have serious consequences on the latent class classification obtained from the specific CDMs. Specification of Q-matrix is usually initiated by test content domain experts, psychometricians have been taking more and more active roles in the Q-matrix validation for several reasons. First, the Q-matrix construction and validation requires tremendous cost in terms of time and human effort. Second, relying only on experts' judgment can lead to Q-matrix construction and validation processes that are highly subjective in nature. Recently, a few research on Q-matrix validation from psychometric perspectives have been undertaken (e.g., Barnes, 2010; de la Torre, 2008; de la Torre & Chiu, 2010; Huo & de la Torre, 2013; Liu, Xu, & Ying, 2012, 2013). However, there have no particular psychometric methods for determining Q-matrices in the multiple-strategy situations thus far. Technically, validating multiple Q-matrices is a straightforward extension of the single Q-matrix validation. However, advancements in Q-matrix construction and validation in a multiple-strategy setting requires theoretically driven judgments that explicitly specify the needs to classify problem-solving skills into more than one strategy. This would be in addition to the subsequent psychometric Q-matrix validation procedures. As such, it is anticipated that constructing and validating multiple Q-matrices in multiple-strategy settings would be more complicated than in single-strategy settings. To the extent that it is worthwhile to incorporate more than one Q-matrix, it would also be necessary to investigate how to validate multiple Q-matrices using psychometric procedures.

Last, the authors want to address the identifiability issue associated with the MS-DINA model in particular, and the CDMs in general. Compared with the SS-DINA model, the MS-DINA model has greater complexity both in terms of the number of attributes involved, and how the latent or ideal response η_{ij} is determined. However, once the ideal response has been determined, both models have the same number of item parameters per item (i.e., two), and the two models have the same degree of estimation complexity. In the experience of the authors, with a mild constraint (i.e., $1 - s_j > g_j$), item parameter identifiability is not an issue with the two models. However, with more attributes and fixed test length, identifiability of the attribute patterns can be more an issue with the MS-DINA model than it is with the SS-DINA model. In general, regardless of the CDMs, the resolution afforded by a particular Q-matrix will determine the extent to which attribute patterns can be clearly distinguished from each other. A complete Q-matrix implies a one-to-one correspondence between the attribute and ideal response patterns. A practical step to determine how well the attribute patterns can be distinguished from each other is to determine the number of equivalence classes implied by the Q-matrix of the test. As recommended by de la Torre and Douglas (2008), one can examine the equivalence classes to determine the extent to which attribute patterns are identifiable when the MS-DINA model is involved. If the number of equivalence classes is relatively small, they suggested simplifying the model, which include using a single strategy. Another strategy that can be recommended is to include additional items that can further split some of the attributes in the same equivalence classes. Some relevant discussions regarding identifiability of the attribute patterns in the MS-DINA model is presented in Appendix D.

In sum, as more potential applications call for more complex model specifications, identification issues may also emerge. Despite many theoretical efforts dealing with identifiability issues on item parameters and attribute patterns, some empirical approaches can be easily implemented in real data analysis to examine model problems with identification. For example, in the fraction subtraction data analysis, real data estimates were used as the generating parameters to simulate item responses. The results showed that the item parameters can be recovered very well.[3] As such, the simulation study provided evidence that model identification is not an issue in fitting the MS-DINA model to the fraction subtraction data.

Appendix A

Illustration of Multiple Problem-Solving Strategies

Fuson et al. (1997) identified three strategies, which were invented by children at elementary schools rather than taught by teachers in class, to solve problems in the domain of multidigit addition and subtraction. The sequential strategy and the combining-units-separately strategy are similar in that they both separate the entire number into two parts (i.e., tens and ones). Unlike the former that adds values to the sum in sequence, the latter performs addition in each part and combines the two parts afterward. However, the compensating strategy seeks to adjust the numbers to simplify the calculation. These three strategies can be redefined for illustration purpose with four attributes, which are given in Table A1.

Table A1. Attributes in Multidigit Addition and Subtraction Domain.

Attribute	Description
1	Basic addition skills
2	Separating the whole number into tens and ones
3	Combining two parts
4	Compensating to make tens

Table A2 shows how these four attributes can be implemented within the three strategies in solving the addition problem 38 + 26 used in the article by Fuson et al. (1997). Clearly, the three strategies share some common attributes and also possess certain unique attributes. For example, they all involve Attribute 1, whereas only Strategy C requires Attribute 4. It is not necessary for children to master all the four attributes to solve the problem correctly. At the minimum, mastering Attributes 1 and 2 (Strategy A) or Attributes 1 and 4 (Strategy C) will allow children to answer this problem correctly. Children who master Attribute 3, in addition to Attributes 1 and 2, have an additional strategy (i.e., Strategy B) at their disposal.

Table A2. Three Strategies in Solving the Problem 38 + 26

Strategy	Steps	Required attribute
A	38 = 30 + 8; 26 = 20 + 6	2
	30 + 20 = 50; 50 + 8 = 58; 58 + 6 = 64	1
B	38 = 30 + 8; 26 = 20 + 6	2
	30 + 20 = 50; 8 + 6 = 14	1
	50 + 14 = 64	3
C	38 + 2 = 40	4
	40 + 24 = 64	1

Appendix B

Some Details of the Expectation–Maximization (EM) Algorithms for the Multiple-Strategy Deterministic, Inputs, Noisy "and" Gate (MS-DINA) Model

For any attribute pattern $\boldsymbol{\alpha}_l$, $l = 1, \ldots, L = 2^K$, the probability of a correct response in Equation 1 can be rewritten as

$$P(Y_{ij} = 1|\boldsymbol{\alpha}_l) = P(Y_{ij} = 1|\eta_{ij}) = (1 - s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}}, \tag{B1}$$

where $\eta_{ij} = 1 - \prod_{m=1}^{M} (1 - I[\alpha_l' \mathbf{q}_{jm} = \mathbf{q}_{jm}' \mathbf{q}_{jm}])$, and $I[\boldsymbol{\alpha}_l' \mathbf{q}_{jm} = \mathbf{q}_{jm}' \mathbf{q}_{jm}]$ is the indicator function that the attribute vector $\boldsymbol{\alpha}_l$ contains the attributes required to solve item j using at least one of the strategies. Once η_{ij} has be determined from $\eta_{ij1}, \ldots, \eta_{ijm}$, the EM algorithm for the MS-DINA model is essentially the same as that for the single-strategy deterministic, inputs, noisy ''and'' gate (SS-DINA) model. If $\boldsymbol{\alpha}_l$ satisfies the requirements of more than one strategy, there is no need to differentiate between which particular strategy was used.

As shown in the article by de la Torre (2009b), the marginalized likelihood of the response data is

$$L(\boldsymbol{Y}) = \prod_{i=1}^{n} L(\boldsymbol{Y}_i) = \prod_{i=1}^{n} \sum_{l=1}^{L} L(\boldsymbol{Y}_i|\boldsymbol{\alpha}_l)p(\boldsymbol{\alpha}_l), \tag{B2}$$

where $L(\boldsymbol{Y}_i)$ is the marginalized likelihood of the response vector of the ith examinee; $L(\boldsymbol{Y}_i|\boldsymbol{\alpha}_l)$ is the conditional likelihood of response vector of examinee i conditioned on the attribute vector $\boldsymbol{\alpha}_l$, $p(\boldsymbol{\alpha}_l)$ is the prior probability of the attribute vector $\boldsymbol{\alpha}_l$, and $L = 2^K$ is the number of possible combinations of K attributes.

Let $\beta_{j\eta}$ denote the item parameters for item j, where $\beta_{j0} = g_j$ and $\beta_{j1} = s_j$. The goal of the maximum marginal likelihood estimation (MMLE) is to find the parameter values that maximize the marginalized likelihood function $L(\boldsymbol{Y})$, or more conveniently, its logarithm form, $l(\boldsymbol{Y}) = \log L(\boldsymbol{Y})$. The first derivative of the $l(\boldsymbol{Y}2)$ with respect to $\beta_{j\eta}$ is

$$\frac{\partial l(\boldsymbol{Y})}{\partial \beta_{j\eta}} = \sum_{i=1}^{N} \frac{1}{L(\boldsymbol{Y}_i)} \sum_{l=1}^{L} p(\boldsymbol{\alpha}_l) \frac{\partial L(\boldsymbol{Y}_i|\boldsymbol{\alpha}_l)}{\partial \beta_{j\eta}}. \tag{B3}$$

After several simplifications, (B3) can be shown to be equal to

$$\begin{aligned}\frac{\partial l(\boldsymbol{Y})}{\partial \beta_{j\eta}} = {} & \frac{\partial g_j}{\partial \beta_{j\eta}} \left[\frac{1}{g_j[1-g_j]}\right] \left[R_j^{(0)} - g_j I_j^{(0)}\right] \\ & + \frac{\partial (1-s_j)}{\partial \beta_{j\eta}} \left[\frac{1}{(1-s_j)s_j}\right] \left[R_j^{(1)} - (1-s_j) I_j^{(1)}\right],\end{aligned} \tag{B4}$$

where $I_j^{(1)}$, the expected number of examinees who satisfy the attribute requirement for item j of at least one of the M strategies, $I_j^{(0)}$, the expected number of examinees who do not satisfy the attribute requirement for item j of any of the M strategies, $R_j^{(1)}$ and $R_j^{(0)}$ are the expected numbers of examinees belonging to $I_j^{(1)}$ and $I_j^{(0)}$, respectively, answering item j correctly. It can be noted that although the same notation is used in this article, $I_j^{(1)}$, $I_j^{(0)}$, $R_j^{(1)}$, and $R_j^{(0)}$ have different interpretations from those in the de la Torre (2009b) article to account for the existence of multiple strategies.

Appendix C

Table C1. The Q-Matrix for $J = 20$.

	Attribute									
	Strategy A					Strategy B				
Item	1	2	3	4	5	3	4	5	6	7
1*	1	1	0	0	0	0	1	0	1	1
2*	1	0	1	0	0	0	0	1	1	1
3*	1	0	0	1	0	0	1	1	0	1
4*	1	0	0	0	1	0	1	1	1	0
5*	0	1	1	0	0	1	0	0	1	1
6	0	1	0	1	0	1	1	0	0	1
7	0	1	0	0	1	1	1	0	1	0
8	0	0	1	1	0	1	0	1	0	1
9	0	0	1	0	1	1	0	1	1	0
10	0	0	0	1	1	1	1	1	0	0
11*	1	1	1	0	0	0	0	0	1	1
12*	1	1	0	1	0	0	1	0	0	1
13*	1	1	0	0	1	0	1	0	1	0
14*	1	0	1	1	0	0	0	1	0	1
15*	1	0	1	0	1	0	0	1	1	0
16	1	0	0	1	1	0	1	1	0	0
17	0	1	1	1	0	1	0	0	0	1
18	0	1	1	0	1	1	0	0	1	0
19	0	1	0	1	1	1	1	0	0	0
20	0	0	1	1	1	1	0	1	0	0

Note. Item with * belongs to the Q-matrix for $J = 10$.

Table C2. DINA Model Parameter Estimates for Data Generated Using the SS-DINA Model (100 replications, N = 2,000, $J = 20$).

	Fitted DINA model											
	Single						Multiple					
			SE						*SE*			
	Estimate		Computed		Empirical		Estimate		Computed		Empirical	
Item	g	$1-s$	g	$1-s$	g	$1-s$	g	$1-s$	g	$1-s$	g	$1-s$
1	0.20	0.80	0.02	0.01	0.02	0.01	0.20	0.80	0.02	0.01	0.02	0.01
2	0.20	0.80	0.02	0.01	0.02	0.01	0.21	0.80	0.02	0.01	0.03	0.01
3	0.20	0.80	0.02	0.01	0.02	0.01	0.20	0.80	0.02	0.01	0.03	0.02
4	0.20	0.80	0.02	0.01	0.02	0.01	0.19	0.80	0.02	0.01	0.03	0.01
5	0.20	0.80	0.02	0.01	0.02	0.01	0.21	0.80	0.02	0.01	0.03	0.01
6	0.19	0.80	0.02	0.01	0.02	0.01	0.21	0.80	0.02	0.01	0.03	0.01
7	0.19	0.80	0.02	0.01	0.02	0.01	0.20	0.80	0.02	0.01	0.03	0.01
8	0.19	0.80	0.02	0.01	0.02	0.01	0.24	0.80	0.02	0.01	0.02	0.01
9	0.20	0.80	0.02	0.01	0.02	0.01	0.24	0.81	0.02	0.01	0.02	0.01
10	0.20	0.80	0.02	0.01	0.02	0.01	0.24	0.80	0.02	0.01	0.02	0.01

(continued)

Table C2. (continued)

	Fitted DINA model											
	Single						Multiple					
			SE						*SE*			
	Estimate		Computed		Empirical		Estimate		Computed		Empirical	
Item	*g*	1 − *s*	*g*	1 − *s*	*g*	1 − *s*	*g*	1 − *s*	*g*	1 − *s*	*g*	1 − *s*
11	0.20	0.80	0.01	0.01	0.01	0.01	0.19	0.80	0.02	0.01	0.01	0.01
12	0.20	0.80	0.01	0.01	0.01	0.01	0.18	0.80	0.02	0.01	0.02	0.01
13	0.20	0.80	0.01	0.01	0.01	0.01	0.18	0.80	0.02	0.01	0.02	0.01
14	0.20	0.80	0.01	0.01	0.01	0.01	0.18	0.80	0.02	0.01	0.02	0.01
15	0.20	0.80	0.01	0.01	0.01	0.01	0.18	0.80	0.02	0.01	0.02	0.01
16	0.20	0.80	0.01	0.01	0.01	0.01	0.19	0.75	0.02	0.01	0.02	0.02
17	0.20	0.80	0.01	0.01	0.01	0.01	0.18	0.81	0.02	0.01	0.02	0.01
18	0.20	0.80	0.01	0.01	0.01	0.01	0.19	0.80	0.02	0.01	0.02	0.01
19	0.20	0.80	0.01	0.01	0.01	0.01	0.19	0.71	0.02	0.01	0.02	0.02
20	0.20	0.80	0.01	0.01	0.01	0.01	0.20	0.75	0.02	0.01	0.02	0.02

Note. SS-DINA = single-strategy deterministic, inputs, noisy "and" gate; DINA = deterministic, inputs, noisy "and" gate.

Table C3. DINA Model Parameter Estimates for Data Generated Using the MS-DINA Model (100 replications, N = 2,000, $J = 20$).

	Fitted DINA model											
	Single						Multiple					
			SE						*SE*			
	Estimate		Computed		Empirical		Estimate		Computed		Empirical	
Item	*g*	1 − *s*	*g*	1 − *s*	*g*	1 − *s*	*g*	1 − *s*	*g*	1 − *s*	*g*	1 − *s*
1	0.21	0.78	0.02	0.01	0.04	0.01	0.19	0.80	0.02	0.01	0.04	0.01
2	0.31	0.80	0.02	0.01	0.03	0.01	0.19	0.80	0.02	0.01	0.03	0.01
3	0.18	0.78	0.02	0.01	0.03	0.01	0.19	0.80	0.02	0.01	0.02	0.01
4	0.19	0.77	0.02	0.01	0.04	0.01	0.19	0.80	0.02	0.01	0.03	0.01
5	0.20	0.78	0.02	0.01	0.03	0.01	0.20	0.80	0.02	0.01	0.02	0.01
6	0.19	0.78	0.02	0.01	0.03	0.01	0.19	0.80	0.02	0.01	0.03	0.01
7	0.29	0.79	0.02	0.01	0.03	0.01	0.19	0.80	0.02	0.01	0.04	0.01
8	0.27	0.79	0.02	0.01	0.03	0.01	0.20	0.80	0.02	0.01	0.02	0.01
9	0.18	0.77	0.02	0.01	0.03	0.01	0.20	0.80	0.02	0.01	0.02	0.01
10	0.15	0.76	0.02	0.01	0.03	0.01	0.20	0.80	0.02	0.01	0.02	0.01
11	0.35	0.79	0.02	0.01	0.02	0.01	0.19	0.80	0.02	0.01	0.04	0.02
12	0.25	0.78	0.02	0.01	0.02	0.01	0.19	0.80	0.02	0.01	0.03	0.01
13	0.33	0.79	0.02	0.01	0.02	0.01	0.19	0.80	0.02	0.01	0.03	0.01
14	0.37	0.80	0.02	0.01	0.02	0.01	0.19	0.80	0.02	0.01	0.03	0.01
15	0.32	0.78	0.02	0.01	0.02	0.01	0.19	0.80	0.02	0.01	0.03	0.01
16	0.22	0.76	0.02	0.01	0.02	0.01	0.20	0.80	0.02	0.01	0.02	0.01
17	0.30	0.79	0.20	0.01	0.02	0.01	0.19	0.80	0.02	0.01	0.03	0.01
18	0.31	0.79	0.02	0.01	0.02	0.01	0.19	0.80	0.02	0.01	0.02	0.01
19	0.30	0.78	0.02	0.01	0.02	0.01	0.20	0.80	0.02	0.01	0.02	0.01
20	0.27	0.77	0.02	0.01	0.02	0.01	0.20	0.80	0.02	0.01	0.02	0.01

Note. DINA = deterministic, inputs, noisy "and" gate; MS-DINA = multiple-strategy deterministic, inputs, noisy "and" gate.

Table C4. Percent of Correct Attribute Classification of α_1 to α_5 (100 replications, $J = 10$).

			Attribute				
N	Generating model	Fitted model	α_1	α_2	α_3	α_4	α_5
500	Single	Single	85.26	85.23	85.09	85.26	85.53
		Multiple	83.93	83.72	82.46	80.72	83.80
	Multiple	Single	80.87	81.02	87.25	81.89	85.83
		Multiple	80.25	80.50	87.16	83.92	86.98
1,000	Single	Single	85.48	85.48	85.70	85.46	85.46
		Multiple	84.47	84.00	82.81	80.77	84.07
	Multiple	Single	81.26	81.82	87.73	82.20	86.42
		Multiple	81.06	81.42	87.86	84.38	87.59
2,000	Single	Single	85.62	85.79	85.90	85.80	85.74
		Multiple	84.69	84.44	83.17	81.48	84.03
	Multiple	Single	81.30	81.73	87.86	82.14	86.40
		Multiple	81.20	81.67	88.31	84.67	87.72

Table C5. Percent of Correct Attribute Classification of α_1 to α_5 (100 replications, $J = 20$).

			Attribute				
N	Generating model	Fitted model	α_1	α_2	α_3	α_4	α_5
500	Single	Single	90.09	90.11	89.91	90.21	90.15
		Multiple	88.82	88.98	88.10	87.74	88.04
	Multiple	Single	83.16	83.53	88.82	87.75	87.87
		Multiple	84.66	84.61	91.59	91.47	91.63
1,000	Single	Single	90.42	90.33	90.58	90.36	90.25
		Multiple	89.27	89.33	88.82	88.34	88.60
	Multiple	Single	83.38	83.51	88.99	88.49	88.08
		Multiple	84.90	84.83	92.04	92.05	91.80
2,000	Single	Single	90.73	90.72	90.70	90.78	90.69
		Multiple	89.60	89.69	89.20	89.00	89.30
	Multiple	Single	83.49	83.69	89.50	88.66	88.29
		Multiple	85.10	85.03	92.45	92.42	92.17

Table C6. Percent of Correct Attribute Classification of α_{6-7} and $\boldsymbol{\alpha}_{1-7}$ of the MS-DINA Model (100 replications).

Condition		SS-DINA data			MS-DINA data		
J	N	α_6	α_7	$\boldsymbol{\alpha}_{1-7}$	α_6	α_7	$\boldsymbol{\alpha}_{1-7}$
10	500	64.53	65.71	35.45	81.05	79.89	40.14
	1,000	64.46	67.22	36.40	81.83	80.93	41.60
	2,000	63.75	66.81	36.22	81.78	81.04	41.30
20	500	62.96	63.81	39.42	84.22	84.61	51.46
	1,000	62.36	62.45	39.18	84.78	84.71	51.67
	2,000	61.78	61.85	39.27	85.17	85.12	51.75

Note. MS-DINA = multiple-strategy deterministic, inputs, noisy "and" gate; SS-DINA = single-strategy deterministic, inputs, noisy "and" gate.

Table C7. Attributes for Fractions Items.

Attribute	Description
(1)	Basic fraction subtraction
(2)	Simplify/reduce
(3)	Separate whole number from fraction
(4)	Borrow one from whole number to fraction
(5)	Convert whole number to fraction
(6)	Convert mixed number to fraction
(7)	Column borrow in subtraction

Table C8. The Q-Matrices for the Fraction Subtraction Data.

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

Appendix D

Further Discussions Regarding Attribute Pattern Identification

Chiu, Douglas, and Li (2009) showed that, in the case of the SS-DINA model, having a complete Q-matrix is a necessary and sufficient condition for the attribute patterns to be identifiable. In general, regardless of the cognitive diagnosis models (CDMs), the resolution afforded by a particular Q-matrix will determine the extent to which attribute patterns can be clearly distinguished from each other. A complete Q-matrix implies a one-to-one correspondence between the attribute and ideal response patterns. A practical step to determine how well the attribute patterns can be distinguished from each other is to determine the number of equivalence classes implied by the Q-matrix of the test. For a particular Q-matrix, attribute patterns in the same equivalence class will result in the same ideal response pattern, therefore, cannot be distinguished from each other. When a complete Q-matrix is involved, the number of equivalence classes is 2^K, each containing a single attribute pattern; with incomplete Q-matrices, fewer equivalence classes are created, resulting in some classes with multiple, indistinguishable attribute patterns. Depending on the desired inference, this may or may not be an issue. For example, assume a test that measures two attributes, where α_1 is assumed to be more basic compared with α_2. A Q-matrix with the specifications 10 and 11 will create only three equivalence classes, one of which will contain both the attribute patterns 00 and 01. If researchers are not particularly interested in identifying individuals who have not mastered α_1 but have mastered α_2 (i.e., 01), such a Q-matrix would suffice. However, if the attribute patterns 00 and 01 need to be further distinguished from each other, items that measure α_2 but α_1 not need to be added. For obvious reasons, this is not a plausible remedy when CDMs are retrofitted to extant data. Moreover, depending on the nature of the attributes, creating items that require an advanced attribute, but not a more basic attribute, at least implicitly, may not be theoretically feasible.

As recommended by de la Torre and Douglas (2008), one can examine the equivalence classes to determine the extent to which attribute patterns are identifiable when the MS-DINA model is involved. If the number of equivalence classes is relatively small, they suggested simplifying the model, which include using a single strategy. Another strategy that can be recommended is to include additional items that can further split some of the attributes in the same equivalence classes. However, due to the more complex nature of the MS-DINA model, in addition to the attribute relationships, the relationship between the strategies needs to also be examined to determine whether additional items can be created to allow for greater identifiability of the attribute patterns. By presenting two versions of an alternative strategy, whose Q-matrix rows are permutations of each other, de la Torre and Douglas also highlighted the importance of careful test construction in ensuring that a maximum number of equivalence classes is created, particularly in situations where the theoretical maximum (i.e., 2^K) cannot be achieved. However, the authors would like to underscore that optimizing the number of equivalence classes by means of a more deliberate test construction process is a challenge not unique to the MS-DINA model—this is an issue that all CDMs requiring a Q-matrix need to deal with, albeit to a lesser extent.

The numbers of equivalence classes in the fraction subtraction data computed using the Q-matrices for the SS-DINA and MS-DINA models are given in Table D1. The model shows that only nine equivalence classes can be identified using the single-strategy Q-matrix. With the introduction of an additional strategy, this number was increased to 13, indicating that students can potentially be further distinguished from each other using the multiple-strategy model. If the test can be extended by adding two new items, namely, ''Express 9/12 in simplest terms,'' and ''Write 2 in fractional form,'' which measures α_2 only and α_5 only, respectively, the number of equivalence classes can be increased dramatically. As Table D1 shows, the number of equivalence classes will more than double for either the SS-DINA or the MS-DINA model with the addition of these two particular items.

Table D1. Number of Equivalence Classes in the Fraction Subtraction Data.

CDM	Q-Matrix	
	Original	Extended
SS-DINA	9	19
MS-DINA	13	27

Note. CDM = cognitive diagnosis model; SS-DINA = single-strategy deterministic, inputs, noisy "and" gate; MS-DINA = multiple-strategy deterministic, inputs, noisy "and" gate.

Authors' Note

The opinions expressed herein are those of the authors, and do not necessarily represent the views of the National Science Foundation or the National Institute on Alcohol Abuse and Alcoholism.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the National Science Foundation Grant DRL-0744486 and the National Institute on Alcohol Abuse and Alcoholism Grant R01AA019511.

Notes

1. The formulas are $\mathrm{MAD}_g = \frac{\sum_j^J |\hat{g}_j - g_j|}{J}$ and $\mathrm{MAD}_{1-s} = \frac{\sum_j^J |\hat{s}_j - s_j|}{J}$.
2. The formulas in computing the Akaike information criterion (AIC) and Bayesian information criterion (BIC) model fit indices for the single-strategy deterministic, inputs, noisy ''and'' gate (SS-DINA) and multiple-strategy deterministic, inputs, noisy ''and'' gate (MS-DINA) models are expressed as $\mathrm{AIC} = -2\ln(L) + 2 \times (2 \times J + 2^K - 1)$ and $\mathrm{BIC} = -2\ln(L) + (2 \times J + 2^K - 1) \times \ln(n)$, respectively. In addition to $-2\ln(L)$, the AIC and the BIC contain additional terms that serve as penalty for model complexity.
3. In the SS-DINA model, the $\mathrm{MAD}_g = 0.02$ and $\mathrm{MAD}_{1-s} = 0.01$. The percent of correct attribute classification of $\alpha_{1-5} = 80.97\%$. In the MS-DINA model, the $\mathrm{MAD}_g = 0.02$ and $\mathrm{MAD}_{1-s} = 0.01$. The percent of correct attribute classification of $\alpha_{1-5} = 82.84\%$, $\alpha_{1-7} = 72.95\%$.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716-723.

Barnes, T. (2010). Novel derivation and application of skill matrices: The q-matrix method. In C. Ramero, S. Vemtora, M. Pechemizkiy, & R. S. J. de Baker (Eds.), *Handbook on educational data mining* (pp. 159-172). Boca Raton, FL: Chapman & Hall.

Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis* (2nd ed.). New York, NY: Chapman & Hall.

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

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, *35*, 8-26.

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, *45*, 343-362.

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

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

de la Torre, J., & Chiu, C.-Y. (2010, April). *General empirical method of Q-matrix validation*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Denver, CO.

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.

de la Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement*, *47*, 227-249.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, *39*, 1-38.

Doornik, J. A. (2009). Object-Oriented Matrix Programming using Ox (Version 6) [Computer software]. London, England: Timberlake Consultants Press.

Fuson, K. C., Wearne, D., Hiebert, J., Human, P., Murray, H., Olivier, A., Carpenter, T. P., & Fennema, E. (1997). Children's conceptual structures for multidigit numbers and methods of multidigit addition and subtraction. *Journal for Research in Mathematics Education*, *28*, 130-162.

Gamerman, D. (1997). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. London, England: Chapman & Hall.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, R. B. (2003). *Bayesian data analysis* (2nd ed.). London, England: Chapman & Hall.

Haertel, E. (1984). An application of latent class models to assessment data. *Applied Psychological Measurement*, *8*, 333-346.

Huebner, A., Wang, B., & Lee, S. (2009). Practical issues concerning the application of the DINA model to CAT data. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing*. Retrieved from http://publicdocs.iacat.org/cat2010/cat09huebner.pdf

Huo, Y., & de la Torre, J. (2013). *Data-driven Q-matrix specification for subsequent test forms*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

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.

Kendall, M. G., & Stuart, A. (1973). *The advanced theory of statistic* (Vol. 2). New York, NY: Hafner.

Lee, J., Grigg, W., & Dion, G. (2007). *The Nation's Report Card: Mathematics 2007* (NCES 2007-494). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/itemmaps/?subj=Mathematics&year=2007&grade=4

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

Liu, J., Xu, G., & Ying, Z. (2013). Theory of the self-learning Q-matrix. *Bernoull*, *19*, 1790-1817.

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

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464.

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

Tatsuoka, K. K. (1987). Validation of cognitive sensitivity for item response curves. *Journal of Educational Measurement*, *24*, 233-245.

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