DINA Models for Multiple-Choice Items With Few Parameters: Considering Incorrect Answers

Applied Psychological Measurement
2015, Vol. 39(6) 431–447
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146621615574693
apm.sagepub.com

Koken Ozaki¹

Abstract

The deterministic-input, noisy “and” gate (DINA) model can judge whether an individual examinee has mastered each skill that is needed to answer an item correctly. This information is useful for students to know their deficits and for teachers to teach effectively. The DINA model is a statistical model for binary (correct or incorrect) data. However, recently a DINA model for multiple-choice items was developed by de la Torre. The model is aimed at obtaining information about students’ skills from incorrect answers. In the present study, new DINA models for multiple-choice items are developed that need many fewer parameters while still being able to express various answering probabilities without any restrictions on the form of the Q-matrix. Simulations using a Markov chain Monte Carlo method are performed to demonstrate the efficacies of the proposed models compared with the DINA model for binary data and the model of de la Torre for multiple-choice items, if appropriate starting values are set.

Keywords

diagnostic testing, MCMC, multiple-choice item

Introduction

Cognitive diagnosis models (CDMs) are models for estimating examinees’ mastery or nonmastery of skills (referred to as “attributes”) that are needed to answer items correctly. Attributes are latent, and so CDMs are latent variable models. Many CDMs have been developed, including the deterministic-input, noisy “and” gate (DINA; Junker & Sijtsma, 2001) model, the higher order DINA model (de la Torre & Douglas, 2004), the reduced reparameterized unified model (R-RUM; Hartz, 2002), the deterministic inputs, noisy “or” gate model (DINO; Templin & Henson, 2006), and the noisy inputs, deterministic “and” gate model (NIDA; Maris, 1999).

All of these CDMs require a Q-matrix (Embretson, 1984; Tatsuoaka, 1983) and response data to estimate examinees’ cognitive profiles. Table 1 shows the Q-matrix for three items from well-known fraction subtraction data (DeCarlo, 2011; de la Torre & Douglas, 2004; Tatsuoaka, 1990). In the Q-matrix, a 1 means that the attribute is needed to correctly answer the item and a

¹University of Tsukuba, Tokyo, Japan

Corresponding Author:

Koken Ozaki, Associate Professor, Graduate School of Business Sciences, University of Tsukuba, 3-29-1 Otsuka, Bunkyo-ku, Tokyo 112-0012, Japan.
Email: koken@gssm.otsuka.tsukuba.ac.jp

Table 1. Q-Matrix for Fraction Subtraction Data, Three Items, Three Skills.

Attribute	α_1	α_2	α_3
Item 1 ($\frac{5}{3} - \frac{3}{4}$)	1	1	1
Item 2 ($\frac{3}{4} - \frac{3}{8}$)	1	0	1
Item 3 ($\frac{6}{7} - \frac{4}{7}$)	0	0	1

Note. α_1 = find a common denominator; α_2 = column borrow to subtract the second numerator from the first; α_3 = subtract numerators.

0 means that it is not needed. For example, the first item requires all three attributes (α_1 , α_2 , and α_3) to correctly answer; however, the second and third items require only the first and third attributes (α_1 and α_3) and the third attribute (α_3), respectively. Because all three items require α_3 , it might be naively thought that the simple sum of the 3 item scores would indicate whether an examinee has α_3 (if the score is 3) or not (otherwise). However, this is actually incorrect. Because Items 1 and 2 require other attributes, an incorrect answer to Items 1 and 2 does not indicate that the examinee does not have α_3 (Tatsuoka, 2009). A correct answer to the third item indicates that the examinee may have α_3 . However, it is not appropriate to have strong confidence, because the examinee might correctly answer the item only by chance. However, an incorrect answer to the third item does not necessarily indicate that the examinee does not have α_3 , because the examinee might have only carelessly made a mistake on the item. Therefore, probabilistic models are needed to estimate examinees' cognitive profiles from the Q-matrix and response data.

DINA Model

The models presented in this article are based on the DINA model, which is the most popular model among CDMs. In all the discussed models, let j be an item, i be an examinee, and α_i be the knowledge state of examinee i . The k th element of α_i is expressed as α_{ik} , which is 1 when examinee i has mastered attribute $k = (1, 2, \dots, K)$ and 0 otherwise. Furthermore, let s_j be the slip parameter of item j , which is the probability that an examinee who has all of the knowledge needed to correctly answer the item fails to answer it correctly. Finally, let g_j be the guessing parameter for item j , which is the probability that an examinee who does not have all of the knowledge needed to correctly answer the item nevertheless does correctly answer it. The probability that examinee i correctly answers item j ($Y_{ij} = 1$) is expressed in the DINA model as follows:

$$P(Y_{ij} = 1 | \alpha_i) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}}, \tag{1}$$

where

$$\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}. \tag{2}$$

Here, q_{jk} is the element of the Q-matrix for item j . If attribute k is needed to answer item j , then $q_{jk} = 1$, and otherwise $q_{jk} = 0$.

Table 2. Selection Probabilities for an Item With Five Options.

	α_1	1	1	1	0	1	0	0	0
	α_2	1	1	0	1	0	1	0	0
	α_3	1	0	1	1	0	0	1	0
Option 1	(1,1,1)	$p(1 \mathbf{A}_1)$	$p(1 \mathbf{A}_2)$	$p(1 \mathbf{A}_3)$	$p(1 \mathbf{A}_4)$	$p(1 \mathbf{A}_5)$	$p(1 \mathbf{A}_6)$	$p(1 \mathbf{A}_7)$	$p(1 \mathbf{A}_8)$
Option 2	(1,1,0)	$p(2 \mathbf{A}_1)$	$p(2 \mathbf{A}_2)$	$p(2 \mathbf{A}_3)$	$p(2 \mathbf{A}_4)$	$p(2 \mathbf{A}_5)$	$p(2 \mathbf{A}_6)$	$p(2 \mathbf{A}_7)$	$p(2 \mathbf{A}_8)$
Option 3	(1,0,0)	$p(3 \mathbf{A}_1)$	$p(3 \mathbf{A}_2)$	$p(3 \mathbf{A}_3)$	$p(3 \mathbf{A}_4)$	$p(3 \mathbf{A}_5)$	$p(3 \mathbf{A}_6)$	$p(3 \mathbf{A}_7)$	$p(3 \mathbf{A}_8)$
Option 4	(0,0,1)	$p(4 \mathbf{A}_1)$	$p(4 \mathbf{A}_2)$	$p(4 \mathbf{A}_3)$	$p(4 \mathbf{A}_4)$	$p(4 \mathbf{A}_5)$	$p(4 \mathbf{A}_6)$	$p(4 \mathbf{A}_7)$	$p(4 \mathbf{A}_8)$
Option 5	(0,0,0)	$p(5 \mathbf{A}_1)$	$p(5 \mathbf{A}_2)$	$p(5 \mathbf{A}_3)$	$p(5 \mathbf{A}_4)$	$p(5 \mathbf{A}_5)$	$p(5 \mathbf{A}_6)$	$p(5 \mathbf{A}_7)$	$p(5 \mathbf{A}_8)$

Note. α_k is attribute k , and $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_8$ are knowledge states, for example, $\mathbf{A}_1 = (\alpha_1, \alpha_2, \alpha_3) = (1, 1, 1)$.

Multiple-Choice Item

As described above, the DINA model is for binary (correct or incorrect) data; however, we are proposing DINA models for multiple-choice items that are in some sense simplified forms of the Multiple-choice DINA (MC-DINA) model (de la Torre, 2009a) explained below. In the example below, let us consider an item with five options that requires three attributes to answer correctly. However, the notation of the equations accommodates the general situation for the number of attributes. Table 2 shows the probabilities that an examinee with each possible knowledge state selects each option. Option 1, the correct option, requires three attributes to select. Option 2 requires the first and the second attributes to select, and the remaining options can be similarly explained. Because three attributes are required to correctly answer the item, there are $2^3 = 8$ patterns of the knowledge state, which are expressed as $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_8$. Therefore, $p(1|\mathbf{A}_1)$ is the probability that an examinee with $\mathbf{A}_1$ selects the first option and $p(1|\mathbf{A}_2)$ is the probability that an examinee with $\mathbf{A}_2$ selects the first option. Note that the scalar α_1 indicates the first attribute and the vector $\mathbf{A}_1 = (\alpha_1, \alpha_2, \alpha_3) = (1, 1, 1)$.

Multiple-choice items as shown in Table 2 can be analyzed using the DINA model, by coding 1 when Option 1 is selected and 0 otherwise (e.g., see Birenbaum, Tatsuoaka, & Xin, 2005; Tatsuoaka, Corter, & Tatsuoaka, 2004). When the DINA model is fit to the multiple-choice item shown in Table 2, $p(1|\mathbf{A}_1) = 1 - s$, $\sum_{c \neq 1} p(c|\mathbf{A}_1) = s$, $p(1|\mathbf{A}_l) = g$ for $l (= 2, \dots, 8)$, and $\sum_{c \neq 1} p(c|\mathbf{A}_l) = 1 - g$ for $l (= 2, \dots, 8)$. Therefore, for example, for an examinee with $\mathbf{A}_2$, the sum of the selection probabilities for Options 2, 3, 4, and 5 is $(1 - g)$, and there are no differences in selection probabilities between Options 2, 3, 4, and 5. However, in Table 2, $p(2|\mathbf{A}_2)$ would be greater than $p(5|\mathbf{A}_2)$ because this examinee has a sufficient mastery level not to select Option 5. And when the DINA model is fit to the multiple-choice item shown in Table 2, $p(1|\mathbf{A}_2) = p(1|\mathbf{A}_8) = g$. However, $p(1|\mathbf{A}_2)$ would be greater than $p(1|\mathbf{A}_8)$ in Table 2 because an examinee with $\mathbf{A}_2$ has mastered more attributes than an examinee with $\mathbf{A}_8$.

Therefore, when a multiple-choice item is coded as binary data and the DINA model is fit, obtaining sufficiently accurate information on examinees' knowledge states is difficult (de la Torre, 2009a). This is why CDMs for multiple-choice items are necessary. The developed models can capture examinees' knowledge states better than the DINA model by obtaining information from incorrect answers (options). It is expected that a smaller number of items than when using the DINA model will be sufficient to correctly estimate examinees' knowledge states.

de la Torre's Model

Most of the existing CDMs are for binary (correct or incorrect) data. One of the exceptions is that of de la Torre (2009a), which allows multiple-choice items to be analyzed when incorrect options are coded by attributes, which means that some (but not all) of the required attributes

are assumed to be required for each different option. For example, let us consider the item $2\frac{4}{12} - \frac{7}{12}$ with three options: $2\frac{3}{12}$, $1\frac{9}{12}$, and $1\frac{3}{4}$. This item requires three attributes, which are “borrow one from whole number to fraction” (α_1), “basic fraction subtraction” (α_2), and “reduce/simplify” (α_3); this item is a slightly modified version of the example of Figure 1 of de la Torre (2009a). If an examinee selects the third option, he or she has mastered α_1 , α_2 , and α_3 . And although both $2\frac{3}{12}$ and $1\frac{9}{12}$ are incorrect options, they have different meanings. $2\frac{3}{12}$ is the option that examinees who have only α_2 are most likely to select, and $1\frac{9}{12}$ is the option that examinees who have mastered only α_1 and α_2 are most likely to select. In this case, both of the incorrect options are coded by attributes.

de la Torre’s (2009a) general model estimates all of the probabilities, $p(1|\mathbf{A}_1)$, $p(1|\mathbf{A}_2)$, $\dots$, $p(5|\mathbf{A}_7)$, and $p(5|\mathbf{A}_8)$, shown in Table 2; however, note that some of the attribute vectors belonging to the same latent groups have common probability estimates. As is stated in his article, the total number of item parameters in the model is $\sum_{j=1}^J C(C_j^* + 1) - \sum_{j=1}^J C_j^* - J$. Here, C is the number of options for each item ($C_j = C$ for all j is considered) and C_j^* is the number of options coded with different attribute patterns. Therefore, too many parameters have to be estimated when fitting the model, although it should be noted that de la Torre (2009a) mentioned that simplifying assumptions will reduce the number of parameters. In the present study, three probability models are developed that require many fewer parameters than that of de la Torre’s (2009a) general model while still expressing sufficiently complex selection probability patterns. Moreover, their parameters can be estimated much more easily and the models are much more parsimonious.

Another feature of de la Torre’s (2009a) model is that it uses items that can distinguish examinees into one of the $C_j^* + 1$ groups by means of the coded options. The term *group* here has the same meaning as a class of knowledge state. Therefore, for example, if two of the options of an item require separately only α_1 or α_2 and the item does not have an option that requires both α_1 and α_2 , then this model cannot be used, because examinees who have both α_1 and α_2 can be classified into both groups. de la Torre (2009a) used marginal maximum likelihood estimation with the estimation–maximization (EM) algorithm, and in each cycle, $p_j(c|g)$, which is a selection probability of latent group g for option c of item j , is updated. Therefore, because $p_j(c|g)$, not each $p_j(c|\alpha_k)$, is the parameter in this model, such options that cannot distinguish more than two latent groups cannot be used. The developed models avoid this limitation. In Simulation Study 1 of this article, the Markov chain Monte Carlo (MCMC; Patz & Junker, 1999) method is used in the estimation of de la Torre’s (2009a) model. However, neither MCMC nor the EM algorithm can estimate the parameters of such items that cannot distinguish more than two latent groups, because of $p_j(c|g)$ being the parameter in this model.

The purpose of the present study is (a) to simplify de la Torre’s model so that the resulting simplified models require significantly fewer parameters, (b) can be used to analyze multiple-choice items even when each of them cannot distinguish more than two latent groups, and also (c) to develop a more complex model, structured DINA model for multiple-choice items 3 (MC-S-DINA3), which will turn out not to be easily expressed as a simplification of de la Torre’s model. These three points are the differences with de la Torre’s (2009a) model.

Multiple-Choice Models

Three models will be developed, referred to as the structured DINA models for multiple-choice items: Structured DINA model for multiple-choice items 1 (MC-S-DINA1), Structured DINA model for multiple-choice items 2 (MC-S-DINA2), and MC-S-DINA3 (in order of development and complexity). de la Torre’s (2009a) model is referred to as the MC-DINA model. The difference between the proposed models and the MC-DINA model is whether the selection

probability is structured (proposed models) or not (MC-DINA model). The proposed models are structured because the selection probabilities are organized using parameters. However, in de la Torre's (2009a) model, selection probabilities themselves are parameters.

MC-S-DINA1

MC-S-DINA1 is the model on which the other two proposed models are based. In MC-S-DINA1, the probability of selecting option c of item j by examinee i is expressed as

$$P(Y_{ij} = c | \alpha_i) = \gamma_{ij} (1 - \delta_j)^{\eta_{ijc}} \zeta_j^{1 - \eta_{ijc}} + \frac{(1 - \gamma_{ij})}{C}. \quad (3)$$

Here,

$$\eta_{ijc} = \prod_{k=1}^{K_j} (2 - 2^{(\alpha_{ik} - q_{jkc})^2}), \quad (4)$$

$$\gamma_{ij} = \sum_{c=1}^C \eta_{ijc} \left(1 - \prod_{k=1}^{K_j} (1 - \alpha_{ik}) \right), \quad (5)$$

where K_j is the number of required attributes for item j and the attributes are ordered such that these attributes are the first K_j attributes (de la Torre, 2011). In Equation 4, $q_{jkc} = 1$ indicates that attribute k is needed to select option c of item j . If attribute k is not needed, $q_{jkc} = 0$. $2 - 2^{(\alpha_{ik} - q_{jkc})^2} = 1$ when $\alpha_{ik} = q_{jkc}$, namely, when $(\alpha_{ik}, q_{jkc}) = (1, 1)$ or $(\alpha_{ik}, q_{jkc}) = (0, 0)$. And $2 - 2^{(\alpha_{ik} - q_{jkc})^2} = 0$ when $(\alpha_{ik}, q_{jkc}) = (1, 0)$ or $(\alpha_{ik}, q_{jkc}) = (0, 1)$. Therefore, $\eta_{ijc} = 1$ when examinee i has exactly the attributes needed for option c of item j , and $\eta_{ijc} = 0$ otherwise. In other words, $\eta_{ijc} = 1$ when examinee i 's knowledge state perfectly matches the attribute vector of option c of item j .

In Equation 5, $\gamma_{ij} = 1$ when $1 - \prod_{k=1}^{K_j} (1 - \alpha_{ik}) = 1$ (examinee i has at least one attribute needed to correctly answer item j) and $\sum_{c=1}^C \eta_{ijc} = 1$, and otherwise $\gamma_{ij} = 0$. In MC-S-DINA1, when $\gamma_{ij} = 1$, the first term determines the selection probability, and when $\gamma_{ij} = 0$, the second term determines the selection probability. When $\gamma_{ij} = 0$, $(\alpha_1, \alpha_2, \alpha_3)$ are $(1, 0, 1)$, $(0, 1, 1)$, $(0, 1, 0)$, or $(0, 0, 0)$ in the case of Table 2), $(1 - \gamma_{ij})/C = 1/C$, which is the relatively strong assumption noted in de la Torre's (2009a) model. Table 3 shows η_{ijc} and γ_{ij} for item j with five options.

Conditional on α_i , examinee i has relatively high selection probabilities for some options; nevertheless, he or she may select one of the other options. The probability of the latter event is

Table 3. η and γ Values for Five Options With Three Attributes.

	α_1	1	1	1	0	1	0	0	0
	α_2	1	1	0	1	0	1	0	0
	α_3	1	0	1	1	0	0	1	0
Option 1	(1, 1, 1)	1	0	0	0	0	0	0	0
Option 2	(1, 1, 0)	0	1	0	0	0	0	0	0
Option 3	(1, 0, 0)	0	0	0	0	1	0	0	0
Option 4	(0, 0, 1)	0	0	0	0	0	0	1	0
Option 5	(0, 0, 0)	0	0	0	0	0	0	0	1
γ		1	1	0	0	1	0	1	0

δ_j , namely, δ_j is the probability that examinee i with α_i selects an option other than the most likely one, given his or her knowledge state. Therefore, $1 - \delta_j$ is the probability that an examinee whose α_i matches a required attribute pattern for an option of item j actually selects that option. ζ_j is the probability that an examinee i whose α_i does not match a required attribute pattern for an option of item j selects that option, and therefore, there is a relationship between δ_j and ζ_j . Specifically, when an examinee i with α_i who is expected to select option c does not select c (with probability δ_j), he or she selects any one other option with probability ζ_j , and so we can set $\zeta_j = \delta_j / (C - 1)$ or equivalently $\delta_j = \zeta_j(C - 1)$. This is the mild assumption described by de la Torre (2009a).

Therefore, MC-S-DINA1 can be reexpressed as

$$P(Y_{ij}=c|\alpha_i) = \gamma_{ij}(1 - (C - 1)\zeta_j)^{\eta_{ijc}} \zeta_j^{1-\eta_{ijc}} + \frac{(1 - \gamma_{ij})}{C}, \tag{6}$$

or equivalently as

$$P(Y_{ij}=c|\alpha_i) = \gamma_{ij}(1 - \delta_j)^{\eta_{ijc}} \frac{\delta_j}{C - 1}^{1-\eta_{ijc}} + \frac{(1 - \gamma_{ij})}{C}. \tag{7}$$

Thus, the number of item parameters is 1 (δ_j or ζ_j) for each item.

The selection probabilities of MC-S-DINA1 are shown in Table 4 for $\delta_j = 0.2$. As the table shows, an examinee who has all of the attributes selects Options 1, 2, 3, 4, and 5 with probabilities .8, .05, .05, .05, and .05, respectively.

MC-S-DINA2

In Table 4, the probability that an examinee whose α_i matches a required attribute pattern for an option of item j selects that option is $1 - \delta_j$ regardless of α_i (except for the case $\alpha_i = 0$). However, the probability of selecting Option 1 by an examinee with $\alpha = (1,1,1)$ would be higher than that of selecting Option 2 by an examinee with $\alpha = (1,1,0)$. The reason is revealed by the following example.

Let us reconsider the item $2\frac{4}{12} - \frac{7}{12}$ with the three options $2\frac{3}{12}$, $1\frac{9}{12}$, and $1\frac{3}{4}$, which requires the three attributes “borrow one from whole number to fraction” (α_1), “basic fraction subtraction” (α_2), and “reduce/simplify” (α_3). An examinee who has mastered these three attributes would select $1\frac{3}{4}$ with high probability because it is the only option such an examinee might logically choose. However, an examinee who has only α_1 and α_2 is most likely to select $1\frac{9}{12}$,

Table 4. Selection Probabilities for MC-S-DINA1.

	α_1	1	1	1	0	1	0	0	0
	α_2	1	1	0	1	0	1	0	0
	α_3	1	0	1	1	0	0	1	0
Option 1	(1,1,1)	.80	.05	.20	.20	.05	.20	.05	.20
Option 2	(1,1,0)	.05	.80	.20	.20	.05	.20	.05	.20
Option 3	(1,0,0)	.05	.05	.20	.20	.80	.20	.05	.20
Option 4	(0,0,1)	.05	.05	.20	.20	.05	.20	.80	.20
Option 5	(0,0,0)	.05	.05	.20	.20	.05	.20	.05	.20

Note. MC-S-DINA1= Structured DINA model for multiple-choice items 1.

although still not with high probability. The reason for this is that, because such an examinee does not have α_3 , he or she may mistakenly simplify the fraction and then select one of the other options. In other words, $1\frac{9}{12}$ is not the only option such an examinee might choose. MC-S-DINA2 addresses this point. In this model, δ_j takes different parameters according to the attribute vector of option c . Because δ_j differs by option, it must be expressed as δ_{jc} , which is the parameter for option c of item j .

Now β_{ij} is defined as follows:

$$\beta_{ij} = \sum_{c=1}^C \delta_{jc} \eta_{ijc}. \quad (8)$$

β_{ij} is the probability that examinee i with α_i selects an option other than the option whose attribute pattern matches α_i . Note that although $C = 5$, for the item shown in Tables 2, 3, and 4, it is not necessary to estimate δ_{j5} , because when $\alpha = 0$, δ_{j5} is not required for the calculation of the selection probability, as can be seen in Equation 9. β_{ij} is δ_{jc} for the option c which examinee i is most likely to select.

MC-S-DINA2 is expressed as follows:

$$P(Y_{ijc} = 1 | \alpha_i) = \gamma_{ij} (1 - \delta_{jc})^{\eta_{ijc}} \left(\frac{\beta_{ij}}{C - 1} \right)^{1 - \eta_{ijc}} + \frac{(1 - \gamma_{ij})}{C}. \quad (9)$$

Compared with MC-S-DINA1, only selection probabilities when $\gamma_{ij} = 1$ are different. The number of parameters is less than the number of options C . The selection probabilities of MC-S-DINA2 are shown in Table 5 for $\delta_1 = 0.2$, $\delta_2 = 0.3$, and $\delta_3 = \delta_4 = 0.4$. Because two options (Options 3 and 4) require one attribute, $\delta_3 = \delta_4$ may not be a strict constraint. Note that the δ_{jc} parameter associated with the key need not be the smallest. A distractor that is carefully constructed may in fact have a higher chance of attracting examinees with a specific attribute pattern.

MC-S-DINA3

In both MC-S-DINA1 and MC-S-DINA2, the selection probabilities for options other than the option which examinee i is most likely to select are the same for each α_i . In MC-S-DINA3, this restriction is relaxed by setting the selection probabilities according to the closeness between α_i and the attribute vector for option c of item j . Closeness is measured by inner product.

Let r_{ijc} be the inner product of α_i and the attribute vector for option c of item j ; therefore, $r_{ijc} = \alpha_i \mathbf{q}'_{jc}$, where the elements of $\mathbf{q}_{jc}$ are q_{jkc} , and α_i and $\mathbf{q}_{jc}$ are row vectors. If r_{ijc} is large,

Table 5. Selection Probabilities for MC-S-DINA2.

	α_1	1	1	1	0	1	0	0	0
	α_2	1	1	0	1	0	1	0	0
	α_3	1	0	1	1	0	0	1	0
Option 1	(1,1,1)	.800	.075	.200	.200	.100	.200	.100	.200
Option 2	(1,1,0)	.050	.700	.200	.200	.100	.200	.100	.200
Option 3	(1,0,0)	.050	.075	.200	.200	.600	.200	.100	.200
Option 4	(0,0,1)	.050	.075	.200	.200	.100	.200	.600	.200
Option 5	(0,0,0)	.050	.075	.200	.200	.100	.200	.100	.200

Note. MC-S-DINA2= Structured DINA model for multiple-choice items 2.

examinee i is expected to select option c of item j with high probability. However, if r_{ijc} is small, examinee i is not expected to select option c of item j .

ω_{ijc} is defined as follows:

$$\omega_{ijc} = \begin{cases} \beta_{ij} \frac{(1-\eta_{ijc})(1+r_{ijc})}{\sum_{c=1}^C (1-\eta_{ijc})(1+r_{ijc})} & \text{if } \gamma_{ij} = 1 \\ \frac{(1-\eta_{ijc})(1+r_{ijc})}{\sum_{c=1}^C (1-\eta_{ijc})(1+r_{ijc})} & \text{if } \gamma_{ij} = 0. \end{cases} \quad (10)$$

For the $\gamma_{ij} = 1$ case, Equation 10 shows that β_{ij} is distributed to ω_{ijc} according to the closeness between α_i and the attribute vector for option c of item j , where the probability is expressed as ω_{ijc} . If $\gamma_{ij} = 0$, Equation 10 shows that the number 1 is distributed to ω_{ijc} according to the closeness.

MC-S-DINA3 is expressed as follows:

$$P(Y_{ijc} = 1 | \alpha_i) = \left(1 - \prod_{k=1}^{K_j} (1 - \alpha_{ik}) \right) (1 - \delta_{jc})^{\eta_{ijc}} (\omega_{ijc})^{1-\eta_{ijc}} + \frac{\prod_{k=1}^{K_j} (1 - \alpha_{ik})}{C}. \quad (11)$$

γ_{ij} does not appear explicitly in Equation 11, because the second term should only be calculated for examinee i if $\alpha_i = 0$. The selection probabilities of MC-S-DINA3 are shown in Table 6 for $\delta_1 = 0.2$, $\delta_2 = 0.3$, and $\delta_3 = \delta_4 = 0.4$. In this case, the selection probabilities for options other than the option which an examinee is most likely to select are not all the same. For example, if $\alpha_i = (1,1,1)$ for examinee i , then the selection probability of Option 2 (.075) is one and a half times those of Options 3 (.05) and 4 (.05) because 1 plus the inner product of $\alpha_i = (1,1,1)$ and the attribute vector of Option 2 is 3, which is one and a half times as large as 1 plus the inner products of $\alpha_i = (1,1,1)$ and the attribute vectors of Options 3 and 4 (= 2). The number of parameters is less than the number of options C , which is the same as for MC-S-DINA2. Specifically, there are four parameters: δ_{j1} , δ_{j2} , δ_{j3} , and δ_{j4} in this example.

Note that MC-S-DINA1 and MC-S-DINA2 are simplified forms of de la Torre's (2009a) model, with the difference that MC-S-DINA1 and MC-S-DINA2 can analyze multiple-choice items even when each of them cannot distinguish more than two latent groups. In contrast, MC-S-DINA3 is difficult to express as a simplification of de la Torre's (2009a) model. In practical application of the models, some goodness-of-fit indices are necessary to determine the best-fit model. Appendix A of an online supplement shows a method of model comparison.

Table 6. Selection Probabilities for MC-S-DINA3.

	α_1	1	1	1	0	1	0	0	0
	α_2	1	1	0	1	0	1	0	0
	α_3	1	0	1	1	0	0	1	0
Option 1	(1,1,1)	.800	.129	.300	.333	.133	.286	.160	.200
Option 2	(1,1,0)	.075	.700	.200	.222	.133	.286	.080	.200
Option 3	(1,0,0)	.050	.086	.200	.111	.600	.143	.080	.200
Option 4	(0,0,1)	.050	.043	.200	.222	.067	.143	.600	.200
Option 5	(0,0,0)	.025	.043	.100	.111	.067	.143	.080	.200

Note. MC-S-DINA3= Structured DINA model for multiple-choice items 3.

Parameter Estimation

As in item response theory, the joint maximum likelihood estimation method, the marginal maximum likelihood estimation method, and the MCMC method can also be adopted to estimate parameters. However, as is well known for item response theory, using the joint marginal maximum likelihood estimation method leads to inconsistent estimates of item parameters (Baker, 1992; de la Torre, 2009b). Therefore, the marginal maximum likelihood estimation method using the EM algorithm (de la Torre, 2009a, 2009b, 2011) or the MCMC method (de la Torre, 2009b; de la Torre & Douglas, 2004; Junker & Sijtsma, 2001) should be used. In estimating the parameters of the developed models, MCMC with the Metropolis–Hastings (M-H) within Gibbs algorithm is adopted, because the full conditional distributions cannot be sampled directly. Furthermore, although the EM algorithm is the original estimation method for the MC-DINA model, in Simulation Study 1 of this article, MCMC is adopted so that the differences of the estimation methods between models do not contaminate the results. Therefore, for all models shown in this article, MCMC is used. However, this is one of the limitations of the study, and different results may be obtained when a different algorithm is involved. Appendix B of an online supplement shows MCMC algorithms.

Initial Values for α_i

The initial values of α_i are important in the MCMC algorithm because if the initial α_i is close to the true α_i , the chain quickly converges to a stationary distribution; however, if bad initial values are used, much time is required for convergence. In a test where items are in multiple-choice form, Q-vectors, such as those used in the simulation studies, shown in Table 7 for 30 items and four options (de la Torre, 2009a), can be set. The numbers in Table 7 indicate the number of times each attribute is required for the set of options. For example, the Q-vector for Item 21 is (3,2,1,0,0) because the correct option requires α_1 , α_2 , and α_3 , one of the incorrect coded option requires α_1 and α_2 , another incorrect coded option requires only α_1 , and the last incorrect option is not coded. The sum of the Q-vectors for the correct options ((1,1,1,0,0) for Item 21) is also a vector whose elements are the number of the items each attribute is required for. For the case shown in Table 7, this sum is (12,12,12,12,12), which means that each attribute is required 12 times (i.e., by 12 items). However, the sum of the attribute vectors of the options each examinee has selected is also a vector. The ratios of the elements of the latter vector to the elements of the former vector can be considered as rough acquisition probabilities of the examinee for each attribute. For example, for the items shown in Table 7, if the sum of attribute vectors of the options an examinee has selected is (10,7,3,2,9), these ratios are (0.83,0.58,0.25,0.17,0.75). If a ratio is more than 0.5, the initial value for α_{ik} is 1, and otherwise $\alpha_{ik} = 0$. For this examinee, the initial value for α_i is (1,1,0,0,1). This method is used for the setting of the initial values for the three developed models and the MC-DINA model.

For the DINA model, only the correct option is used to calculate the ratio. If this ratio is larger than the geometric mean of the 0.5^{K_j} values, the initial value is set to 1, and otherwise it is set to 0.

Simulation Studies

The same Q-vectors as those used by de la Torre (2009a), shown in Table 7 for 30 items for four options, were used in the simulation study. The purposes of the simulation study were the following: (a) to compare the accuracy of the estimates and the fit to data of the DINA, MC-S-DINA1, MC-S-DINA2, MC-S-DINA3, and MC-DINA models when data are generated from the MC-S-DINA1, MC-S-DINA2, MC-S-DINA3, and MC-DINA models, respectively; (b) to

Table 7. Q-Vectors for Multiple-Choice Items.

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

examine the effect of the number of items on the estimates of examinee and item parameters of the above five models; and (c) to examine whether MC-S-DINA1, MC-S-DINA2, and MC-S-DINA3 can estimate α_{ik} and δ_{jc} for items even when each of them cannot distinguish more than two latent groups. The first point was examined in Simulation Study 1, the second point was examined in Simulation Studies 1 and 2, and the third point was examined in Simulation Study 2. In both of the simulations, the estimation program was written in R (R Development Core Team, 2005).

In Simulation Study 1, data were generated from the MC-S-DINA1, MC-S-DINA2, MC-S-DINA3, and MC-DINA models and analyzed using the DINA, MC-S-DINA1, MC-S-DINA2, MC-S-DINA3, and MC-DINA models. Three conditions for the number of items were examined: 10, 20, and 30. Items 21 to 30 were used for 10 items; Items 11 to 30 were used for 20 items. The number of examinees was 1,000 for each condition. The number of MCMC samples was 5,000 and the burn-in was 2,000 for the three developed models and the MC-DINA model, and 4,000 and 1,000 for the DINA model. These numbers were determined using the convergence criterion of Gelman and Rubin (1992). By running five parallel chains, the criterion was satisfied for all item parameters.

Table 8. Coded Options for Multiple-Choice Items.

Item 21					Item 26				
1	1	1	0	0	1	0	0	1	1
0	1	0	0	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	1
0	0	0	0	0	1	0	0	0	0
Item 22					Item 27				
1	1	0	1	0	0	1	1	1	0
0	1	0	0	0	0	1	0	0	0
0	0	0	1	0	0	0	1	0	0
0	0	0	0	0	0	0	0	1	0
Item 23					Item 28				
1	1	0	0	1	0	1	1	0	1
1	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0
Item 24					Item 29				
1	0	1	1	0	0	1	0	1	1
0	0	0	1	0	0	1	0	0	0
0	0	1	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1
Item 25					Item 30				
1	0	1	0	1	0	0	1	1	1
1	0	0	0	0	0	0	0	0	1
0	0	1	0	0	0	0	0	1	0
0	0	0	0	1	0	0	1	0	0

In Simulation Study 2, the same items were used as in Simulation Study 1 except for Items 21 to 30. Table 8 shows coded options for these items, which cannot be used in the MC-DINA model. In Table 8, the rows represent the options and the columns represent the attributes. In this study, MC-S-DINA1, MC-S-DINA2, and MC-S-DINA3 were used to compare the results. The other conditions were the same as those of Study 1.

In both simulations, 25 repetitions were performed for each condition and the true mastery rates for all attributes were randomly generated from Uniform(0.5,0.9) in each condition. When the true structure was MC-S-DINA1, the true δ_j were generated from Uniform(0,0.2). When the true structures were MC-S-DINA2 and MC-S-DINA3, the true δ_{j1} were generated from Uniform(0,0.2), the true δ_{j2} were generated from Uniform(0.2,0.4), and the true δ_{j3} were generated from Uniform(0.4,0.6). When the true structure was the MC-DINA model, if $\eta_{ijc} = 1$, $p(c|\alpha_i)$ was set to the largest probability and for other options probabilities were set randomly, and if $\eta_{ijc} = 0$ for all options, $p(1|\alpha_i)$, $p(2|\alpha_i)$, $p(3|\alpha_i)$, and $p(4|\alpha_i)$ were set randomly. Chi-square values are calculated for the MC-S-DINA1, MC-S-DINA2, MC-S-DINA3, and MC-DINA models. When the number of coded options was 1, three incorrect options were pooled and two categories (correct or incorrect) were used to calculate the chi-square value (using 2×2 cross-tables). Similarly, when the numbers of coded options were 2 and 3, three categories (using 2×3 cross-tables) and four categories (using 2×4 cross-tables) were used, respectively.

Table 9. Average Recovery Rate of Examinee Parameters for Five Analysis Models (Simulation Study 1).

True Model	DINA	MC1	MC2	MC3	MC
10 items					
MC1	0.686	0.865	0.917	0.881	0.921
MC2	0.676	0.795	0.834	0.798	0.827
MC3	0.655	0.750	0.843	0.836	0.848
MC	0.531	0.537	0.533	0.502	0.574
20 items					
MC1	0.889	0.991	0.991	0.984	0.986
MC2	0.879	0.942	0.953	0.932	0.948
MC3	0.836	0.941	0.951	0.950	0.938
MC	0.588	0.630	0.586	0.521	0.607
30 items					
MC1	0.977	0.996	0.996	0.994	0.994
MC2	0.945	0.983	0.986	0.980	0.985
MC3	0.914	0.975	0.982	0.982	0.979
MC	0.736	0.865	0.917	0.881	0.921

Note. MC1, MC2, and MC3 are the MC-S-DINA1, MC-S-DINA2, and MC-S-DINA3 models, respectively, and MC is de la Torre's (2009a) model.

Results

The results of Simulation Study 1 are shown in Tables 9 through 13. In Table 9, the average correct recovery rate of examinee parameters is shown. For example, when MC-S-DINA1 was the true model and the DINA model was fit to 10 items, the average rate was 0.686. The table illustrates that generally the four multiple-choice models can estimate α_{ik} much better than the DINA model. In particular, when the number of items is small (10 items), the differences are very large. In addition, the average rate for the DINA model becomes worse as the true model becomes more complicated. The other four models also have the same tendencies (except for the case when the true model is MC-S-DINA3) but the average rates never become worse than the case of the DINA model. Therefore, using the information from the incorrect coded options is extremely useful for estimating α_{ik} .

In general, the true model estimates α_i better than the other models. However, even when MC-S-DINA1 or MC-S-DINA3 is the true model, MC-S-DINA2 estimates α_i better than or equally well as either. Therefore, in this simulation setting, MC-S-DINA2 is the better choice than MC-S-DINA1 or MC-S-DINA3. Then, MC-S-DINA2 and MC-DINA should be compared. When the number of items is 10 and the true model is MC-S-DINA1 or MC-S-DINA3, the MC-DINA model estimates α_i better than MC-S-DINA2. Therefore, the MC-DINA model is useful for the case where the number of items is small. The reason why the MC-DINA model estimates α_i better than the other models in the case of a small number of items may be that the MC-DINA model captures small fluctuations in simulation data better than the other models.

Table 10 shows the results of the item parameters. "Bias" is the average of "The number of items $\times$ The number of item parameters for each item $\times$ 25 (repetitions)" biases and "root mean square error" (RMSE) is the average of the 25 root mean square errors. MC(ALL) is the case when calculating using the estimates of all item parameters for each item, and MC(Key) is the case when calculating using the estimates of only the item parameters for the key options. The table illustrates that for all cases, biases are close to 0. And except for the MC-DINA model, RMSEs are small when the number of items is large. Therefore, the developed three models estimate more accurate item parameters than the MC-DINA model. However, when the number

Table 10. Biases and RMSEs of Item Parameters for Four Analysis Models (Simulation Study 1).

Model	Bias	RMSE
10 items		
MC1	−0.012	0.021
MC2	0.019	0.083
MC3	0.016	0.140
MC(ALL)	0.000	0.157
MC(Key)	0.009	0.160
20 items		
MC1	−0.002	0.009
MC2	0.000	0.029
MC3	0.007	0.035
MC(ALL)	0.000	0.172
MC(Key)	0.005	0.191
30 items		
MC1	0.000	0.009
MC2	0.001	0.024
MC3	−0.003	0.026
MC(ALL)	0.000	0.178
MC(Key)	−0.007	0.203

Note. RMSE = root mean square error.

Table 11. Number of Items Each Model Fit Best Using Chi-Square Values (10 Items, Simulation Study 1).

True model	MC1	MC2	MC3	MC
All items				
MC1	4.04	5.56	0.40	0.00
MC2	1.32	7.68	1.00	0.00
MC3	0.00	0.00	9.72	0.28
MC	0.72	3.28	1.08	4.92

of items is 10, MC-S-DINA3 has apparently larger RMSE than in the case where the number of items is 20 or 30. This indicates that 10 items is too few for MC-S-DINA3 to accurately estimate item parameters. Because the key options have the largest selection probabilities, RMSEs for MC(Key) are larger than those for MC(ALL).

Tables 11, 12, and 13 show the results of the model comparisons. The numbers in the tables are the average number of items of the fit best of each model, based on χ^2 (the best-fit model was taken to be the one which gives the smallest chi-square value). These tables show the results separately for the different numbers of coded options as well as for all items. If a model is often selected using χ^2 , this means that the model has sufficient power to describe data.

Generally, the true model is selected. Some exceptions are as follows. MC-S-DINA2 is sometimes selected even when the true model is the MC-DINA model. This tendency is stronger for a smaller number of options. For example, Table 12 shows that for the one-option case, MC-S-DINA2 is selected by 5.84 items on average, whereas MS-DINA is selected by 2.24 items on average. However, Table 12 also shows that for the two-option case, MC-S-DINA2 is selected by 1.84 items on average, whereas MS-DINA is selected by 7.12 items on average. This indicates that when the number of coded options is 1, MC-S-DINA2 is sometimes

Table 12. Number of Items Each Model Fit Best Using Chi-Square Values (20 Items, Simulation Study 1).

True model	MC1	MC2	MC3	MC
All items				
MC1	10.28	9.16	0.56	0.00
MC2	3.00	15.56	1.44	0.00
MC3	0.28	0.52	19.20	0.00
MC	0.64	7.68	2.32	9.36
1 option; Items 1-10				
MC1	5.12	4.32	0.56	0.00
MC2	0.88	7.68	1.44	0.00
MC3	0.28	0.52	9.20	0.00
MC	0.08	5.84	1.84	2.24
2 options; Items 11-20				
MC1	5.16	4.84	0.00	0.00
MC2	2.12	7.88	0.00	0.00
MC3	0.00	0.00	10.00	0.00
MC	0.56	1.84	0.48	7.12

Table 13. Number of Items Each Model Fit Best Using Chi-Square Values (30 Items, Simulation Study 1).

True model	MC1	MC2	MC3	MC
All items				
MC1	10.88	10.80	3.12	5.20
MC2	4.48	13.36	2.68	9.48
MC3	3.00	2.56	15.04	9.40
MC	0.20	3.92	2.88	23.00
1 option; Items 1-10				
MC1	3.12	2.96	2.76	1.16
MC2	2.88	3.28	2.16	1.68
MC3	3.00	2.16	3.04	1.80
MC	0.00	2.64	2.00	5.36
2 options; Items 11-20				
MC1	3.80	3.80	0.36	2.04
MC2	0.72	5.48	0.48	3.32
MC3	0.00	0.40	5.60	4.00
MC	0.00	1.00	0.72	8.28
3 options; Items 21-30				
MC1	3.96	4.04	0.00	2.00
MC2	0.88	4.60	0.04	4.48
MC3	0.00	0.00	6.40	3.60
MC	0.20	0.28	0.16	9.36

sufficient even when the true model is the MC-DINA model. However, when the number of coded options is 3 (and perhaps when it is greater than 3), MC-S-DINA2 is not sufficient when the true model is the MC-DINA model.

The results of the examinee parameters for Simulation Study 2 are shown in Table 14. MC1-2, MC2-2, and MC3-2 are the average recovery rates for the developed three models in Simulation Study 2. The results of Simulation Study 1 (MC1, MC2, MC3) are included for comparison. As shown, MC1-2, MC2-2, and MC3-2 can estimate α_{ik} very well, in particular, equally well as MC1, MC2, and MC3, respectively. Therefore, it has been empirically

Table 14. Average Recovery Rate of Examinee Parameters for Five Analysis Models (Simulation Study 2).

Items	MCI-2	MCI	MC2-2	MC2	MC3-2	MC3
10	0.867	0.865	0.827	0.834	0.840	0.836
20	0.990	0.991	0.957	0.953	0.957	0.950
30	0.996	0.996	0.985	0.986	0.980	0.982

Table 15. Biases and RMSEs of Item Parameters for Three Developed Models (Simulation Study 2).

Model	Bias	RMSE
10 items		
MCI	0.002	0.029
MC2	0.005	0.149
MC3	0.003	0.189
20 items		
MCI	0.005	0.016
MC2	−0.019	0.074
MC3	0.017	0.079
30 items		
MCI	0.012	0.021
MC2	0.014	0.061
MC3	−0.017	0.063

Note. RMSE = root mean square error.

demonstrated that the developed models can estimate α_{ik} even for items that cannot distinguish examinees into one of $C_j^* + 1$ groups by the coded options. The results of the item parameters for Simulation Study 2 are shown in Table 15. This table shows that the biases are close to 0s. However, RMSEs are slightly larger than the results in Table 10. This indicates that the parameters for items that cannot distinguish more than two latent groups have slightly more variance than the items that can distinguish more than two latent groups.

In Simulation Study 2, the means of the correlations between the true α_k and initial α_k^0 for each k in MC-S-DINA1 (the correlations between the true α_k and initial α_k^0 for each k in the DINA model are given in parentheses) are .862 (.594) for 10 items, .868 (.822) for 20 items, and .897 (.860) for 30 items. Because the same simulation data were analyzed, these correlation values are identical to those for the other two developed models. Therefore, good initial values could generally be found, which indicates that the number of MCMC samplings could be reduced.

Discussion

Although the developed models are intended to be used for multiple-choice items, they can also be used for tests for which examinees write how they arrive at their answers. Although the examinee would not have all the attributes needed to correctly answer the item, he or she should be judged as having some of the attributes to answer the item correctly. For example, if an examinee writes $2\frac{4}{12} - \frac{7}{12} = 1\frac{9}{12}$, he or she may have only α_1 and α_2 .

Attributes are assumed to be independent in all three developed models and the MC-DINA model in the simulations. However, this is a strict assumption because if an examinee has an attribute, he or she may be thought of as having a high academic level; therefore, it is natural to

think that he or she may have other attributes. Assuming global academic level θ as in item response theory, de la Torre and Douglas (2004) developed the higher order latent trait model for cognitive diagnosis (HODINA). Therefore, it is necessary to elaborate the developed models by assuming correlations between attributes as in HODINA to model more realistic situations. To do so, a model for the general joint distribution of the attributes is needed, and a different MCMC algorithm would be needed to estimate more general attribute structures. And note that the current findings of this study may not apply to other attribute structures.

The results of the simulation studies presented herein indicate that the developed models have much higher probabilities of correctly estimating examinees' true attribute vectors than the DINA model. This is analogous to the results of the nominal response model (Bock, 1972), where higher information values for examinees' latent trait θ can be obtained than the one-, two-, or three-parameter logistic model. The reason for this is also the same: Both the developed models and the nominal response model make use of the information on which options the examinee selected.

The results of the model comparisons indicate that the best-fit model tends to differ according to the number of coded options. Therefore, when analyzing multiple-choice items, the best-fit model should be examined item by item. This can be done by performing $4^{\text{number of items}}$ times as many analyses in Simulation Study 1 of this article. However, this would have a large cost time. Therefore, a more efficient method following the idea of the generalized-DINA model (de la Torre, 2011) should be used.

Typically, more complicated models require larger sample sizes to be well estimated. Because the sample size was fixed at 1,000 for each case in the simulation studies, these studies cannot distinguish whether simpler models would do better when the sample size is small. And because of the same reason, these studies cannot determine the sample size needed for the chi-square difference test for model selection to be reliable. To examine this point, another simulation study, which compares different sample size cases, is needed.

Although two simulation studies were performed, some real-data studies are necessary to examine the performance of the models in practical settings. When incorrect options are also coded by attributes, effort by the item writers becomes more necessary, as does cooperation between psychometricians and item writers.

The DINA model is not the only CDM. The DINO, NIDA, and R-RUM models are also CDMs. These models are suitable in some situations, depending on the cognitive process for answering binary items. However, the ideas behind these models may also be suitable for the cognitive process for answering multiple-choice items. Therefore, although the three developed models are based on the DINA model, the corresponding CDMs based on the DINO, NIDA, and R-RUM models for multiple-choice items are also needed. And like the generalized-DINA model, the generalization of these models for multiple-choice items is also left for future study.

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) received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

The online appendix is available at <http://apm.sagepub.com/supplemental>

References

- Baker, F. B. (1992). *Item response theory: Parameter estimation techniques*. New York, NY: Dekker.
- Birenbaum, M., Tatsuoka, C., & Xin, T. (2005). Large-scale diagnostic assessment: Comparison of eighth graders' mathematics performance in the United States, Singapore and Israel. *Assessment in Education Principles Policy and Practice*, 12, 167-181.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- 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. (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. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199.
- de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69, 333-353.
- Embretson, S. (1984). A general latent trait model for response processes. *Psychometrika*, 49, 175-186.
- Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation using Multiple Sequences. *Statistical Science*, 7, 457-511.
- Hartz, S. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality* (Doctoral dissertation). University of Illinois at Urbana-Champaign.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187-212.
- Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational Measurement*, 24, 146-178.
- R Development Core Team. (2005). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- 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. (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: Lawrence Erlbaum.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. New York, NY: Routledge Academic.
- Tatsuoka, K. K., Corter, J., & Tatsuoka, C. (2004). Patterns to diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. *American Educational Research Journal*, 41, 901-906.
- Templin, J., & Henson, R. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11, 287-305.