JEM JOURNAL of EDUCATIONAL MEASUREMENT NCME national council on measurement in education

*Journal of Educational Measurement*
*Spring 2010, Vol. 47, No. 1, pp. 115–127*

# A Note on the Invariance of the DINA Model Parameters

**Jimmy de la Torre**
*Rutgers, The State University of New Jersey*
**Young-Sun Lee**
*Teachers College-Columbia University*

*Cognitive diagnosis models (CDMs), as alternative approaches to unidimensional item response models, have received increasing attention in recent years. CDMs are developed for the purpose of identifying the mastery or nonmastery of multiple fine-grained attributes or skills required for solving problems in a domain. For CDMs to receive wider use, researchers and practitioners need to understand the basic properties of these models. The article focuses on one CDM, the deterministic inputs, noisy "and" gate (DINA) model, and the invariance property of its parameters. Using simulated data involving different attribute distributions, the article demonstrates that the DINA model parameters are absolutely invariant when the model perfectly fits the data. An additional example involving different ability groups illustrates how noise in real data can contribute to the lack of invariance in these parameters. Some practical implications of these findings are discussed.*

Although unidimensional item response models (IRMs) are useful for scaling and ordering students on a latent proficiency continuum, they do not allow evaluation of students' specific strengths and weaknesses. For testing and assessment to address needs in practical educational settings and to answer questions of greater relevance, alternative approaches to unidimensional IRMs need to be considered. One promising alternative can be found in latent variable models for mastery vector or cognitive diagnosis. Generally, these models are referred to as cognitive diagnosis models (CDMs) and are developed primarily for the purpose of identifying the mastery or nonmastery of multiple fine-grained attributes required for solving problems on a test. In the cognitive diagnosis parlance, attribute is used as a generic term to refer to a skill, cognitive process, or problem-solving step. Mastery and nonmastery of the attributes are represented by a vector of binary latent variables. From these binary vectors profiles can be generated for individual students or group of students to indicate which attributes students have mastered, and which attributes they have yet to master.

Recent applications and works of cognitive diagnosis modeling show researchers' increasing awareness and recognition of the potential benefits of this approach (e.g., de la Torre, 2006; de la Torre & Lee, 2008; Gorin & Embretson, 2006; Jang, 2009; Roussos, Templin, & Henson, 2007). However, for CDMs to receive wider use, researchers and practitioners need to understand the basic properties of these models. In particular, it is important to understand how the parameters of these models are affected by changes in the characteristics of the underlying attribute distribution. In addition, it is also of value to document how the presence of noise in real data can skew these properties.

One model that can be used for cognitive diagnosis is the *deterministic inputs, noisy "and" gate* (DINA; Junker & Sijtsma, 2001). In this article we investigate the issue of item parameter invariance as it pertains to the DINA model. That is, we examined how the DINA model item parameter estimates vary as the characteristics of the underlying latent distribution change. Specifically, we used simulated data to demonstrate that when the model perfectly fits the data, the item parameter invariance property can be observed; however, when the data contain additional noise attributable to the less than perfect model-data fit, this property does not hold. The DINA model was chosen in this study because of its simple, but interpretable formulation, and it has been the foundation of the other models discussed by Doignon and Falmagne (1999), Haertel (1989), Macready and Dayton (1977), and Tatsuoka (2002). Despite its simplicity, it provides good model-data fit (de la Torre & Douglas, 2008). In addition, as de la Torre and Douglas have shown, a straightforward modification of the DINA model can allow multiple strategies to be easily incorporated in the model. The importance and relevance of the DINA model are further underscored by the recent introduction of more complex and general CDMs based on the DINA model (de la Torre, 2008b, 2009a; de la Torre & Liu, 2008).

## The DINA Model

This section provides a brief background on the DINA model. Let $\mathbf{y}_i$ denote the vector of dichotomous item responses of student $i$ to $J$ items. We assume that the components of $\mathbf{y}_i$ are statistically independent given $\boldsymbol{\alpha}_i = \{\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK}\}'$, the vector of attributes required for the test. The DINA model is a multiple discrete IRM that relates the distribution of $\mathbf{y}_i$ to the attribute vector $\boldsymbol{\alpha}_i$. (For the purposes of this article, CDMs will be used to refer to IRM with underlying multivariate discrete latent variables.) Like many CDMs, implementation of the DINA model requires the construction of a Q-matrix (Tatsuoka, 1983). A Q-matrix is a matrix with $J$ rows and $K$ columns of ones and zeros, where $K$ denotes the number of attributes measured by the test. The $jk$th element of the Q-matrix is one (i.e., $q_{jk} = 1$) if and only if the $k$th attribute is necessary to correctly answer the $j$th item. As a conjunctive model, the DINA model explicitly specifies that students must possess all the required attributes in the $j$th row of the Q-matrix to maximize their probability of a correct response to item $j$.

In the DINA model, a latent response $\eta_{ij}$ is defined deterministically through the equation $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$. The latent response $\eta_{ij}$ assumes a value of one or zero, where $\eta_{ij} = 1$ indicates that student $i$ possesses all the attributes required for item $j$, and $\eta_{ij} = 0$ indicates that student $i$ lacks at least one of the attributes required for item $j$. However, given the stochastic nature of the DINA model, students who possess all the required attributes for an item may slip and answer the item incorrectly; conversely, students who lack at least one of the required attributes for an item may guess and answer the item correctly. The probabilities of slipping and guessing are denoted by $s_j = P(Y_{ij} = 0 \mid \eta_{ij} = 1)$ and $g_j = P(Y_{ij} = 1 \mid \eta_{ij} = 0)$. Under the ideal condition, both the guessing and slip parameters are zero. That is, students who have all the required attributes will not slip and always answer an item correctly, whereas students who lack at least one of the required attributes cannot guess and will always fail to respond correctly. De la Torre (2008a) proposed $\delta_j$ as a discrimination index of item quality that accounts for both the slip and guessing parameters (i.e., $\delta_j = 1 - s_j - g_j$). An item that perfectly discriminates between students with latent responses 0 and 1 has a discrimination index of $\delta_j = 1$ (i.e., $s_j = g_j = 0$). Items with high guessing and slip parameters have smaller discrimination index (i.e., are less discriminating).

Given $s_j$ and $g_j$, the item response function that relates the conditional distribution of the response to item $j$ and the attribute vector $\boldsymbol{\alpha}_i$ is defined as

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

As can be seen from (1), the DINA model is a parsimonious model that requires only two interpretable parameters for the conditional distribution of each item. This property holds true for any value of $K$.

The complete formulation of the DINA model requires specifying the joint distribution of the attribute vector. Two formulations salient to this study are the saturated and higher-order formulations. In the saturated formulation, the probability associated with each of the attribute combinations is considered, thus requiring $2^K$ probabilities to be estimated. In some applications, it is reasonable to impose a hierarchical structure on the attributes (Leighton, Gierl, & Hunka, 2004). In doing so, the number of attribute combinations, hence probabilities, to be considered can be greatly reduced. A hierarchically structured formulation of the attributes represents a special case of the saturated unstructured formulation. In the higher-order formulation, the elements of $\boldsymbol{\alpha}_i$ can be assumed to be conditionally independent by positing a higher-order latent proficiency $\theta_i$. The relationship between the attributes and higher-order proficiency can be expressed using the latent logistic regression model

$$P(\boldsymbol{\alpha}_i \mid \theta_i) = \prod_{k=1}^{K} \left\{ \frac{\exp[1.7\lambda_1(\theta_i - \lambda_{0k})]}{1 + \exp[1.7\lambda_1(\theta_i - \lambda_{0k})]} \right\}. \tag{2}$$

Equation 2 is the familiar one-parameter logistic model except that in this setup, the data matrix to be modeled involves unobserved latent "responses" (i.e., $\alpha_{ik}$), and $\lambda_1$ and $\lambda_{0k}$ are the latent discrimination and difficulty parameters, respectively. In this formulation, attributes with higher $\lambda_{0k}$ are deemed more difficult to master. The DINA model with $\theta_i$ is referred to as the higher-order DINA (HO-DINA; de la Torre & Douglas, 2004) model. Detailed and didactic discussion of these two joint distribution formulations and their corresponding parameter estimation algorithms can be found in de la Torre (2009b) and de la Torre and Douglas (2008). Other formulations that have been considered for the joint distribution of the attributes include the independence and tetrachoric formulations (de la Torre & Douglas, 2004; Hartz, 2002).

Table 1
*Q-Matrix for the 20-Item Test*

| | Attribute | | | | |
|---|---|---|---|---|---|
| Item | 1 | 2 | 3 | 4 | 5 |
| 1 | 1 | 1 | 0 | 0 | 0 |
| 2 | 1 | 0 | 1 | 0 | 0 |
| 3 | 1 | 0 | 0 | 1 | 0 |
| 4 | 1 | 0 | 0 | 0 | 1 |
| 5 | 0 | 1 | 1 | 0 | 0 |
| 6 | 0 | 1 | 0 | 1 | 0 |
| 7 | 0 | 1 | 0 | 0 | 1 |
| 8 | 0 | 0 | 1 | 1 | 0 |
| 9 | 0 | 0 | 1 | 0 | 1 |
| 10 | 0 | 0 | 0 | 1 | 1 |
| 11 | 1 | 1 | 1 | 0 | 0 |
| 12 | 1 | 1 | 0 | 1 | 0 |
| 13 | 1 | 1 | 0 | 0 | 1 |
| 14 | 1 | 0 | 1 | 1 | 0 |
| 15 | 1 | 0 | 1 | 0 | 1 |
| 16 | 1 | 0 | 0 | 1 | 1 |
| 17 | 0 | 1 | 1 | 1 | 0 |
| 18 | 0 | 1 | 1 | 0 | 1 |
| 19 | 0 | 1 | 0 | 1 | 1 |
| 20 | 0 | 0 | 1 | 1 | 1 |

## Simulated Data

### Description of the Data

The primary goal of this section is to demonstrate that, when the DINA model fits the item responses, the parameter estimates, $\hat{s}_j$ and $\hat{g}_j$, will be invariant regardless of the nature of the attribute distribution. In this study, four attribute distributions were considered: three distributions resulting from the HO-DINA model with differing higher-order characteristics and one distribution where the attribute vectors are assumed to be equally probable, as in, the attribute vectors are uniformly distributed. It should be noted that other methods, such as assuming that the attributes are hierarchically structured (Leighton et al., 2004) or the underlying latent traits have a multivariate normal distribution (Hartz, 2002), can also be employed to manipulate the characteristics of the attribute distribution.

Several aspects of the simulated data were held constant: the number of attributes was fixed to $K = 5$, test length to $J = 20$, the slip and guessing parameters to $s = g = .10$, and sample size to $N = 1000$. The Q-matrix used in simulating the data is given in Table 1. There were two types of items based on the number of required attributes (i.e., two- and three-attribute items), and the item types appeared an equal number of times. For the distributions that involve the HO-DINA model, three ability distributions with means $\mu_\theta = \{-1.0, 0.0, 1.0\}$ and a common standard

Table 2
*Mean Proportion of Masters across the Five Attributes*

| Attribute Distribution | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ |
|---|---|---|---|---|---|
| Uniform | .50 | .50 | .50 | .50 | .50 |
| $\theta \sim N(-1, 1)$ | .50 | .36 | .24 | .15 | .08 |
| $\theta \sim N(0, 1)$ | .76 | .64 | .50 | .36 | .24 |
| $\theta \sim N(1, 1)$ | .92 | .85 | .76 | .64 | .50 |

deviation $\sigma_\theta = 1.0$ were used. The higher-order parameters were fixed to $\lambda_{0k} = \{-1.0, -.5, .0, .5, 1.0\}$ and $\lambda_1 = 1.0$ across all conditions. This resulted in the attributes that were ordered in terms of ease of their mastery (i.e., $\alpha_1$ was the easiest to master whereas $\alpha_5$ was the hardest). In this setup, $\theta_i$ was first sampled from a normal distribution with the appropriate mean and standard deviation. The sampled $\theta_i$ was used to generate $\alpha_i$ via Equation 2; $\mathbf{y}_i$ in turn was generated using $\boldsymbol{\alpha}_i$ and the DINA model. For conditions involving equiprobable attribute vectors, $\boldsymbol{\alpha}_i$ was sampled directly using a uniform multinomial distribution.

One hundred data sets were generated for each of the four attribute distribution conditions. The item parameters were estimated via an EM implementation of the marginal maximum likelihood estimation (MMLE/EM) that was implemented using the computer program Ox (Doornik, 2002). The console version of Ox is available free of charge for academic research and teaching purposes, and the MMLE/EM code for the DINA model used in this study can be made available upon request.

## Results

Table 2 shows the proportion of examinees mastering the attributes for each of the four distributions. As expected, the table shows that each of the attributes was mastered by 50% of the examinees when the attribute vectors were uniformly distributed. The table also shows that for a fixed attribute the proportion of masters increased with higher $\mu_\theta$. For example, the mean proportion of examinees who mastered $\alpha_1$ increased from .50 to .76 to .92 as $\mu_\theta$ changed from $-1$ to 0 to 1. The same pattern can be observed for the remaining attributes as well. Consequently, the group with the higher $\mu_\theta$ also had higher overall mastery across the five attributes. Moreover, given the increasing values of $\lambda_{0k}$ used in the HO-DINA model, it was not surprising to observe that, for a fixed value of $\mu_\theta$, the proportion of examinees mastering the different attributes decreased as one moves from $\alpha_1$ through $\alpha_5$. Note also that, although the distributions of the proportion of masters across the different attributes were not the same, examinees from the uniform and $\mu_\theta = 0$ distributions had identical overall mastery.

Table 3 presents the mean absolute bias (MAB) of the guessing and slip parameter estimates by item type (i.e., two- vs. three-attribute items). MAB was calculated by averaging the absolute difference between the true and estimated parameters across the 10 items for each item type. The table shows that both the guessing and slip

Table 3
*Mean Absolute Bias of the Guessing and Slip Parameter Estimates by Item Type*

| Attribute Distribution | Attributes Required | $\overline{\lvert\hat{g} - g\rvert}$ | $\overline{\lvert\hat{s} - s\rvert}$ |
|---|---|---|---|
| Uniform | Two | .000 | .000 |
| | Three | .000 | .001 |
| $\theta \sim N(-1, 1)$ | Two | .006 | .000 |
| | Three | .000 | .002 |
| $\theta \sim N(0, 1)$ | Two | .001 | .000 |
| | Three | .005 | .000 |
| $\theta \sim N(1, 1)$ | Two | .001 | .000 |
| | Three | .001 | .000 |

parameters were accurately estimated across the four attribute distributions. The minimum and maximum MAB for the guessing parameters were .000 and .006, whereas minimum and maximum MAB for the slip parameters were .000 and .002. The results of the simulated data analysis indicate that the invariance property of the DINA model parameters held across attribute distributions. It is worth noting that unlike traditional IRM, the invariance property of the DINA model is absolute in that the parameter estimates obtained using different calibration samples did not require any transformation for them to be comparable.

We would like to point out that the findings above are not limited to the specific conditions and estimation method considered in this article. Our preliminary results showed that the invariance property of the DINA model parameters held for other sample sizes (e.g., 500 examinees), test lengths (e.g., 40 items), and sizes of the guessing and slip parameters (e.g., $s = g = .25$). Moreover, we also verified that when the data were generated and analyzed using the HO-DINA model via the Markov chain Monte Carlo algorithm instead of the MMLE/EM, the property still held.

## Real Data—Fraction Subtraction

### Description of the Data

The data, which represented a subset of the data originally used and described by Tatsuoka (1990) and recently analyzed by Tatsuoka (2005), contained responses to 15 fraction subtraction items by 536 middle school students. Based on the number of correct responses, the examinees were divided into two clusters, below (258 examinees, 48.1%) and above (251 examinees, 46.8%) the median (i.e., eight); students whose scores are exactly at the median (27 examinees, 5%) were randomly assigned to create equal clusters. Three disparate groups of equal size, low-, high-, and average-ability, were formed based on these clusters. For the low-ability group, 225 and 150 examinees were randomly selected from the clusters with number correct below the median and above the median, respectively. For the high-ability group, the number of examinees sampled from each cluster was the reverse of the low-ability group. For the average-ability group, the equal number of examinees was randomly

Table 4
*The Q-Matrix for the Fraction Subtraction Data*

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

Table 5
*Mean and Standard Deviation of Number Correct by Group across Replications*

| Group | Mean | *SD* |
|---|---|---|
| Low Ability | 7.11 | .03 |
| Average Ability | 8.00 | .03 |
| High Ability | 8.90 | .03 |

selected from each cluster. Therefore, about 70% of the 536 observations were used for each group. The process was repeated 1,000 times to produce item parameter bootstrap confidence intervals. Given the Q-matrix in Table 4, the data were calibrated using the same MMLE/EM code in Ox to see how the DINA model parameters lose their invariance property when real data were involved.

The mean and standard deviation of the number correct for the three ability groups across 1,000 replications are given in Table 5. The low- and high-ability groups were approximately 1.8 points apart with respect to the number correct score, with identical standard deviations. Although the mean of the average-ability group was half way between those obtained from the low- and high-ability groups, the standard deviation was the same for the other groups. Figure 1 shows the relative frequency distributions of the mean number correct for each group across the possible number of correct scores. Compared to the other groups, the low-ability group had more observations on the lower end of the distribution; the reverse was true for the high-ability group; the average-ability group was always in between the two groups across all 16 possible scores.

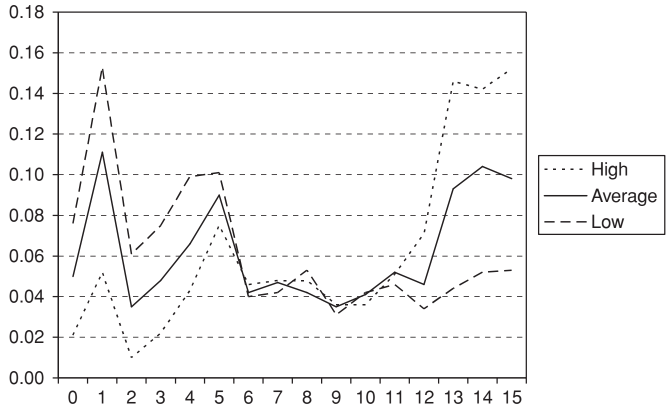


*Figure 1.* Relative frequency distribution: number correct across the three groups (over 1,000 replications).

Table 6
*Points and Interval Estimates of the Guessing Parameter*

| Item | Complete $g$ | Low Ability $g_{.025}$ | Low Ability $g_{.975}$ | Average Ability $g_{.025}$ | Average Ability $g_{.975}$ | High Ability $g_{.025}$ | High Ability $g_{.975}$ |
|---|---|---|---|---|---|---|---|
| 1 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| 2 | .211 | .195 | .218 | .195 | .226 | .194 | .238 |
| 3 | .139 | .066 | .118 | .091 | .190 | .124 | .250 |
| 4 | .125 | .100 | .121 | .112 | .138 | .130 | .158 |
| 5 | .335 | .302 | .342 | .305 | .366 | .310 | .391 |
| 6 | .032 | .026 | .036 | .025 | .039 | .026 | .044 |
| 7 | .072 | .061 | .076 | .062 | .083 | .064 | .091 |
| 8 | .157 | .107 | .152 | .117 | .198 | .133 | .244 |
| 9 | .080 | .040 | .073 | .048 | .110 | .061 | .144 |
| 10 | .170 | .135 | .162 | .154 | .186 | .182 | .218 |
| 11 | .103 | .067 | .102 | .074 | .130 | .082 | .161 |
| 12 | .031 | .020 | .034 | .021 | .039 | .024 | .046 |
| 13 | .134 | .109 | .130 | .120 | .147 | .138 | .169 |
| 14 | .022 | .010 | .021 | .015 | .028 | .022 | .038 |
| 15 | .010 | .006 | .013 | .006 | .014 | .005 | .016 |

## Results

Tables 6 and 7 give the 95% bootstrap confidence limits based on the percentile method for the guessing and slip parameters, respectively, across the three ability groups. For comparison purposes, the estimates obtained using all 536 observations are also presented. Table 6 shows that the guessing parameter estimates obtained using the average-ability group were similar to those obtained using the entire sample in that the confidence intervals of the average-ability group contained all the estimates based on the entire sample. However, interval estimates for eight and three items in

Table 7
*Points and Interval Estimates of the Slip Parameter*

| Item | Complete $s$ | Low Ability $s_{.025}$ | Low Ability $s_{.975}$ | Average Ability $s_{.025}$ | Average Ability $s_{.975}$ | High Ability $s_{.025}$ | High Ability $s_{.975}$ |
|---|---|---|---|---|---|---|---|
| 1 | .277 | .325 | .354 | .258 | .293 | .204 | .239 |
| 2 | .118 | .107 | .143 | .105 | .130 | .104 | .123 |
| 3 | .038 | .038 | .052 | .031 | .045 | .027 | .039 |
| 4 | .131 | .108 | .162 | .110 | .150 | .109 | .136 |
| 5 | .248 | .248 | .278 | .233 | .263 | .222 | .247 |
| 6 | .226 | .213 | .260 | .207 | .244 | .207 | .231 |
| 7 | .078 | .068 | .099 | .067 | .090 | .064 | .081 |
| 8 | .048 | .036 | .059 | .039 | .057 | .042 | .054 |
| 9 | .062 | .065 | .085 | .053 | .071 | .045 | .061 |
| 10 | .070 | .052 | .088 | .055 | .082 | .058 | .076 |
| 11 | .105 | .109 | .134 | .092 | .117 | .082 | .102 |
| 12 | .133 | .116 | .154 | .117 | .147 | .120 | .140 |
| 13 | .158 | .143 | .181 | .142 | .174 | .144 | .164 |
| 14 | .197 | .172 | .231 | .174 | .218 | .175 | .205 |
| 15 | .182 | .165 | .208 | .165 | .201 | .166 | .190 |

the low- and high-ability groups did not contain the guessing parameter estimates based on the complete data. It can be noted that, relative to the estimates based on the entire data, the interval estimates for the guessing parameters of items 3, 4, 8, 9, 10, 11, 13, and 14 using the low-ability students were underestimates, whereas the interval estimates for the guessing parameters of items 4, 10, and 13 using the high-ability students were overestimates.

Similar to the case of the guessing parameters, Table 7 shows that the interval estimates based on the average-ability group captured the slip parameter estimates based on the entire data. However, instead of the low-ability group, the high-ability group had more interval estimates that did not capture the complete-data slip parameter estimates, albeit the difference was small (i.e., three versus four). In addition, the estimates from the low-ability group (items 1, 9, and 11) were overestimates, whereas those from the high-ability group (items 1, 5, 9, and 11) were underestimates.

The impact of the different ability distributions on the posterior probability of the attribute vector was also examined. Table 8 summarizes the posterior probabilities for the 32 attribute vectors, and based on these posterior probabilities, the fraction subtraction data resulted in only nine unique attribute classes. For example, the row where only $\alpha_1 = 0$ is displayed indicates that students who have not mastered the first attribute cannot be further distinguished from each other. Using the entire data, the proportion of students in any of the 16 attribute vectors where $\alpha_1 = 0$ is .0127; equivalently, the proportion of students that have not mastered $\alpha_1$ is $.0127 \times 16 = .2030$. For all four distributions, the bulk of the probabilities (at least .88) were concentrated on the following attribute vectors: $\alpha_1 = 0$, $\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\alpha_1 =$

Table 8
*Posterior Probability Distribution for Three Ability Groups*

| $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ | Complete | Low | Average | High |
|---|---|---|---|---|---|---|---|---|
| 0 | | | | | .0127 | .0142 | .0127 | .0109 |
| 1 | 0 | 0 | | | .0073 | .0091 | .0073 | .0059 |
| 1 | 0 | 1 | 0 | | .0197 | .0262 | .0198 | .0144 |
| 1 | 0 | 1 | 1 | 0 | .0066 | .0069 | .0066 | .0061 |
| 1 | 0 | 1 | 1 | 1 | .0016 | .0024 | .0017 | .0008 |
| 1 | 1 | 0 | | | .0048 | .0065 | .0048 | .0029 |
| 1 | 1 | 1 | 0 | | .1142 | .1321 | .1142 | .0977 |
| 1 | 1 | 1 | 1 | 0 | .1038 | .0873 | .1038 | .1238 |
| 1 | 1 | 1 | 1 | 1 | .3687 | .2977 | .3686 | .4356 |

$\alpha_2 = \alpha_3 = \alpha_4 = 1$, and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 1$. The starkest difference between these distributions can be observed by comparing the $P(\alpha_1 = 0 \mid \mathbf{X})$ and $P(\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 1 \mid \mathbf{X})$. Although the former was always smaller than the latter, $P(\alpha_1 = 0 \mid \mathbf{X})/P(\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 1 \mid \mathbf{X})$ was larger for groups with lower abilities: the ratios were .76, .55, and .40 for the low-, average-, and high-ability groups, respectively.

The results of the fraction subtraction data analysis indicate that when real data were involved, the invariance property of the DINA model parameters may not hold. In addition, the resulting attribute distribution determines how many and which parameters will be affected, and what the pattern of underestimation and overestimation will be. To verify that the above findings were not artifacts of the Q-matrix, we generated data using the item parameter estimates based on the entire data and the different posterior distributions of the attribute vectors above for a fixed sample size of 375. With 1,000 replications for each ability group, we verified that the DINA model parameters were also invariant for this Q-matrix and these attribute distributions.

## Summary and Discussion

This article demonstrates that the DINA model parameters are absolutely invariant, but this invariance property holds only to the extent that the model fits the data. With the simulated data where the model perfectly fits the data, the same item parameter estimates were obtained across calibration samples with dramatically different underlying attribute distributions; with the presence of noise, as in the fraction subtraction data, discrepancies in the item parameter estimates were observed with varying attribute distributions. However, this should not be a reason to downplay the practical usefulness of the DINA model. A paradoxical result has been observed with more common IRMs—although more complex models such as the three-parameter logistic model provide better model-data fit, simpler models such as the one-parameter logistic model yield parameter estimates that are more invariant (e.g., Custer et al., 2008; Fan & Ping, 1999; Forsyth, Saisangian, & Gilmer, 1981; Wright, 1992). Compared to other CDMs, the DINA model represents one of the simplest, if not the simplest model for cognitive diagnosis. To the extent that the relationship between model fit and item parameter invariance in IRMs applies to CDMs, one can expect the DINA model to provide relatively invariant item parameter estimates even when the model does not perfectly fit the data.

The invariance property of the DINA model vis-à-vis model- and nonmodel-fitting data has several practical implications. First, the invariance property, or lack thereof, of the DINA model parameters can be used as an indirect measure of model-data fit. In situations where sufficiently large sample size is available, two disparate groups can be created to calibrate the tests separately. By examining the discrepancy between the two sets of item parameter estimates, one can gauge the extent to which the posited model deviates from the data. Second, in situations where the model does not perfectly fit the data, as one would expect in practice, the impact of model misspecification on parameter estimates can be minimized by avoiding samples that represent extreme underlying distributions. As the real data analysis shows, this type of sample resulted in more extreme item parameter estimates. However, even if the model's fit to the data is adequate, it would still be prudent to employ representative samples whenever possible, not so much to minimize bias, but to increase the precision of the estimates—estimates based on more extreme distributions, in addition to being biased, have been shown to have larger standard errors (de la Torre, 2005).

This article has also demonstrated that explicit equating is not necessary with the DINA model under two conditions: when the model is correctly specified, or when samples with similar characteristics are involved. However, these conditions may not always be satisfied in many applied settings. Consequently, one needs to devise methods by which item parameters can be made comparable when the calibration samples involved are markedly disparate and model-data fit is less than ideal. For example, one can consider the different methods used by Xu and von Davier (2008) as they apply to DINA model.

As a final note, it would be important to underscore that although the current work was carried out in the context of one specific model—the DINA model—its implications can be far-reaching. This is so because, as alluded to earlier, several recently developed CDMs, which have greater complexity and generality, are explicitly based on this model. These include DINA-based models for continuous response, nominal response, cognitively based multiple choice options, multiple strategies, and a generalization of the DINA model (de la Torre, 2008b, 2009a; de la Torre & Liu, 2008; Templin, Henson, Rupp, & Jang, 2008). It would be instructive to examine how the findings of this study extend to these new models.

## References

Custer, M., Sharairi, S., Yamazaki, K., Signatur, D., Swift, D., & Frey, S. (2008, March). A paradox between IRT invariance and model-data fit when utilizing the one-parameter and three-parameter models. Paper presented at the meeting of the American Educational Research Association, New York, NY.

de la Torre, J. (2005, April). Model fit and parameter invariance: The case of the higher-order DINA model. Paper presented at the meeting of the National Council on Measurement in Education, Montreal, Canada.

de la Torre, J. (2006, June). Attribute vector profile comparisons at the state level: An application and extension of cognitive diagnosis modeling in NAEP. Paper presented at the international meeting of the Psychometric Society, Montreal, Canada.

de la Torre, J. (2008a). 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. (2008b, July). The generalized DINA model. Paper presented at the international meeting of the Psychometric Society, Durham, NH.

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

de la Torre, J., & Douglas, J. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, *74*, 595–624.

de la Torre, J., & Lee, Y. -S. (2008, March). Cognitive diagnosticity of IRT-constructed assessment: An empirical investigation. Paper presented at the meeting of the National Council on Measurement in Education, New York, NY.

de la Torre, J., & Liu, Y. (2008, March). A cognitive diagnosis model for continuous response. Paper presented at the meeting of the National Council on Measurement in Education, New York, NY.

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

Doornik, J. A. (2002). *Object-oriented matrix programming using Ox* (Version 3.1). [Computer software]. London, England: Timberlake Consultants Press.

Fan, X., & Ping, Y. (1999, April). Assessing the effect of model-data misfit on the invariance property of IRT parameter estimates. Paper presented at the meeting of the American Educational Research Association, Montreal, Canada.

Forsyth, R., Saisangian, U., & Gilmer, J. (1981). Some empirical results related to the robustness of the Rasch model. *Applied Psychological Measurement*, *5*, 175–186.

Gorin, J. S., & Embretson, S. E. (2006). Item difficulty modeling of paragraph comprehension items. *Applied Psychological Measurement*, *30*, 394–411.

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

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

Jang, E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for fusion model application to language assessment. *Language Testing*, *26*, 31–73.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*, 258–272.

Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement*, *41*, 205–236.

Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, *33*, 379–416.

Roussos, L., Templin, J., & Henson, R. (2007). Skills diagnosis using IRT-based latent class models. *Journal of Educational Measurement*, *44*, 293–311.

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, C. (2005). Data-analytic methods for latent partially ordered classification models: Corrigendum. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*, 465–467.

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. Safto (Eds.). *Monitoring attribute vector and knowledge acquisition* (pp. 453–488). Hillsdale, NJ: Erlbaum.

Templin, J., Henson, R., Rupp, A., & Jang, E. (2008, April). Cognitive diagnosis models for nominal response data. Paper presented at the meeting of the National Council on Measurement in Education, New York, NY.

Wright, B. D. (1992). IRT in the 1990s: Which models work best? *Rasch Measurement Transactions*, *6*, 196–200.

Xu, X., & von Davier, M. (2008). *Linking for the general diagnostic model* (ETS Research Report No. RR-08–08). Princeton, NJ: Educational Testing Service.

## Authors

JIMMY DE LA TORRE is Associate Professor of Educational Psychology at Rutgers University, 10 Seminary Place, New Brunswick, NJ, 08901; j.delatorre@rutgers.edu. His primary research interests include item response theory, cognitive diagnosis, Bayesian analysis, and the use of diagnostic assessments to support classroom instruction and learning.

YOUNG-SUN LEE is Associate Professor of Psychology and Education, Teachers College-Columbia University, 525 West 120th Street, New York, NY 10027; yslee@tc.columbia.edu. Her primary research interests include applications of item response theory and cognitive diagnosis modeling.