Bayesian DINA Modeling Incorporating Within-Item Characteristic Dependency

Applied Psychological Measurement

2019, Vol. 43(2) 143–158

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0146621618781594

journals.sagepub.com/home/apm

Peida Zhan¹ , Hong Jiao², Manqian Liao² and Yufang Bian¹

Abstract

The within-item characteristic dependency (WICD) means that dependencies exist among different types of item characteristics/parameters within an item. The potential WICD has been ignored by current modeling approaches and estimation algorithms for the deterministic inputs noisy “and” gate (DINA) model. To explicitly model WICD, this study proposed a modified Bayesian DINA modeling approach where a bivariate normal distribution was employed as a joint prior distribution for correlated item parameters. Simulation results indicated that the model parameters were well recovered and that explicitly modeling WICD improved model parameter estimation accuracy, precision, and efficiency. In addition, when potential item blocks existed, the proposed modeling approach still demonstrated good performance and high robustness. Furthermore, the fraction subtraction data were analyzed to illustrate the application and advantage of the proposed modeling approach.

Keywords

cognitive diagnosis, within-item characteristic dependency, cognitive diagnosis model, DINA model, Bayesian estimation

Introduction

Currently, a number of cognitive diagnosis models (CDMs) have been developed (see, for example, Rupp, Templin, & Henson, 2010). The deterministic-inputs, noisy “and” gate (DINA) model (Junker & Sijtsma, 2001) is the most popular one because of its interpretability. Different parameterizations of the DINA model (e.g., DeCarlo, 2011; de la Torre, 2011; Henson, Templin, & Willse, 2009; Junker & Sijtsma, 2001; von Davier, 2014) and estimation algorithms (e.g., Culpepper, 2015; de la Torre, 2009) have been proposed. However, the potential within-item characteristic dependency (WICD; Fox, 2010) has been ignored by these modeling approaches and estimation algorithms, which may underestimate the real-world complexity. The WICD means that dependencies exist among different item characteristics/parameters

¹Zhejiang Normal University, Zhejiang, China

²University of Maryland, College Park, USA

Corresponding Authors:

Peida Zhan, Department of Psychology, College of Teacher Education, Zhejiang Normal University, No. 688 Yingbin Road, Jinhua, Zhejiang, 321004, P. R. China.

Email: pdzhan@gmail.com

Yufang Bian, Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, No. 19, Xijiekouwai St., Haidian District, Beijing, 100875, P. R. China.

Email: bianyufang66@126.com

within an item. More specifically, the guessing and slipping parameters in the DINA model are not independent of each other. A correlation could be used to quantify such a relationship between the guessing the slipping parameters in the DINA model.

The correlation among item parameters in the DINA model was supported by many empirical studies. For example, a correlation coefficient of $-.892$ was found between the guessing and slipping parameters based on the examination for the certificate of proficiency in English data in a study of Templin and Hoffman (2013); a correlation coefficient of $-.802$ was found for the revised Purdue spatial visualization tests—visualization of rotations data in Culpepper (2015); a correlation coefficient of $-.686$ was found for the FCAT (Florida Comprehensive Assessment Test) 2003 ninth grade mathematics data in Lee, de la Torre, and Park (2012); furthermore, a correlation coefficient of $-.415$ was found for the fraction subtraction data in de la Torre (2009). It can be seen that the estimated guessing and slipping parameters in the DINA model were negatively correlated based on the real data from reading, mathematics, and cognitive psychological tests although the WICD was ignored in the modeling and estimation process. Thus, this study aims at accounting for WICD in the DINA model within the Bayesian framework.

The rest of the article starts with a review of the DINA model. Then, it demonstrates a way to account for WICD in the DINA model, which is estimated with the Bayesian approach. Three simulation studies evaluating the feasibility of the proposed modeling approach are followed. Finally, an empirical example is presented to illustrate the application of the proposed approach for WICD.

Standard Bayesian DINA Modeling

Let $\mathbf{Y}_{ni}$ ($\mathbf{Y}$ is an N -by- I matrix) be the observed response of person n ($n = 1, \dots, N$) to item i ($i = 1, \dots, I$). In the DINA model, the relationship among attributes and observed responses can be written as follows:

$$p_{ni} = P(\mathbf{Y}_{ni} = 1 | \boldsymbol{\alpha}_n = \boldsymbol{\alpha}_c, \boldsymbol{\Omega}_i) = g_i + (1 - s_i - g_i) \prod_{k=1}^K \alpha_{ck}^{q_{ik}}, \quad (1)$$

where p_{ni} is the probability of person n with attribute profile c correctly responding to item i ; α_{ck} is the k th element of attribute profile c ($c = 1, \dots, C$; $k = 1, \dots, K$), where $\alpha_{nk} = \alpha_{ck} = 1$ if person n masters the k th attribute of attribute profile c , and $\alpha_{nk} = \alpha_{ck} = 0$ otherwise; $\boldsymbol{\Omega}_i = (g_i, s_i)'$ is a vector of item parameters, where s_i and g_i are the slipping and guessing probabilities of item i , respectively, which describe the item-level aberrant responses; $(1 - s_i - g_i)$ is the item discrimination index (IDI; de la Torre, 2008a). The Q-matrix (Tatsuoka, 1983) is an I -by- K matrix with element q_{ik} indicating whether attribute k is required to correctly answer item i (i.e., $q_{ik} = 1$ if the attribute is required, and 0 otherwise).

Let $\pi_c = P(\boldsymbol{\alpha}_c)$ be the marginal probability of attribute profile c in the population (i.e., the mixing proportion) and $\boldsymbol{\pi} = (\pi_1, \dots, \pi_C, \dots, \pi_C)'$ be a C -dimensional probabilistic vector of attribute profiles. Given the local independence assumption, the likelihood of the observed item responses based on the DINA model (Culpepper, 2015; Macready & Dayton, 1992) can be written as follows:

$$P(\mathbf{Y} | \boldsymbol{\pi}, \boldsymbol{\Omega}) = \prod_{n=1}^N \sum_{c=1}^C \pi_c \prod_{i=1}^I p_{ni}^{Y_{ni}} (1 - p_{ni})^{1-Y_{ni}}. \quad (2)$$

In the Bayesian framework, the posterior distribution of the model parameters is proportional to the product of their prior distributions and the likelihood of item response data, which is given by,

$$P(\boldsymbol{\pi}, \boldsymbol{\Omega} | \mathbf{Y}) \propto P(\mathbf{Y} | \boldsymbol{\pi}, \boldsymbol{\Omega}) P(\boldsymbol{\pi}) P(\boldsymbol{\Omega}), \quad (3)$$

where $P(\mathbf{Y} | \boldsymbol{\pi}, \boldsymbol{\Omega})$ is the likelihood of the item response data for the DINA model (see Equation 2), $P(\boldsymbol{\pi})$ is the prior distribution of mixing proportions, and $P(\boldsymbol{\Omega})$ is the prior distributions of item parameters. A common practice for estimating the DINA model parameters is to use univariate priors for item parameters, which assumes that the item guessing and slipping parameters are independent of each other in the prior distributions:

$$P(\boldsymbol{\Omega}) = \prod_{i=1}^I P(\boldsymbol{\Omega}_i) = \prod_{i=1}^I P(s_i) P(g_i), \quad (4)$$

where $P(g_i)$ and $P(s_i)$ are the prior of item guessing and slipping parameters, respectively. In general, a monotonicity restriction ($g_i < (1 - s_i)$) can be imposed on Equation 4. The monotonicity restriction imposed a within-item scale/range dependence upon slipping and guessing parameters, but it is different from WICD.

As the guessing and slipping parameters are on the probability scale, the Beta distributions are often specified as their prior distributions,

$$s_i \sim \text{Beta}(a_{s_i}, b_{s_i}) \text{ and } g_i \sim \text{Beta}(a_{g_i}, b_{g_i}), \quad (5)$$

where $a_{s_i}, b_{s_i}, a_{g_i}, b_{g_i}$ are the scale parameters. As suggested by Culpepper (2015), the scale parameters are assigned as $a_{s_i} = b_{s_i} = a_{g_i} = b_{g_i} = 1$ for each item. In a word, univariate uniform priors of item parameters are used by the standard Bayesian DINA modeling approach.

Bayesian DINA Modeling Incorporating the WICD

Although a priori independence assumption does not imply the parameter independence in the posterior distribution (Levy & Mislevy, 2016), the WICD is not explicitly modeled. To account for the WICD, a multivariate prior distribution could be specified for the item parameters (Mislevy, 1986). In the item response theory framework, some studies (e.g., Glas & van der Linden, 2003; van der Linden, 2007) have explored to account for WICD using the multivariate normal distribution. The same method could be applied to the DINA model to account for WICD. In this study, a bivariate normal distribution is employed as a joint prior distribution of item parameters, as follows:

$$\begin{pmatrix} \text{logit}(g_i) \\ \text{logit}(s_i) \end{pmatrix} = \begin{pmatrix} \beta_i \\ \delta_i \end{pmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (6)$$

where $\text{logit}(x) = \log(x / (1 - x))$, such transformation frees up the restriction of range for the original g_i and s_i parameters (i.e., 0-1); $\boldsymbol{\mu} = (\mu_\beta, \mu_\delta)'$ is a mean vector and $\boldsymbol{\Sigma}$ is a variance-covariance matrix, and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_\beta^2 & \rho_{\beta\delta} \sigma_\beta \sigma_\delta \\ \rho_{\beta\delta} \sigma_\beta \sigma_\delta & \sigma_\delta^2 \end{pmatrix}, \quad (7)$$

where $\rho_{\beta\delta}$ is the correlation coefficient between the logit-transformed item parameters.

Overall, the main difference between the standard and the proposed Bayesian DINA modeling approaches is the difference in the priors of the item parameters, that is, a truncated bivariate uniform prior versus a joint bivariate normal prior distribution. For illustrative purposes, the proposed modeling approach is denoted as the W-DINA and the regular one is denoted as the

U-DINA throughout this article. It is noted that the W-DINA and the U-DINA have the same item response function but the former uses a more informative prior for item parameters than the latter to account for the WICD.

Bayesian Parameter Estimation

The Bayesian Markov chain Monte Carlo (MCMC) method was used to estimate model parameters. In this study, the JAGS (version 4.3.0; Plummer, 2015) were used to automate the estimation process. The JAGS uses a default option of the Gibbs sampler (Gelfand & Smith, 1990). JAGS code for the W-DINA and the U-DINA are provided in Table S1 in online supplements.

To begin with, given local independence, $\mathbf{Y}_{ni} \sim \text{Bernoulli}(p_{ni})$. Furthermore, the prior distribution of the person parameters is assumed to be $\boldsymbol{\pi} = (\pi_1, \dots, \pi_C) \sim \text{Dirichlet}(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_C)'$, $\sum_{c=1}^C \pi_c = 1$, $c \sim \text{Categorical}(\boldsymbol{\pi})$, where C is the number of possible attribute profiles; $\boldsymbol{\pi}$ is assumed to follow a Dirichlet distribution, an uninformative prior is used by setting $\lambda_1 = \dots = \lambda_C = 1$; $c \in \{1, \dots, C\}$, indicating person n 's attribute profile, is assumed to follow a categorical distribution, with the probability of membership in the c th profile, then, $\boldsymbol{\alpha}_n = \boldsymbol{\alpha}_c$ (i.e., $\alpha_{nk} = \alpha_{ck}$). The posterior mode is used as the estimated value of each attribute given its categorical nature.

In addition, in the W-DINA, the prior of the item parameters is assumed to be a joint bivariate normal distribution as follows,

$$\begin{pmatrix} \text{logit}(g_i) \\ \text{logit}(s_i) \end{pmatrix} = \begin{pmatrix} \beta_i \\ \delta_i \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_\beta \\ \mu_\delta \end{pmatrix}, \boldsymbol{\Sigma} \right),$$

with the hyper priors specified as $\mu_\delta \sim N(-2.197, 2)$, $\mu_\beta \sim N(-2.197, 2)$, $\boldsymbol{\Sigma} \sim \text{InvWishart}(\mathbf{R}, 2)$, where $\mathbf{R}$ is a two-dimensional identity matrix. The posterior mean is used as the estimated value of each item parameter given its continuous nature.

Simulation Study

Three simulation studies were conducted. The purposes of Study 1 are two-fold. First, to examine whether parameters of the W-DINA can be recovered accurately; second, to evaluate the consequences of ignoring WICD, in which the data were simulated from the W-DINA, but analyzed with the U-DINA. However, the method of generating item parameters in Study 1 yields values in a subset of the identified space. It is, therefore, expected that an informative prior would outperform a uniform prior. Thus, the main purpose of Study 2 is to evaluate the performance of the W-DINA when the item parameters were simulated from a uniform distribution of the whole identified space. In addition, in practice, items in the test may be grouped into several item blocks. For instance, content constraints may be taken into consideration in test construction (Yi & Chang, 2003). It is possible that items in the same item block have more similar characteristics. Thus, Study 3 mainly investigated the impact of item blocks on parameter estimation of the W-DINA.

Simulation Study 1

Design and data generation. In Study 1, five independent variables were manipulated including (a) the degree of the WICD ($\rho_{\beta\delta}$) at two levels of -0.3 and -0.9 ; (b) the values of μ_β and μ_δ ($\boldsymbol{\mu}$) at two levels of $\mu_\beta = \mu_\delta = -2.197$ and $\mu_\beta = \mu_\delta = -1.386$, indicating that the mean quality of all items is relatively high ($\bar{g} \approx \bar{s} \approx 0.1$) and relatively low ($\bar{g} \approx \bar{s} \approx 0.2$), respectively; (c)

Table 1. Recovery of Item Parameters in Study 2.

<i>I</i>	<i>N</i>	<i>g_i</i>				<i>s_i</i>			
		MA_bias		M_RMSE		MA_bias		M_RMSE	
		W-DINA	U-DINA	W-DINA	U-DINA	W-DINA	U-DINA	W-DINA	U-DINA
15	200	0.011	0.009	0.054	0.050	0.021	0.019	0.115	0.106
	1,000	0.006	0.005	0.032	0.030	0.009	0.008	0.069	0.064
30	200	0.008	0.007	0.042	0.040	0.014	0.013	0.094	0.087
	1,000	0.004	0.004	0.023	0.022	0.008	0.007	0.051	0.047

Note. *I* = test length; *N* = sample size; MA_Bias = mean absolute value of bias across all items; M_RMSE = mean value of root mean square error across all items; W = W-DINA; U = U-DINA; DINA = deterministic inputs noisy “and” gate.

sample sizes (*N*) at two levels of 200 and 1000; (d) test length (*I*) at two levels of 15 and 30; and (e) estimation approaches (EA) at two levels of the W-DINA and the U-DINA. Then, *g_i* and *s_i* were generated from a bivariate normal distribution (Equation 6) with variance of 1. Five attributes (*K* = 5) were measured and the simulated Q-matrices are presented in Figure 1. The true attribute profile of each person is randomly chosen from all possible patterns (*C* = 32) with equal probability.

Analysis. Thirty replications were implemented in each simulated study condition. For each replication, two Markov chains with random starting points were used and 10,000 iterations were run for each chain. The first 5,000 iterations in each chain were discarded as burn-in. Finally, 10,000 iterations were retained for model parameter inferences. The potential scale reduction factor (PSRF; Brooks & Gelman, 1998) was computed to assess the convergence of each parameter. In our studies, the PSRFs were generally less than 1.01, suggesting good convergence in the specified settings.¹ To evaluate parameter recovery, the bias and root mean square error (RMSE) of the item parameter estimates were computed. The pattern correct classification rates (PCCRs) were computed to evaluate the classification accuracy. In addition, the number of iterations for convergence, the effective sample sizes (Kass, Carlin, Gelman, & Neal, 1998), and the overall computing time² for the same total number of iterations were employed to evaluate the difference in computational efficiency between the two modeling approaches. The ratio of the computing time for the U-DINA over that for the W-DINA was reported.

Results. Figures 2 and 3 present the recovery of item parameters for the W-DINA and the U-DINA in conditions with a test length of 30 items. Due to page limit, the results for conditions with 15 items were presented in Figures S1 and S2 in online supplements. First, the W-DINA and the U-DINA are similar in the patterns of the item parameter recovery. For the *s_i* parameter, more required attributes lead to higher bias and RMSE. By contrast, the estimation of the *g_i* parameters was less biased and more accurate than the *s_i* parameters, and the variability of the *g_i* parameter estimates was slightly smaller for items that required more attributes. As expected, the bias and RMSE both decreased as sample size increased. The mean quality of items slightly affects the recovery of item parameters; its impact was more evident in conditions with short test length, and the impact on the U-DINA was more noticeable than that on the W-DINA.

Regarding the difference between the two modeling approaches, the recovered *g_i* parameters between them were quite similar; however, the bias and RMSE of the W-DINA were slightly smaller than those of the U-DINA, and such advantages increased as sample size decreased. By contrast, the recovery of the *s_i* parameters for the W-DINA was more accurate and more precise

Figure 1. *K*-by-*I* Q matrix for Study I.
Note. Blank means “0,” gray means “1”; “*” denotes items used in *I* = 15 conditions. *K* = the number of attributes; *I* = test length; Q = Q-matrix.

Figure 2. The recovery of the guessing parameters across *I* = 30 conditions in Study I.
Note. The values were estimated from 30 replications. *I* = test length; *N* = sample size; μ = mean quality of items; ρ = the strength of WICD; WICD = within-item characteristic dependency; W = W-DINA; DINA = deterministic inputs noisy “and” gate; U = U-DINA; RMSE = root mean square error.

than the U-DINA, and such advantages also increased as sample size decreased. The W-DINA has outperformed the U-DINA in terms of the s_i parameters in items with multiple attributes. Such results were within expectation because the recovery of the s_i parameters in items with multiple attributes was worst in the U-DINA, which means the room for potential improvement was the greatest. Likewise, the W-DINA also slightly outperformed the U-DINA in terms of the g_i parameters in items with fewer attributes.

Obviously, the sample size and the strength of WICD had more impact on the difference between these two modeling approaches. The differences for the conditions with a small sample size were significantly larger than those with a large sample size. A possible explanation is that when the sample size is large enough, response data provide enough information to assure better accuracy in item parameter estimation. Thus, the influence of the prior distribution decreases. In addition, the W-DINA led to higher parameter estimation accuracy as the strength of WICD increased. Essentially, modeling the WICD borrow information among item parameters in estimation. That is, this joint distribution is an augmented approach in utilizing extra information in the g_i parameters when estimating the s_i parameters.

For the recovery of attribute profiles see Figures S3 and S4 in online supplements. First, the patterns in the recovery of attribute profiles for the W-DINA were similar to those for the U-

Figure 3. The recovery of the slipping parameters across $l = 30$ conditions in Study 1.
 Note. The values were estimated from 30 replications. l = test length; N = sample size; μ = mean quality of items; ρ = the strength of WICD; WICD = within-item characteristic dependency; W = W-DINA; DINA = deterministic inputs noisy “and” gate; U = U-DINA; RMSE = root mean square error.

DINA. In almost all the conditions, the PCCR decreased as the magnitude of the WICD increased. A larger number of items and higher item mean quality improved the attributes recovery for both models. Regarding the difference between the two modeling approaches. For a given condition, the W-DINA was slightly better than the U-DINA. Although the advantages in the PCCR of the W-DINA were less than 0.015, the advantages increased as the test length increased, the extent of the WICD increased, the mean quality of items decreased, or the sample size decreased.

For the number of iterations needed to reach convergence, the multivariate potential scale reduction factor (MPSRF) was computed (Brooks & Gelman, 1998) for each replication. As similar patterns were found in different conditions, Figure 4 presents the MPSRF in several conditions for illustration. First, for a given condition, the pattern of the MPSRF for the W-DINA was quite similar to that for the U-DINA, which means that imposing the WICD does not evidently affect the convergence of the posterior distribution. Second, test length affects convergence, whereas the sample size, the strength of the WICD, or the mean quality of items does not. Although employing a burn-in of 1,000 iterations is adequate for both models, a burn-in of 5,000 iterations was employed in this study to be conservative.

In addition, Figure 5 plots the mean, minimum, and maximum effective sample sizes of each item parameter across 30 replications to provide evidence regarding the degree of autocorrelation in the Markov chains. Larger effective sample sizes are associated with smaller autocorrelation. As similar patterns were found in different conditions, only limited conditions were presented. The effective sample size was computed from two chains with all iterations, resulting in an expected value of 20,000. The results indicated that imposing WICD has no significant effect on effective sample sizes.

The sample size had some effect on the ratio of overall computing time of the U-DINA over the W-DINA, and the ratios were almost the same for conditions with the same sample size. The mean ratio is 1.418 ($SD = 0.011$) for conditions with a sample size of 1,000 and 1.313

Figure 4. Plots of Brooks–Gelman MPSRF for selected conditions in Study 1.

Note. For the W-DINA, MPSRF assesses convergence of item parameters, mixing proportions, mean vector, and variance and covariance matrix; for the U-DINA, MPSRF assesses convergence of item parameters and mixing proportions; each panel includes plots of MPSRF from 30 replications; the plots are based upon chain lengths from 50 to 750 in increments of 50. MPSRF = multivariate potential scale reduction factor; DINA = deterministic inputs noisy “and” gate; I = test length; N = sample size; μ = mean quality of items; ρ = the strength of within-item characteristic dependency.

($SD = 0.004$) for conditions with a sample size of 200, respectively. Consequently, the W-DINA is more efficient in parameter estimation than the U-DINA.

Overall, both the W-DINA and the U-DINA parameters were well recovered. Explicitly, modeling the WICD can further improve model parameter estimation accuracy, precision, and efficiency, especially when the sample size is small.

Simulation Study 2

Design and data generation. In Study 2, two independent variables were manipulated including (a) the test length (I) at two levels of 15 and 30 and (b) the sample sizes (N) at two levels of 200 and 1000. The Q-matrices in Figure 1 were still used. The true attribute profile of each person was still randomly chosen from 32 possible patterns with equal probability. To shed light on the potential limitations or the performance in boundary conditions of the proposed modeling approach, item parameters were generated from $s_i \sim U(0, 1)$ and $g_i | s_i \sim U(0, 1 - s_i)$. When item parameters were uniformly generated, it means the items were developed without careful screening for quality though such study conditions may rarely exist in practice.

Results. Table 1 summarizes the recovery of item parameters. The disadvantages of the W-DINA were around 0.002 in mean absolute bias and were less than 0.01 in mean RMSE, respectively. Table 2 presents the recovery of attribute parameters. The disadvantages of the W-DINA were less than 0.01 in PCCR across all conditions. Note that such a low PCCR was expected because of unrealistically low item quality. Overall, the W-DINA can provide a comparable recovery of model parameters to the U-DINA in the simulated boundary conditions.

Figure 5. The mean, minimum, and maximum effective sample sizes for item parameters in selected conditions in Study 1.

Note. The values were estimates from 30 replications. DINA = deterministic inputs noisy “and” gate; l = test length; N = sample size; μ = mean quality of items; ρ = the strength of WICD; WICD = within-item characteristic dependency.

Simulation Study 3

Design and data generation. In Study 3, two independent variables were manipulated including (a) the number of potential item blocks (B) at two levels of 2 and 3 and (b) the sample sizes (N) at two levels of 200 and 1,000. Let subscript b ($b = 1, \dots, B$) denote the b th item block. Considering the operability, items were assumed to be grouped by content domains containing different sets of attributes. There are six attributes in total measured by 30 items. The corresponding Q-matrices are presented in Figure 6. In Q1, there are two item blocks, and three attributes ($K_b = 3$) were assessed by 15 items ($I_b = 15$) in each item block. In Q2, there are three item blocks and two attributes ($K_b = 2$) were assessed by 10 items ($I_b = 10$) in each item block. For both Q-matrices, item parameters in each item block were separately generated from,

$$(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) = \left(\begin{pmatrix} \mu_{\beta,b} \\ \mu_{\delta,b} \end{pmatrix}, \begin{pmatrix} \sigma_{\beta,b}^2 & \rho_{\beta\delta,b} \sigma_{\beta,b} \sigma_{\delta,b} \\ \rho_{\beta\delta,b} \sigma_{\beta,b} \sigma_{\delta,b} & \sigma_{\delta,b}^2 \end{pmatrix} \right) = \begin{cases} \text{for } b=1 & \left(\begin{pmatrix} -2.197 \\ -1.386 \end{pmatrix}, \begin{pmatrix} 1.00 & \\ -0.45 & 0.25 \end{pmatrix} \right) \\ \text{for } b=2 & \left(\begin{pmatrix} -1.734 \\ -1.386 \end{pmatrix}, \begin{pmatrix} 0.25 & \\ -0.16 & 0.16 \end{pmatrix} \right), \\ \text{for } b=3 & \left(\begin{pmatrix} -1.386 \\ -2.197 \end{pmatrix}, \begin{pmatrix} 0.25 & \\ -0.30 & 1.00 \end{pmatrix} \right) \end{cases}$$

where $\rho_{\beta\delta,b}$ is -0.9 , -0.8 , and -0.6 in three item blocks, respectively. The true attribute profile of each person was randomly chosen from all possible patterns ($C = 64$) with equal probability. Note that a common $\boldsymbol{\mu}$ and a common $\boldsymbol{\Sigma}$ in Equation 6 were still employed in the analysis.

Results. Note that the results of the conditions with different Q-matrices should be viewed separately because they have different structures and characteristics. Figures 7 and 8 present the bias

Figure 6. K-by-I Q matrices for simulation Study 3.
Note. Blank means “0,” gray means “1”; K = the number of attributes; I = test length; Q = Q-matrix; B = the number of item blocks; b = bth item block.

Table 2. Pattern Correct Classification Rate in Study 2.

I	N	W-DINA	U-DINA
15	200	0.138	0.146
	1,000	0.165	0.169
30	200	0.272	0.277
	1,000	0.300	0.301

Note. I = test length; N = sample size; W = W-DINA; U = U-DINA; DINA = deterministic inputs noisy “and” gate.

and RMSE of the estimated item parameters for the W-DINA across all conditions. Virtually, all conditions had unbiased estimates, except a few items in $N = 200$ conditions had small bias less than ± 0.03 . Although the RMSE for item parameter estimates increased as the sample size decreased, the largest RMSE was smaller than 0.06. The PCCRs were around 0.63 for Q1 (i.e., $B = 2$) and 0.75 for Q2 (i.e., $B = 3$), respectively. Higher PCCRs in Q2 than those in Q1 could be attributed to the fact that the Q2 has a simpler structure than Q1. For instance, 24 out of the 30 items require single attribute in Q2, whereas 12 out of the 30 items require single attribute in Q1. In addition, the W-DINA worked slightly better than the U-DINA with potential item blocks (see details in Tables S2 and S3 in online supplements). Overall, item blocks have little effect on parameter estimation of the W-DINA.

An Empirical Example

Data Description and Analysis

To further demonstrate the application of the proposed modeling approach, a fraction subtraction data from de la Torre (2009) were analyzed. The data set contained a total of 536 people responding to 15 items measuring five required attributes. The total number of possible attribute profiles is 32. The Q-matrix can be found in de la Torre (2009). The W-DINA and the U-DINA were fitted to this data set. The analysis was specified in the same way as the previous simulation study. The deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002) was computed for each model to evaluate the model data fit.

Figure 7. Recovery of the guessing parameters across all conditions in Study 3.
Note. The values were estimated from 30 replications. Q = Q-matrix; B = the number of item blocks; N = sample size; RMSE = root mean square error.

Figure 8. Recovery of the slipping parameters across all conditions in Study 3.
Note. The values were estimated from 30 replications. Q = Q-matrix; B = the number of item blocks; N = sample size; RMSE = root mean square error.

Results

Figure 9 shows the plots of the Brooks–Gelman MPSRF. Compared to the MPSRF for the U-DINA, the MPSRF for the W-DINA assessed the convergence of two additional elements, μ and Σ . Despite tiny differences, both models yielded MPSRF < 1.1 around 375 iterations. Furthermore, the U-DINA and the W-DINA required 23.626 min and 18.487 min, respectively, to converge, and the ratio was approximately 1.278 indicating the W-DINA is more efficient. The DIC was 6,263.638 for the W-DINA and 6,361.466 for the U-DINA, respectively, which identified the W-DINA as the better fitting model.

Table 3 presents the estimated item parameters. According to the IDI_i , the quality of these 15 items was high. Most of the estimates of the U-DINA were larger than those of the W-DINA, which was consistent with the patterns of the bias found in Simulation Study 1. In addition, all the standard errors of the U-DINA were equal or larger than those of the W-DINA, which was consistent with the patterns of RMSE found in Simulation Study 1. Table 4 presents the estimated μ and Σ for the W-DINA. The estimated correlation between the β_s and δ_s was -0.495 . According to the estimated values, the mean guessing and mean slipping probabilities of the domain of items were approximately 0.071 and 0.120, respectively. An empirical correlation between the estimated guessing and slipping parameters is $-.546$ for the W-DINA and $-.415$ for the U-DINA. This indicates that the independence of prior distributions for different item parameters assumed by the U-DINA would underestimate the association between the guessing and slipping parameters.

Figure 9. Plots of Brooks–Gelman MPSRF in the fraction subtraction data.

Note. The plots are based on chain lengths from 50 to 750 in increments of 50. MPSRF = multivariate potential scale reduction factor; W = W-DINA; DINA = deterministic inputs noisy “and” gate.

Table 3. Estimated Item Parameters for the Fraction Subtraction Data.

Item	W-DINA					U-DINA				
	$\hat{g}$	$\hat{s}$	$SD(\hat{g})$	$SD(\hat{s})$	IDI_i	$\hat{g}$	$\hat{s}$	$SD(\hat{g})$	$SD(\hat{s})$	IDI_i
1	0.016	0.278	0.011	0.023	0.706	0.011	0.277	0.011	0.024	0.712
2	0.206	0.113	0.024	0.019	0.681	0.213	0.121	0.025	0.021	0.666
3	0.124	0.045	0.049	0.010	0.831	0.146	0.039	0.055	0.010	0.815
4	0.122	0.125	0.018	0.026	0.753	0.127	0.137	0.019	0.029	0.736
5	0.135	0.238	0.059	0.022	0.627	0.215	0.249	0.062	0.022	0.535
6	0.034	0.220	0.011	0.026	0.745	0.036	0.229	0.012	0.027	0.735
7	0.075	0.081	0.016	0.017	0.844	0.076	0.080	0.016	0.018	0.844
8	0.151	0.054	0.041	0.013	0.794	0.172	0.051	0.045	0.014	0.776
9	0.091	0.068	0.033	0.014	0.841	0.109	0.061	0.037	0.014	0.830
10	0.168	0.073	0.022	0.019	0.759	0.170	0.075	0.023	0.021	0.755
11	0.101	0.104	0.031	0.016	0.795	0.123	0.102	0.035	0.017	0.775
12	0.035	0.135	0.012	0.021	0.830	0.034	0.137	0.013	0.023	0.830
13	0.130	0.151	0.020	0.022	0.718	0.137	0.161	0.021	0.024	0.702
14	0.026	0.193	0.009	0.031	0.781	0.025	0.203	0.010	0.033	0.772
15	0.017	0.182	0.008	0.024	0.800	0.015	0.185	0.008	0.025	0.800

Note. DINA = deterministic inputs noisy “and” gate; IDI = item discrimination index.

Conclusion and Discussion

To explicitly model WICD, this study proposed a modified Bayesian DINA modeling approach, in which a bivariate normal distribution was employed as a joint prior distribution for item parameters. Simulation results indicated that the model parameters for the W-DINA were well

Table 4. Estimated Item Mean Vector and Variance and Covariance Matrix for the Fraction Subtraction Data.

Parameter		Posterior <i>M</i>	<i>SD</i>	Correlation
Σ	Σ_{11}	1.233	0.645	1.000
	Σ_{12}	−0.415	0.301	−.495
	Σ_{22}	0.571	0.260	1.000
μ	μ_{β}	−2.564	0.305	—
	μ_{δ}	−1.995	0.201	—

recovered and that explicit modeling of WICD can further improve model parameter estimation. The W-DINA outperforms the regular U-DINA (i.e., more accurate, precise, and efficient model parameter estimates, especially for small sample size conditions) in most normal testing conditions. Even under boundary conditions, the W-DINA can provide a comparable recovery of model parameters to the U-DINA. In addition, the proposed modeling approach with a common bivariate normal distribution also worked well when potential item blocks presented. Furthermore, the fraction subtraction data were analyzed to illustrate the application of the proposed modeling approach. Overall, given the results of the simulation and empirical studies, the advantage of explicitly modeling WICD was evident. Although the improvement in parameter estimation accuracy and precision decreased as the sample size increased, the proposed modeling approach still improved the efficiency in parameter estimation.

In general, the use of a multivariate normal distribution between item parameters serves as an augmentation approach that the estimation of one item parameter could borrow information from the other item parameter. Such an augmentation approach for item parameter estimation is quite similar to some proposed augmentation approaches for person parameter estimation (e.g., de la Torre, 2008b; Wang, Chen, & Cheng, 2004). In addition, in many simulation studies related to the DINA model, the guessing and slipping parameters were generated independently, which may not be the case in reality. The item parameter generation approach in this study could be a possible approach to generate correlated within-item parameters in future simulation studies. Furthermore, according to the results, the advantage of using the proposed model was more evident in item parameters when the sample size is small. The practical gains of using the proposed model would be reflected in some applications where items are precalibrated with sufficient precision and low item exposure rate, for example, the cognitive diagnostic computerized adaptive testing (Cheng, 2009).

The work represented in this article is an initial attempt to better understand the WICD in CDMs. Despite promising results, some limitations still exist. Although only the DINA model was used in this study, the proposed modeling method can be easily extended to the Deterministic Input Noisy Output “OR” (DINO) model (Templin & Henson, 2006), because the DINA and the DINO models share a “dual” relation (Köhn & Chiu, 2016). However, unlike the DINA and the DINO models that contain $2I$ item parameters, different items in more general CDMs (e.g., de la Torre, 2011; Henson et al., 2009; von Davier, 2008) can have different numbers of parameters. Thus, employing a common multivariate normal distribution for all items does not seem feasible. How to consider the WICD in generalized CDMs is an interesting topic. In addition, for potential item blocks existed, it is valuable to include multiple multivariate normal distributions in further exploration. In such cases, the idea of item-family (Glas, van der Linden, & Geerlings, 2016) can be imposed. Currently, the W-DINA only focuses on dichotomous items, but it is meaningful and practical to extend it to polytomous response items

(e.g., Ma & de la Torre, 2016; von Davier, 2008). Although the WICD issue was explored in the full Bayesian estimation framework in this study, the WICD issue also exists in current non-Bayesian algorithms for the U-DINA (e.g., the MMLE/EM (Marginal Maximum Likelihood Estimation/ expectation maximization) approach, see Figures S5 and S6 and Tables S4 and S5 in online supplements), which need to be further studied. Furthermore, in the empirical example, the Q-matrix does not satisfy identifiability conditions (Xu & Zhang, 2016), which may yield biased diagnostic classifications. Thus, further analyses are needed based on a modified Q-matrix (Chen, Culpepper, Chen, & Douglas, 2018; de la Torre & Chiu, 2015). Finally, it should be noted that although a multivariate normal distribution for item parameters is assumed to account for WICD in this study, caution should be exercised when claiming universal validity for such assumption.

Acknowledgments

The authors thank Dr. Hua-Hua Chang, Dr. Jimmy de la Torre, and two anonymous reviewers for their constructive comments on earlier drafts of this article. In addition, the first author would also like to express his heartfelt thanks to Dr. Wen-Chung Wang for his selfless guidance in the past few years.

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 study was partly supported by Natural Science Foundation of Zhejiang Province, China (Grant No. LY16C090001).

Supplemental Material

Supplemental material is available for this article online.

Notes

1. Note that, in all studies in this article, the convergence was examined via the (univariate) potential scale reduction factor (PSRF). However, the PSRF focuses on only one variable (e.g., a guessing parameter of Item 1), thus, due to the space limitation, it is impossible for us to present all PSRFs. Instead, the multivariate potential scale reduction factor (MPSRF; Brooks & Gelman, 1998), which can be taken as a summary of PSRFs for multiple parameters, was computed for each replication to evaluate the number of iterations needed to reach convergence (Figure 4).
2. All calculations reported in this article were conducted on an msi GT72VR 6RD DOMINATOR laptop with a 2.6GHz Intel Core i7 6700HQ CPU, 2133MHz 32 GB of memory, and 256 GB SanDisk z400s Solid State Drive.

ORCID iD

Peida Zhan <https://orcid.org/0000-0002-6890-7691>

References

- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434-455.

- Chen, Y., Culpepper, S. A., Chen, Y., & Douglas, J. (2018). Bayesian estimation of the DINA Q-matrix. *Psychometrika*, 83, 89-108.
- Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika*, 74, 619-632.
- Culpepper, S. A. (2015). Bayesian estimation of the DINA model with Gibbs Sampling. *Journal of Educational and Behavioral Statistics*, 40, 454-476.
- 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. (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). Multidimensional scoring of abilities: The ordered polytomous response case. *Applied Psychological Measurement*, 32, 355-370.
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34, 115-130.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199.
- de la Torre, J., & Chiu, C.-Y. (2015). A general method of empirical Q-matrix validation. *Psychometrika*, 81, 253-273.
- Fox, J.-P. (2010). *Bayesian item response modeling: Theory and applications*. New York, NY: Springer.
- Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.
- Glas, C. A. W., & van der Linden, W. J. (2003). Computerized adaptive testing with item cloning. *Applied Psychological Measurement*, 27, 247-261.
- Glas, C. A. W., van der Linden, W. J., & Geerlings, H. (2016). Item-family models. In W. J. van der Linden (Ed.), *Handbook of item response theory, Volume one: Models*. CRC press 437-447. Boca Raton, FL: Chapman and Hall/CRC.
- Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74, 191-210.
- 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.
- Kass, R. E., Carlin, B. P., Gelman, A., & Neal, R. M. (1998). Markov Chain Monte Carlo in practice: A roundtable discussion. *The American Statistician*, 52, 93-100.
- Köhn, H.-F., & Chiu, C.-Y. (2016). A proof of the duality of the DINA model and the DINO model. *Journal of Classification*, 33, 171-184.
- Lee, Y.-S., de la Torre, J., & Park, Y. S. (2012). Cognitive diagnosticity of IRT-constructed assessment: An empirical investigation. *Asia Pacific Education Review*, 13, 333-345.
- Levy, R., & Mislevy, R. J. (2016). *Bayesian psychometric modeling*. CRC Press
- Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology*, 69, 253-275.
- Macready, G. B., & Dayton, C. M. (1992). The application of latent class models in adaptive testing. *Psychometrika*, 57, 71-88.
- Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*, 51, 177-195.
- Plummer, M. (2015). JAGS Version 4.0.0 user manual. Lyon, France. Retrieved from <https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/>
- Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York, NY: Guilford Press.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64, 583-639.
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345-354.
- Templin, J., & Henson, R. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11, 287-305.
- Templin, J., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus. *Educational Measurement: Issues and Practice*, 32, 37-50.

- van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, 72, 287-308.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61, 287-307.
- von Davier, M. (2014). The DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology*, 67, 49-71.
- Wang, W.-C., Chen, P.-H., & Cheng, Y.-Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods*, 9, 116-136.
- Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, 81, 625-649.
- Yi, Q., & Chang, H. H. (2003). a-Stratified CAT design with content blocking. *British Journal of Mathematical and Statistical Psychology*, 56, 359-378.