Article

Journal of Educational and Behavioral Statistics
2018, Vol. 43, No. 1, pp. 57–87
DOI: 10.3102/1076998617719727

Tracking Skill Acquisition With Cognitive Diagnosis Models: A Higher-Order, Hidden Markov Model With Covariates

Shiyu Wang
University of Georgia

Yan Yang
Steven Andrew Culpepper
Jeffrey A. Douglas
University of Illinois at Urbana-Champaign

A family of learning models that integrates a cognitive diagnostic model and a higher-order, hidden Markov model in one framework is proposed. This new framework includes covariates to model skill transition in the learning environment. A Bayesian formulation is adopted to estimate parameters from a learning model. The developed methods are applied to a computer-based assessment with a learning intervention. The results show the potential application of the proposed model to track the change of students' skills directly and provide immediate remediation as well as to evaluate the efficacy of different interventions by investigating how different types of learning interventions impact the transitions from nonmastery to mastery.

Keywords: *cognitive diagnostic models; higher order; hidden Markov model; longitudinal; skill change; Markov chain Monte Carlo; spatial cognition*

Introduction

A fundamental goal of educational research is to assess students' mastery of skills from their performance on test items. Cognitive diagnostic modeling is a useful tool to generate a fine breakdown of the skills or attributes possessed by students, for the purpose of more complete feedback, which can hopefully enhance students' learning. Research continues to document the benefits of cognitive diagnosis models (CDMs) as a framework for classifying students into educationally relevant skill profiles. In fact, CDMs have been used to study spatial skills (Culpepper, 2015), English-language proficiency (Chiu & Köhn, 2015; J. L. Templin & Hoffman, 2013), fraction subtraction (de la Torre & Douglas, 2004), pathological gambling (J. L. Templin & Henson, 2006), and skills found in large-scale testing programs (H. Li, Hunter, & Lei, 2015; Ravand, 2015). Moreover, recent developments of methods for estimating the CDM Q-matrix have further broadened the application of CDMs to social anxiety disorders (Chen, Liu, Xu, & Ying, 2015).

In the aforementioned applications, CDMs provided a useful psychometric and measurement framework for the classification of skills or attributes at a given point in time. However, in the context of education, students usually learn and master content over time. In the case where students learn, the application of traditional CDM methods to this problem would consist of sequentially applying a CDM model to classify students at several time points. One limitation of this strategy is that traditional CDMs are not designed to model the learning process. That is, student learning in particular content domains may follow common patterns or trajectories that traditional methods would not directly uncover. In fact, educational researchers and practitioners may be interested in theoretical questions concerning the pathways students take toward mastery or the effect of interventions on learning. Theoretical questions about the learning process are a fundamentally different set of questions than understanding mastery at a given point in time. Instead, questions about learning trajectories necessarily require a longitudinal perspective on attribute mastery, and there is a corresponding methodological opportunity to extend the CDM framework to model the skill acquisition process.

This article focuses on the use of CDMs to model students' learning trajectories in an effort to provide researchers information about the learning process. In fact, research has only recently considered the role of CDMs to track learning and skill acquisition in a longitudinal fashion (Kaya & Leite, 2017; F. Li, Cohen, Bottge, & Templin, 2016; Studer, 2012; Ye, Fellouris, Culpepper, & Douglas, 2016). For instance, Studer (2012) proposed two methods using a CDM framework, that is, the parameter-driven process for change method and knowledge tracing (KT; Corbett & Anderson, 1994) with cognitive diagnostic modeling. The former models learning indirectly by tracking student membership in latent states that drive the distributions of the student parameter in the static portion of the model, whereas the latter directly tracks students' attribute changes. F. Li, Cohen, Bottge, and Templin (2016) used the deterministic input, noisy, "and" gate (DINA; Junker & Sijtsma, 2001) model as a measurement model in the latent transition model framework and applied the proposed method in the analysis of effects of an instructional intervention. Kaya and Leite (2017) evaluated the combination of a latent transition model and the DINA and deterministic input, noisy, "or" gate model (J. L. Templin & Henson, 2006) in simulation studies. Furthermore, Ye, Fellouris, Culpepper, and Douglas (2016) showed how the CDM framework can be used to adaptively assess learning in an online learning environment.

Prior research has also considered learning from other methodological frameworks. In the educational data mining field, KT (Corbett & Anderson, 1994) is a popular method used in intelligent tutoring systems (ITSs). Conventional KT requires that each item must require exactly one skill and uses conditional probability tables for guess and slip parameters and learning probabilities. Moreover, it assumes that students and items do not vary in abilities or difficulties, so that any two items involving the same skill are assumed to be equivalent (Khajah, Huang, González-Brenes, Mozer, & Brusilovsky, 2014). Many different models have been developed in the KT family, which have extended this method to deal with multiple skills and different item difficulties (e.g., González-Brenes, Huang, & Brusilovsky, 2014; González-Brenes & Mostow, 2013; Pardos & Heffernan, 2010; Y. Xu & Mostow, 2012). In the psychometric field, analysis of changes in latent variables has mainly been developed in the item response theory (IRT) and latent transition analysis frameworks. For example, Andersen (1985) and Embretson (1991) proposed a multidimensional IRT model to measure learning in terms of the increment in unidimensional latent abilities over repeated measures, and Chung, Walls, and Park (2007) used logistic regression in the latent transition model framework to measure the change of latent classes over time.

This article offers at least three contributions to the literature using CDMs to track student learning. The first is to propose a family of learning models that integrates the CDM framework with a higher-order, hidden Markov model (HO-HMM) for attribute transitions (e.g., see Chen, Culpepper, Wang, & Douglas, in press). Second, prior research by Kaya and Leite (2017) recommended developing modeling frameworks that allow for the inclusion of student covariates to explain individual differences in transitions among classes. We use the HO-HMM to model transitions from nonmastery to mastery of an attribute as a function of student covariates. In the context of computer-based assessments, student covariates may include cumulative practice, intervention indicators, and student characteristics. Third, we evaluate the developed HO-HMM through a computerized assessment of spatial rotation ability with embedded learning interventions. The experimental application demonstrates another target of the study, which is to apply the developed methods in a computer-based assessment with learning intervention. The potential applications of the proposed method are to provide individualized learning trajectory for each student, which can be used to design the personalized learning strategy in the future once the model parameters have been trained from the data. In addition, they can be used to evaluate the efficacy of different interventions by investigating how different types of learning interventions impact the transitions from nonmastery to mastery. Furthermore, the estimated learning model parameters can be used as priors in an ITS to predict future students' performances on the same instrument.

The rest of the article is organized in four sections. We first review the common cognitive diagnostic models. Next, a family of learning models is proposed and a specific learning model example and the model estimation procedures are presented. This is followed by the presentation of an assessment with embedded learning interventions that is developed for evaluating spatial reasoning skills and the corresponding experimental design. A simulation study based on this experiment is presented in the same section to demonstrate the accuracy and convergence of the proposed model estimation algorithm. The proposed methods are applied to the data collected from this experiment. Finally, we discuss the results and provide concluding remarks.

Review of Cognitive Diagnostic Models

Latent class models for cognitive diagnosis are generally restricted to reflect some assumptions about the skills by which examinees respond to items. Based on such assumptions, CDMs can be categorized as noncompensatory or compensatory models. Conjunctive models, one type of noncompensatory model, express the notion that all attributes specified in the attribute-by-item Q-matrix (Tatsuoka, 1985) for an item should be required to answer the item correctly but allow for slips and guesses in ways that distinguish the models from one another. Some common conjunctive models include the DINA (Junker & Sijtsma, 2001), noisy input, deterministic, "and" gate model (Maris, 1999), and the reduced reparameterized unified model (Hartz & Roussos, 2008). Unlike noncompensatory models, compensatory models allow an individual to compensate for what is lacked in some measured skills by having mastered other skills. Encompassing the traditional categories of the reduced CDMs, several general models based on different link functions (e.g., Davier, 2008; de la Torre, 2011; Henson, Templin, & Willse, 2009) have been developed to include many of the common compensatory and noncompensatory models.

Different types of CDMs have different forms of item response functions, which model the probability of a correct response given the latent attribute profile. For the purposes of illustration the DINA model, item response function is described. Suppose that there are K attributes associated with J items. Entry q_{jk} in the $J \times K$ matrix $\mathbf{Q}$ indicates whether Item j requires the kth attribute. Let X_{ij} denote the ith $(i = 1, \ldots, N)$ subject's responses to Item j $(j = 1, \ldots, J)$. Let $\boldsymbol{\alpha}_i$ denote this subject's latent attribute pattern, where $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, ..., \alpha_{iK})'$ and each α_{ik} takes a value of either 0 or 1 for $k = 1, 2, \ldots K$. Specifically, α_{ik} is an indicator of whether the ith subject possesses the kth attribute. The ideal response pattern obtained under $\boldsymbol{\alpha}_i$ is denoted as $\boldsymbol{\eta}_i = (\eta_{i1}, \eta_{i2}, ..., \eta_{iJ})'$, where $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$ to indicate whether subject i has mastered all the attributes required by Item j. The DINA model allows for deviations from this pattern according to item parameters for slipping, $s_j = P(X_{ij} = 0 | \eta_{ij} = 1)$, and guessing, $g_j = P(X_{ij} = 1 | \eta_{ij} = 0)$. The item response function of the DINA model is then

$$P(X_{ij} = 1 | \boldsymbol{\alpha}_i, \mathbf{s}, \mathbf{g}) = (1 - s_j)^{\eta_{ij}} g_j^{(1-\eta_{ij})}. \tag{1}$$

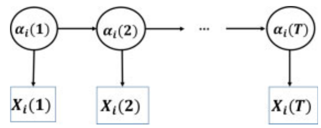

FIGURE 1. *The path diagram for a learning model.*

Models for Learning

As reviewed in the Introduction, the main application of CDMs is on the classification of skills or attributes at a given point in time. We consider a different situation where students may receive some type of feedback between time t and $t+1$, which could promote learning in students' skills and attributes. The path diagram in Figure 1 illustrates such a process. The notation $\boldsymbol{\alpha}_i = \left(\boldsymbol{\alpha}_i(1), \boldsymbol{\alpha}_i(2), ..., \boldsymbol{\alpha}_i(T)\right)'$ in Figure 1 denotes the ith student's attribute profile trajectory, and $\boldsymbol{\alpha}_i(t) = \left(\alpha_{i1}(t), ..., \alpha_{iK}(t)\right)'$ represents the corresponding attribute profile at time t, where $t = 1, 2, \ldots, T$ is used to index attributes over time and $k = 1, 2, \ldots, K$ indicates the kth attribute. Note that $\boldsymbol{\alpha}_i(1)$ is the initial attribute state that reflects student i's performance at $t = 1$. Additionally, let $j = 1, \ldots, J(t)$ index the number of items administered at time t where the total assessment consists of $J = \sum_{t=1}^{T} J(t)$ items. Let $\mathbf{Q}(t)$ be a $J(t) \times K$ matrix with zeros and ones that denote the skills needed to correctly answer the J_t items at time t. Define the elements of $\mathbf{Q}(t)$ as q_{jst}, where $s = 1, 2, \ldots, K$ indexes skills. Let $\mathbf{X}_i(t) = \left(X_{i1}(t), X_{i2}(t), ..., X_{iJ(t)}(t)\right)$ denote the ith student's response vector to the $J(t)$ questions at time t and $X_{ij}(t)$ is his or her response on Item j at time t.

A learning model is proposed to trace the subject's latent attribute profile $\boldsymbol{\alpha}_i(t)$ dynamically as illustrated by Figure 1. It contains two components: One is a measurement model that connects the observed response data $\mathbf{X}_i(t)$ with attribute profile $\boldsymbol{\alpha}_i(t)$ through $P\left(\mathbf{X}_i(t)|\boldsymbol{\alpha}_i(t), \boldsymbol{\beta}_t\right)$; the other is a transition model $P\left(\boldsymbol{\alpha}_i(t+1)|\boldsymbol{\alpha}_i(t)\right)$. Note $\boldsymbol{\beta}_t$ denotes parameters of the items examinee i received at time t. In the current framework, the item parameters in the measurement model are assumed to be time invariant, that is, their values do not change with time. The subscript of $\boldsymbol{\beta}_t$ only indicates that the items examinee i receives may depend on time t not the value of the item parameters. The measurement model can be chosen as one of the common CDMs reviewed in the previous section, and the DINA model is used as an example. The main focus is on modeling the transition kernel of attribute profiles. Several assumptions are imposed to simplify the transition model structure. First, we assume that a student will never lose an attribute once he or she has mastered it. That is, for some time $t \geq 1$, $\forall k$,

$$P\left(\alpha_{ik}(t') = 1|\alpha_{ik}(t) = 1\right) = 1, \forall t' > t \quad \text{and} \quad P\left(\alpha_{ik}(t') = 0|\alpha_{ik}(t) = 1\right) = 0, \forall t' > t. \quad (2)$$

The assumption implied by Equation 2 can help reduce the number of parameters to be estimated in the transition model and is reasonable for training skills in a short time period (e.g., see Corbett & Anderson, 1994; Chen et al., in press). With this assumption, the question now is how to model $P\left(\alpha_{ik}(t+1) = 1|\alpha_{ik}(t) = 0\right)$, the transition probability of nonmastery on one attribute to mastery during the process of training. Second, we assume that the transition probability depends on n independent variables $\mathbf{Y}_i(t) = \left(Y_{i1}(t), \ldots, Y_{in}(t)\right)$, which can be either latent or observed. The t in the parentheses indicates the values of the covariates might change with time t. The general transition probability can be denoted as

$$P\left(\alpha_{ik}(t + 1) = 1|\alpha_{ik}(t) = 0\right) = G\left(Y_{i1}(t), Y_{i2}(t), \ldots, Y_{in}(t)\right), \quad (3)$$

where G is a link function. When we use a logistic link function, Equation 3 can be written as

$$\text{logit}\left[P\left(\alpha_{ik}(t+1) = 1|\alpha_{ik}(t) = 0\right)\right] = \lambda_{0k} + \sum_{m=1}^{n} \lambda_{mk} Y_{im}(t). \quad (4)$$

Finally, an additional assumption is imposed to the transition model. We assume that at time $t+1$, the transition of the components of $\boldsymbol{\alpha}_i(t+1)$ are independent given $\boldsymbol{\alpha}_i(t)$ and other covariates $\mathbf{Y}_i(t)$, that is

$$P\left(\boldsymbol{\alpha}_i(t+1)|\boldsymbol{\alpha}_i(t), \mathbf{Y}_i(t)\right) = \begin{cases} 0, & \text{if for some } k, \alpha_{ik}(t+1) = 0 \quad \text{and} \quad \alpha_{ik}(t) = 1, \\ 1, & \text{for all } k, \alpha_{ik}(t) = 1, \\ \prod_{k:\alpha_{ik}(t)=0} p\left(\alpha_{ik}(t+1)|\alpha_{ik}(t) = 0, \boldsymbol{\alpha}_i(t), \mathbf{Y}_i(t)\right), & \text{otherwise.} \end{cases} \quad (5)$$

Here, $p\left(\alpha_{ik}(t+1)|\alpha_{ik}(t) = 0, \boldsymbol{\alpha}_i(t), \mathbf{Y}_i(t)\right)$ denotes the the distribution for skill k at time $t+1$ conditioned upon skills at time t and the other covariates $Y_{i1}(t), \ldots, Y_{in}(t)$. It can be described as

$$P\left(\alpha_{ik}(t+1) = 1|\alpha_{ik}(t) = 0, \boldsymbol{\alpha}_i(t), \mathbf{Y}_i(t)\right)^{\alpha_{ik}(t+1)}$$
$$P\left(\alpha_{ik}(t+1) = 0|\alpha_{ik}(t) = 0, \boldsymbol{\alpha}_i(t), \mathrm{Y}_i(t)\right)^{1-\alpha_{ik}(t+1)}. \quad (6)$$

A Learning Model Example

In this subsection, we provide a simple learning model that integrates the DINA model as the measurement model and utilizes several latent and observed

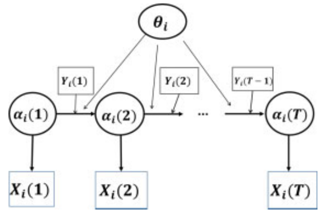

FIGURE 2. *A learning model with the higher-order, hidden Markov model.*

covariates to model the transition probability. A natural assumption regarding the transition probability is that it should depend on a general learning ability, the attributes the student has already mastered, and also the amount of practice he or she has received on that attribute. The path diagram for such a learning model with a HO-HMM can be illustrated by Figure 2.

A transition model incorporating the proposed covariates can be written as

$$\text{logit}\left[P\left(\alpha_{ik}(t+1)=1|\alpha_{ik}(t)=0,\boldsymbol{\alpha}_i(t)\right)\right]=\lambda_{0k}+\lambda_{1k}\theta_i+\lambda_{2k}\sum_{l\neq k}\alpha_{il}(t)$$
$$+\lambda_{3k}\sum_{m=1}^{t}\sum_{j=1}^{J_t}q_{jkm}c_{jk}. \tag{7}$$

In Equation 7, θ_i denotes the general learning ability for subject i, which is time invariant. $\sum_{l\neq k}\alpha_{il}(t)$ represents the number of mastered skills at time t and $\sum_{m=1}^{t}\sum_{j=1}^{J_t}q_{jkm}c_{jk}$ is a covariate that indicates the "practice" this subject i has received on skill k by time t, where q_{jkm} is the (j,k) element of the $\mathbf{Q}(m)$ matrix at time m and c_{jk} is the prespecified benefit a student can gain on learning skill k if he or she practices Item j. c_{jk} is a positive number that may varies across items and skills to quantify the degree to which Item j helps in the learning of skill k. When $c_{jk=1}$, the $\sum_{m=1}^{t}\sum_{j=1}^{J_t}q_{jkm}c_{jk}$ directly reflects the number of questions related to attribute k until time t.

In Equation 7, $\lambda_{0k}\in(-\infty,\infty)$ is the intercept, which is the log odds of becoming mastery of skill k, given the learning ability $\theta_i=0$, nonmastery of any other attributes, and zero practice/training on this skill. We constrain λ_{1k} to be positive to make the sign of θ_i identifiable as $\theta_i\in(-\infty,\infty)$. This also reflects the assumption that the higher learning ability, the greater the chance of mastering attributes in the transition. The contribution of the other learned skills on the mastery of skill k is captured by $\lambda_{2k}\in(-\infty,\infty)$. A positive λ_{2k} means that the skills mastered at time t will accelerate the mastery of the unlearned skills at time $t+1$. The learning rate for skill k is $\lambda_{3k}\in(-\infty,\infty)$, and a positive coefficient indicates that the more practice the student has received on the nonmastered attribute k, the higher probability he or she will master this attribute at time

$t+1$. The transition model coefficient vector $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3)$, where $\boldsymbol{\lambda}_m = (\lambda_{mk})_k^K$, $m = 0, 1, 2, 3$, depends on k, indicating that the impact of the covariates on mastery could vary across different attributes. However, a large sample size might be needed to accurately estimate each λ_{mk}. A parsimonious model can be reduced from this general form by constraining the coefficients to be the same for different k, that is, only $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$ needs to be estimated.

Parameter Estimation

Expectation maximization algorithm is the traditional way to estimate hidden Markov models (Cappé, Moulines, & Rydén, 2005). However, due to the modeling of the transition kernel within the hidden Markov model, it is hard to get the closed form of the maxima of all model parameters in our case. Thus, a fully Bayesian formulation is adopted to estimate parameters of the learning model. In order to illustrate this process, the joint likelihood function for the learning model is presented. The Q-matrix is assumed to be known in our current study. Let $\boldsymbol{\theta} = (\theta_i)_{i=1}^N$ denote the general learning ability vector for N students, which is the only unknown latent covariate in the learning model. The coefficients of the n covariates in the learning model are denoted as $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_0, \ldots, \boldsymbol{\lambda}_n)$. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T)$ be the row vectors of slipping and guessing parameters, where $\boldsymbol{\beta}_t = (\mathbf{s}_t, \mathbf{g}_t)$. Let I_t denote the set of the identity of items at time t, then $\mathbf{s}_t = (s_j)_{\{j \in I_t\}}$ and $\mathbf{g}_t = (g_j)_{\{j \in I_t\}}$. Furthermore, let $\pi_c = P\left(\alpha_c(1)\right)$ be the marginal probability of belonging to class c in the population at the initial time point and define $\boldsymbol{\pi} = (\pi_1, \ldots., \pi_C)'$ as the $C = 2^K$ dimensional vector of class membership probabilities. Let $p(\cdot)$ denote the priors for $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and $\boldsymbol{\pi}$. Based on the assumption that the ith subject's responses at one time point are independent of the previous given the latent attribute profile at that time, the full likelihood function in the Baysian framework can be written as

$$\begin{aligned} L(\mathbf{X}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\beta}) &= p(\boldsymbol{\lambda})p(\boldsymbol{\beta})\prod_{i=1}^{N}\sum_{c=1}^{C}\pi_c L\left(\mathbf{X}_i(1)|\boldsymbol{\alpha}_i(1) = \boldsymbol{\alpha}_c, \boldsymbol{\beta}_1\right)p(\boldsymbol{\pi})p(\theta_i) \\ &\times \left(\prod_{t=1}^{T-1} P\left(\boldsymbol{\alpha}_i(t+1)|\theta_i, \boldsymbol{\alpha}_i(t), \boldsymbol{\lambda}\right) \times L\left(\mathbf{X}_i(t+1)|\boldsymbol{\alpha}_i(t+1), \boldsymbol{\beta}_{t+1}\right)\right), \end{aligned}$$

where $P\left(\boldsymbol{\alpha}_i(t+1)|\boldsymbol{\alpha}_i(t), \theta_i, \boldsymbol{\lambda}\right)$ is constructed based on the proposed learning model (Equation 4) and $L\left(\mathbf{X}_i(t)|\boldsymbol{\alpha}_i(t), \boldsymbol{\beta}\right)$ is the likelihood function constructed by the DINA model (Equation 1). The joint posterior distribution of $\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\beta}$ given $\mathbf{X}$ is

$$P(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\beta}|\mathrm{X}) \propto L(\mathbf{X}|\boldsymbol{\alpha}, \boldsymbol{\beta})P(\boldsymbol{\alpha}|\boldsymbol{\theta}, \boldsymbol{\lambda})P(\boldsymbol{\theta})P(\boldsymbol{\lambda})P(\boldsymbol{\beta}). \tag{8}$$

To outline the specific form for the posterior distribution, let $P_{ijt} = P\left(X_{ij}(t) \ = \ 1|\boldsymbol{\alpha}_i(t), s_j, g_j\right)$, then

$$L(\mathbf{X}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{N}\prod_{t=1}^{T} L\left(\mathbf{X}_i(t)|\boldsymbol{\alpha}_i(t), \boldsymbol{\beta}_t\right) = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J_t} P_{ijt}^{X_{ijt}}(1 - P_{ijt})^{1-X_{ijt}}.$$

The full conditional distribution of the parameters given the data and the rest of the parameters are discussed next. As we cannot get the closed form for the full conditional distributions, sampling is carried out using Markov chain Monte Carlo (MCMC). First, the structure parameter $\boldsymbol{\pi}$, person parameters $\boldsymbol{\alpha}_i(t)$, and the DINA model parameters can be updated using Gibbs sampling. To get the full conditional distribution for $\boldsymbol{\alpha}$ and $\boldsymbol{\pi}$, we assume the prior for $\boldsymbol{\pi}$ is

$$p(\boldsymbol{\pi}) = p(\pi_1, ..., \pi_C) \propto \text{Dirichelet}\,(\boldsymbol{\delta}_0), \boldsymbol{\delta}_0 = (\delta_{01}, ..., \delta_{0C}).$$

Note that an uninformative prior can be implemented by setting elements of $\boldsymbol{\delta}_0$ as 1, while the values of the elements within $\boldsymbol{\delta}_0$ specify prior information about the prevalence of class membership in the population at the initial time point. Based on this prior, we can obtain the full conditional distribution for $\boldsymbol{\alpha}_i(t)$,

$$P\left(\boldsymbol{\alpha}_i(t)|X_i(t), \pi, \boldsymbol{\beta}_t\right) = \prod_{\boldsymbol{\alpha}_c \in V_i(t)} \tilde{\pi}_{ict}^{\mathcal{I}\left(\boldsymbol{\alpha}_i(t) = \boldsymbol{\alpha}_c\right)},$$

where $V_i(t)$ denotes the space for $\boldsymbol{\alpha}_i(t)$ at time t. Due to an assumption about the skill transition (Equation 2), we can obtain $\tilde{\pi}_{ict}$ in the following three cases:

(1) When $t = 1$,

$$\tilde{\pi}_{ic1} = \frac{\pi_c P\left(\mathbf{X}_i(1)|\boldsymbol{\alpha}_i(1) = \boldsymbol{\alpha}_c, \boldsymbol{\beta}_1\right) P\left(\boldsymbol{\alpha}_i(2)|\boldsymbol{\alpha}_i(1), \boldsymbol{\lambda}, \theta_i\right)}{\sum_{\boldsymbol{\alpha}_c \in V_i(1)} \pi_c P\left(\mathbf{X}_i(1)|\boldsymbol{\alpha}_i(1) = \boldsymbol{\alpha}_c, \boldsymbol{\beta}_1\right) P\left(\boldsymbol{\alpha}_i(2)|\boldsymbol{\alpha}_i(1), \boldsymbol{\lambda}, \theta_i\right)}, \quad (9)$$
$$V_i(1) = \{\boldsymbol{\alpha}|\alpha_k \neq 1, \text{if} \quad \alpha_{ik}(2) = 0\}$$

(2) For $1 < t < T$,

$$\tilde{\pi}_{ict} = \frac{P\left(\boldsymbol{\alpha}_i(t+1)|\boldsymbol{\alpha}_i(t), \boldsymbol{\lambda}, \theta_i\right) P\left(\boldsymbol{\alpha}_i(t)|\boldsymbol{\alpha}_i(t-1), \boldsymbol{\lambda}, \theta_i\right) P\left(\mathbf{X}_i(t)|\boldsymbol{\alpha}_i(t) = \boldsymbol{\alpha}_c, \boldsymbol{\beta}_t\right)}{\sum_{\boldsymbol{\alpha}_c \in V_i(t)} P\left(\boldsymbol{\alpha}_i(t+1)|\boldsymbol{\alpha}_i(t), \boldsymbol{\lambda}, \theta_i\right) P\left(\boldsymbol{\alpha}_i(t)|\boldsymbol{\alpha}_i(t-1), \boldsymbol{\lambda}, \theta_i\right) P\left(\mathbf{X}_i(t)|\boldsymbol{\alpha}_i(t) = \boldsymbol{\alpha}_c, \boldsymbol{\beta}_t\right)}$$
$$V_i(t) = \{\boldsymbol{\alpha}|\alpha_k \neq 1, \text{if} \quad \alpha_{ik}(t+1) = 0 \quad \text{and} \quad \alpha_k = 1, \quad \text{if} \quad \alpha_{ik}(t-1) = 1\}, \quad (10)$$

(3) When $t = T$,

$$\tilde{\pi}_{icT} = \frac{P\Big(\mathbf{X}_i(T)|\boldsymbol{\alpha}_i(T) = \boldsymbol{\alpha}_c, \boldsymbol{\beta}_t\Big)P\Big(\boldsymbol{\alpha}_i(T)|\boldsymbol{\alpha}_i(T-1), \boldsymbol{\lambda}, \theta_i\Big)}{\sum_{\boldsymbol{\alpha}_c \in V_i(T)} P\Big(\mathbf{X}_i(T)|\boldsymbol{\alpha}_i(T) = \boldsymbol{\alpha}_c, \boldsymbol{\beta}_T\Big)P\Big(\boldsymbol{\alpha}_i(T)|\boldsymbol{\alpha}_i(T-1), \boldsymbol{\lambda}, \theta_i\Big)} \quad (11)$$
$$V_i(T) = \{\boldsymbol{\alpha}|\alpha_k \neq 0, \text{if} \quad \alpha_{ik}(T-1) = 1\}.$$

With similar arguments as those in Culpepper (2015), we can get the full conditional distribution for $\boldsymbol{\pi}$ as

$$\boldsymbol{\pi}|\boldsymbol{\alpha}_1(1), ..., \boldsymbol{\alpha}_N(1) \propto \text{Dirichlet}(\tilde{\mathbf{N}} + \boldsymbol{\delta}_0), \quad (12)$$

where $\tilde{\mathbf{N}} = (\tilde{N}_1, ..., \tilde{N}_C) = \Big(\sum_{i=1}^N \mathcal{I}_{\boldsymbol{\alpha}_i(1) = \boldsymbol{\alpha}_1}, ..., \sum_{i=1}^N \mathcal{I}_{\boldsymbol{\alpha}_i(1) = \boldsymbol{\alpha}_c}\Big)$.

For the DINA model parameters $\boldsymbol{\beta} = (\boldsymbol{\beta}_1 \ldots, \boldsymbol{\beta}_T)$, we can use the same formulation as that in Culpepper (2015). Remember that I_t represents the set of item identity each student receives at time t, then the full conditional distributions for the item parameters, s_j and g_j, for $j \in I_t$, are truncated bivariate β distributions, which can be written as

$$\begin{aligned} p(s_j, g_j|\mathbf{X}, \boldsymbol{\alpha}) \propto s_j^{\tilde{a}_s - 1}(1 - s_j)^{\tilde{b}_s - 1} g_j^{\tilde{a}_g}(1 - g_j)^{\tilde{b}_g - 1}\mathcal{I}_{\{(s_j, g_j) \in \mathcal{P}\}}, \\ \tilde{a}_s = \mathcal{S}_j + a_s,\ \tilde{b}_s = \mathcal{T}_j - S_j + b_s, \\ \tilde{a}_g = \mathcal{G}_j + a_g,\ \tilde{b}_g = N - \mathcal{T}_j - \mathcal{G}_j + b_g, \\ \mathcal{T}_j = \sum_{i=1}^N \eta_{ij}, \mathcal{S}_j = \sum_{i|X_{ij(t)}=0}^N \eta_{ij}, \mathcal{G}_j = \sum_{i|X_{ij}(t)=1}^N (1 - \eta_{ij}). \end{aligned} \quad (13)$$

Here a_s, b_s, a_g, and b_g represent the parameters in the joint prior distribution for s_j and g_j, and we set each of those four parameters as 1, which is an uninformative prior that is identical to employing a linearly truncated bivariate uniform prior for s_j and g_j. $\mathcal{P}$ is the space for s_j, g_j to restrict $(1 - s_j) > g_j$.

Second, the full conditional distributions for the latent trait θ_i and the learning model coefficients $\boldsymbol{\lambda}$ are

$$P(\boldsymbol{\lambda}|\boldsymbol{\alpha}, \boldsymbol{\theta}) \propto \prod_{i=1}^N \prod_{t=1}^{T-1} P\Big(\boldsymbol{\alpha}_i(t+1)|\boldsymbol{\alpha}_i(t), \theta_i, \boldsymbol{\lambda}\Big)p(\boldsymbol{\lambda}) \quad (14)$$

$$P(\theta_i|\boldsymbol{\alpha}_i, \boldsymbol{\lambda}) \propto \prod_{t=1}^{T-1} P\Big(\boldsymbol{\alpha}_i(t+1)|\boldsymbol{\alpha}_i(t), \boldsymbol{\lambda}, \theta_i\Big)p(\theta_i). \quad (15)$$

The full conditional distributions for $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ are sampled using a Metropolis–Hastings sampling procedure. The prior for the intercept λ_{0k} and the latent trait θ_i are assumed to be from normal distribution. The prior for λ_{1k} is from the lognormal distribution, which constrains it as positive to make θ_i identifiable. The prior for the other coefficients in the learning model can be chosen either from normal or lognormal distribution to reflect the assumption of whether some covariates will make a positive contribution to increase the odds of mastery. The specified priors of the learning model (Equation 7) are provided in the Simulation Study section.

The MCMC Algorithm

In summary, we present the Metropolis–Hastings within Gibbs algorithm to estimate the proposed learning model (Equation 4). We use $p(\cdot)$ to denote the appropriate priors for the learning model parameter and at iteration r,

(1) Update the coefficients vector $\boldsymbol{\lambda}$. Suppose there are n covariates in the transition model, then there are a total of $(n+1)K$ coefficients, including the intercept. For $h \in \{0, 1, 2, \ldots, n\}$, let $\boldsymbol{\lambda}_{-h} =: \{\boldsymbol{\lambda}_m \in \boldsymbol{\lambda}, m < h\}$ and $\boldsymbol{\lambda}_{+h} := \{\boldsymbol{\lambda}_m \in \boldsymbol{\lambda}, m > h\}$. For example, in the learning model example (Equation 7), $n = 3$, when $h = 1, \boldsymbol{\lambda}_{-1} = \{\boldsymbol{\lambda}_0\}$, $\boldsymbol{\lambda}_{+1} = \{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2\}$. Note that $\boldsymbol{\lambda}_{-0}$ and $\boldsymbol{\lambda}_{+n}$ are defined as the empty set. Sample the elements of $\boldsymbol{\lambda}$ sequentially, that is, draw λ^*_{hk} from uniform $(\lambda^{r-1}_{hk} - \delta_{\lambda_{hk}}, \lambda^{r-1}_{hk} + \delta_{\lambda_{hk}})$ and accept $\boldsymbol{\lambda}^*_h = (..\lambda^r_{h,k-1}, \lambda^*_{hk}, \lambda^{r-1}_{h,k+1} \ldots)$ with probability

$$\pi(\boldsymbol{\lambda}^*_h, \boldsymbol{\lambda}^{r-1}_h) = \frac{\prod_{i=1}^{N} \prod_{t=1}^{T-1} P\left(\boldsymbol{\alpha}^{r-1}_i(t+1) | \boldsymbol{\alpha}^{r-1}_i(t), \theta^{r-1}_i, \boldsymbol{\lambda}^*_h, \boldsymbol{\lambda}^{r-1}_{+h}, \boldsymbol{\lambda}^r_{-h}\right) p(\lambda^*_{hk})}{\prod_{i=1}^{N} \prod_{t=1}^{T-1} P\left(\boldsymbol{\alpha}^{r-1}_i(t+1) | \boldsymbol{\alpha}^{r-1}_i(t), \theta^{r-1}_i, \boldsymbol{\lambda}^{r-1}_h, \boldsymbol{\lambda}^{r-1}_{+h}, \boldsymbol{\lambda}^r_{-h}\right) p(\lambda^{r-1}_{hk})}.$$

(2) Update $\boldsymbol{\theta}$. Draw $\theta^*_i \sim N(\theta^{r-1}_i, \sigma^2_{c\theta})$, and accept θ^*_i with probability

$$\pi(\theta^*_i, \theta^{r-1}_i) = \frac{\prod_{t=1}^{T=1} P\left(\boldsymbol{\alpha}^{r-1}_i(t+1) | \boldsymbol{\alpha}^{r-1}_i(t), \boldsymbol{\lambda}^r, \theta^*_i\right) p(\theta^*_i)}{\prod_{t=1}^{T-1} P\left(\boldsymbol{\alpha}^{r-1}_i(t+1) | \boldsymbol{\alpha}^{r-1}_i(t), \boldsymbol{\lambda}^r, \theta^{r-1}_i\right) p(\theta^{r-1}_i)}.$$

Note that the prior distribution for θ_i is set as $N(0, 1)$ to guarantee identifiability of θ_i.

(3) Update $\boldsymbol{\alpha}$. We will update $\boldsymbol{\alpha}^r_i(1)$, $\boldsymbol{\alpha}^r_i(2), \ldots, \boldsymbol{\alpha}^r_i(T)$ sequentially for each examinee $i = 1, 2, \ldots, N$. For examinee i at time point t, draw $\boldsymbol{\alpha}_i(t)$ from the space $V_i(t)$ based on the probability $\tilde{\pi}_{ict}$ in Equation 9 through 11. The full conditional distribution π is updated based on Equation 12.

(4) Update s_j and g_j. For $j \in I_t$, $t = 1, 2, \ldots, T$.

   (a) Sample $g^*_j | s^{r-1}_j \sim \beta(\tilde{a}_g, \tilde{b}_g) \mathcal{I}_{\{0 \leq g^*_j < 1 - s^{r-1}_j\}}$
   (b) Sample $s^*_j | g^r_j \sim \beta(\tilde{a}_s, \tilde{b}_s) \mathcal{I}_{\{0 \leq s^*_j < 1 - g^r_j\}}$.

In the model calibration procedure, we assume students' latent attribute profiles will change with time, while the CDM model parameters are static. At a given time point, Monte Carlo simulation studies suggest CDMs have an invariance property (e.g., see Bradshaw & Madison, 2016; de la Torre & Lee, 2010) that item parameters are independent of the distribution of the examinees' latent attribute profiles as long as the model fits the data. However, the invariance property of CDMs is at the population level, and complications may arise when attributes change over time with a relatively small sample size (e.g., $N < 500$). Namely, if item blocks are not counterbalanced, the estimated DINA model parameters of test blocks in the later stages could be inaccurate. For instance, consider an extreme example where all individuals in a sample mastered the attributes required for the last item. For the DINA model, the estimated guessing parameter may be inaccurate given no individual in the sample guesses on the item, that is, by the final time all individuals are masters and there are no nonmasters on the attribute. A necessary condition for identifying the DINA model parameters (e.g., see G. Xu & Zhang, 2015) is that the population class proportions are nonzero. It is therefore possible that an empirical identifiability problem could arise for estimating guessing parameters for items that receive insufficient exposure to nonmasters of the skills. In our experiment, we use a counterbalanced test design, allowing each item to be positioned throughout all test stages, to guarantee items will be exposed evenly at different time points, having enough masters and nonmasters of the skills exposed to the items. Such a design is also helpful to address other confounding factors such as test fatigue. The details can be found in the next section. When using this design, Equation 13 needs to be modified as

$$\mathcal{T}_j = \sum_{t=1}^{T} \sum_{i|j \in I_t(i)} \eta_{ijt}, \mathcal{S}_j = \sum_{t=1}^{T} \sum_{i|j \in I_t(i), X_{ij(t)}=0}^{N} \eta_{ijt}, \mathcal{G}_j = \sum_{t=1}^{T} \sum_{i|j \in I_t(i), X_{ij}(t)=1}^{N} (1 - \eta_{ijt}), \qquad (16)$$

where $I_t(i)$ denotes the set of the identity of items the student i received at time point t.

Application: A Spatial Reasoning Test With Learning Interventions

A computer-based assessment and training tool were developed to conduct an experiment. The developed methods were applied to the data set collected from this experiment to assess the efficacy of the intervention and the nature of student learning. A simulation study was conducted based on the experimental design to evaluate the proposed MCMC algorithm.

The Spatial Reasoning Test

In the experiment, participants are asked to complete a series of 50 items on a computer-based assessment of spatial rotation ability. The test is developed based on an extended version of the Purdue Spatial Visualization Test (PSVT; Yoon, 2011). The whole test consists of five test blocks each containing 10 questions, followed by a learning intervention block. The five test blocks thus represent five different time points. Participants first answer the questions in a

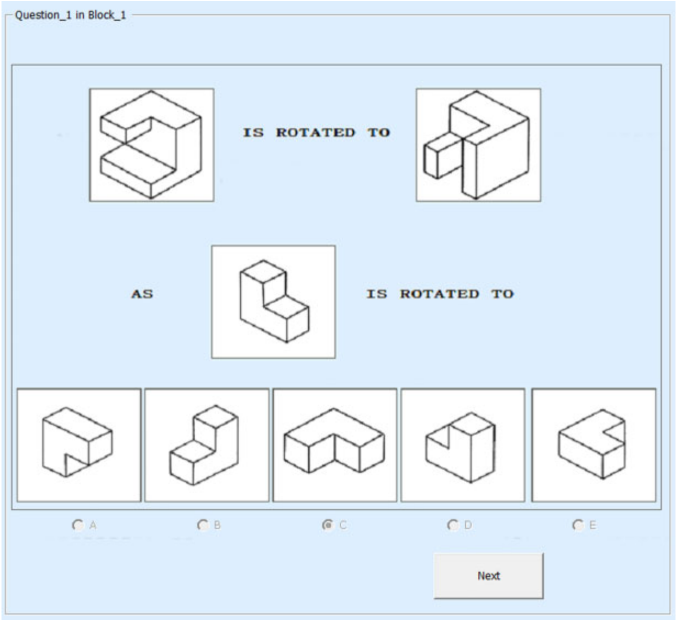

FIGURE 3. *Test block.*

test block then proceed to a learning block in which they can get feedback on their responses to the questions in the previous test block and play with the learning intervention. Items in the test block (Figure 3) include a reference item that is rotated and then subjects are presented a new object and must determine which of five options corresponds to the rotated version. All items include x- and y-axis rotations with objects of varying complexity. Two types of learning interventions were developed. In the first type (Figure 4), for each item in the learning block, a graphical box is provided in which the participant is to use a left-to-right or an up-and-down bar to rotate the 3-D model along either the horizontal or the vertical axis into the correct position. This is the active learning element of the study designed to cause learning of these rotation tasks. In the second type (Figure 5), compared with the first one, another graphical box is provided to show the participants how the reference item is rotated to the correct position. Four mental rotation skills are identified: (1) 90° x-axis, (2) 90° y-axis, (3)180° x-axis, and (4) 180° y-axis.

In order to balance the item positions and avoid empirical identifiability issues, we developed five versions of the test by rotating the five test blocks as the first block in the test (Table 1). During the experiment, those five tests were randomly assigned to participants to guarantee that different test blocks can have

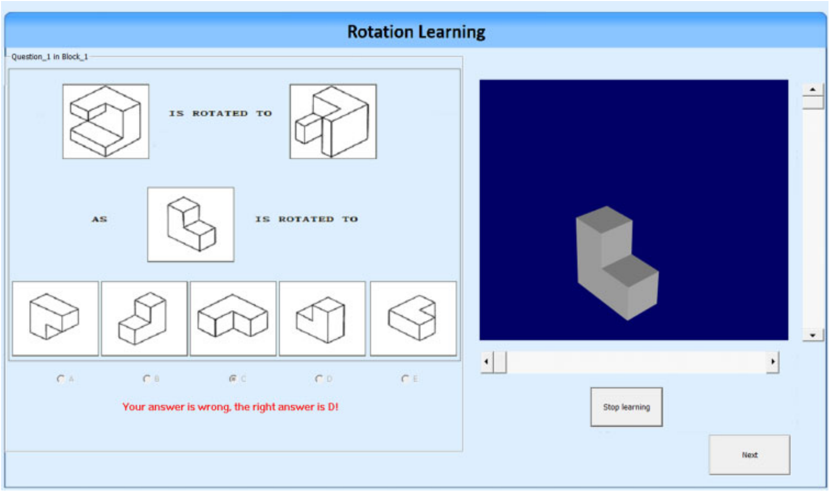

FIGURE 4. *Learning block: Type 1.*

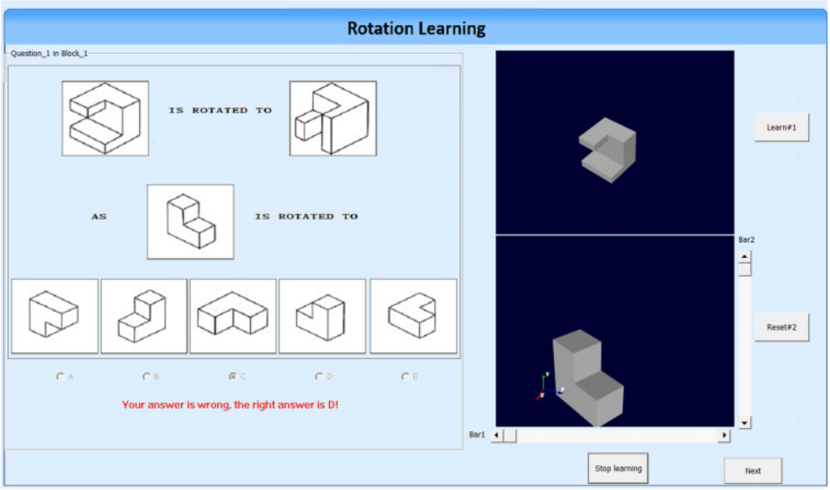

FIGURE 5. *Learning block: Type 2.*

an equal chance to be the first block among the participants. A total of 351 college students participated in this experiment, and among them 177 took the 5 tests with the first type of learning intervention and 174 received the second type of learning intervention. The distribution of the number of participants

TABLE 1.
Balanced Block Design

	order				
Test Version	1	2	3	4	5
Test 1	Test block 1	Test block 2	Test block 3	Test block 4	Test block 5
Test 2	Test block 2	Test block 3	Test block 4	Test block 5	Test block 1
Test 3	Test block 3	Test block 4	Test block 5	Test block 1	Test block 2
Test 4	Test block 4	Test block 5	Test block 1	Test block 2	Test block 3
Test 5	Test block 5	Test block 1	Test block 2	Test block 3	Test block 4

TABLE 2.
Distribution of the Number of Participants

Learning Intervention	Test 1	Test 2	Test 3	Test 4	Test 5	Total
Type 1	40	40	32	36	29	177
Type 2	34	34	43	29	34	174
Total	74	74	75	65	63	351

within each type of learning intervention and each test version is summarized in Table 2.

A Simulation Study

Before analyzing the data set collected from the learning experiment, a Monte Carlo simulation study is presented to demonstrate the accuracy and convergence of the proposed MCMC algorithm. We mimic the experimental design presented above to simulate different conditions in terms of the number of attributes $K = 4,6$ and the number of examinees $N = 500, 1{,}500$. Under each simulation condition, the Q-matrix for each test block has a similar structure and satisfies the identifiability condition (G. Xu & Zhang, 2015). Two cases are considered for the simulation of the initial latent attribute profile $\boldsymbol{\alpha}_i(1)$ for $i = 1, 2, \ldots, N$. The first is that we sampled each examinee's attribute profile uniformly from the 2^K possible attribute patterns. The second way is based on a multivariate normal threshold model (Wang & Douglas, 2015) to mimic a realistic situation in which attributes are correlated and of unequal prevalence. The attribute pattern in the later stage, $\boldsymbol{\alpha}_i(t), t > 1$, was simulated based on the proposed transition model. The learning model example (Equation 7) with a parsimonious transition model is considered. The learning parameters were set as $\lambda_0 = -1.5$, $\lambda_1 = 1$, $\lambda_2 = 0.1$, $\lambda_3 = 0.3$. Examinees' general learning abilities θ_i were simulated

TABLE 3.
Mean and SD *of Learning Model Parameters Estimates Over 50 Independent Replications*

$\lambda_0 = -1.5, \lambda_1 = 1.0, \lambda_2 = 0.1, \lambda_3 = 0.3$										
K	N	$\boldsymbol{\alpha}$	$\hat{\lambda}_0$	$SD\ (\hat{\lambda}_0)$	$\hat{\lambda}_1$	$SD\ (\hat{\lambda}_1)$	$\hat{\lambda}_2$	$SD\ (\hat{\lambda}_{21})$	$\hat{\lambda}_3$	$SD\ (\hat{\lambda}_3)$
4	1,500	Uniform	−1.27	.10	.84	.07	.13	.04	.24	.02
		Normal	−1.29	.09	.81	.06	.08	.03	.27	.02
6		Uniform	−1.20	.09	.82	.05	.10	.03	.24	.02
		Normal	−1.18	.08	.85	.05	.08	.02	.26	.02
4	500	Uniform	−1.31	.16	.74	.12	.15	.06	.22	.03
		Normal	−1.31	.14	.80	.10	.14	.04	.25	.03
6		Uniform	−1.35	.15	.78	.10	.13	.04	.26	.03
		Normal	−1.26	.13	.75	.09	.12	.03	.23	.03

Note. *SD* = standard deviation.

from $N(0, 1)$. The slipping and guessing parameters for the DINA model were fixed at 0.1 for all items. The priors for the parameters were as follows:

$$\lambda_0 \sim \text{Normal}\ (0, 1), \lambda_1 \sim \text{Lognormal}\ (-0.5, 0.5),$$

$$\lambda_2 \sim \text{Lognormal}\ (-1, 0.6), \lambda_3 \sim \text{Lognormal}\ (-1, 0.6)$$

$$\theta_i \sim \text{Normal}\ (0, 1), \boldsymbol{\pi} \sim \text{Dirichlet}(1, ..., 1)$$

$$s_j \sim \text{Beta}\ (1, 1), g_j \sim \text{Beta}(1, 1), j \in I_t, t = 1, \ldots, 5.$$

The MCMC program was written in C++ using the Rcpp version 0.12.11 package (Eddelbuettel, 2013). For each replication, five chains were run and the multivariate potential scale reduction factor $\hat{R}$ was computed (Brooks & Gelman, 1998). Using this criterion, we verified the MCMC program converged with a burn-in of 2,000 followed by 3,000 iterations. Estimation accuracy was assessed by computing the expected values and standard errors for learning parameters, root mean squared error (RMSE) for s_j, g_j and attribute pattern agreement (PAR), $\sum_{i=1}^{N} \frac{I[\widehat{\boldsymbol{\alpha}}_i(t)=\boldsymbol{\alpha}_i(t)]}{N}$, and attribute agreement rate (AAR), $\sum_{i=1}^{N}\sum_{k=1}^{K} \frac{I[\hat{\alpha}_{ik}(t)=\alpha_{ik}(t)]}{NK}$ over 50 replications.

The mean estimates and standard deviations of the estimates for $\boldsymbol{\lambda}$ under different simulation conditions are documented in Table 3. The mean estimates are close to the true values with small standard errors, which indicate a relatively good estimation of those coefficients in the transition model can be obtained. Table 4 presents PAR, AAR, and the correlation between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$. The results indicate the MCMC program can produce accurate estimates of students'

TABLE 4.
Classification Rates

$T = 5$			PAR (%)					AAR (%)					
			T					T					
K	N	$\boldsymbol{\alpha}$	1	2	3	4	5	1	2	3	4	5	$\mathrm{Cov}(\theta, \hat{\theta})$
4	1,500	Uniform	84.2	91.0	95.3	98.1	99.4	95.5	97.6	98.8	99.5	99.9	60.7
		Normal	88.1	91.0	95.4	97.8	99.4	96.7	97.6	98.8	99.5	99.9	60.3
6		Uniform	64.3	78.1	87.9	93.9	98.4	92.5	95.7	97.7	98.9	99.7	69.3
		Normal	72.1	78.9	87.7	93.3	98.2	94.4	95.8	97.6	98.7	99.7	70.3
4	500	Uniform	84.1	91.2	95.6	98.1	99.4	95.5	97.6	98.8	99.5	99.8	60.2
		Normal	87.7	91.0	95.0	99.2	95.9	96.6	97.6	98.7	99.4	99.9	64.6
6		Uniform	62.3	77.2	86.3	93.7	98.4	92.1	95.5	97.4	98.8	99.7	69.2
		Normal	71.5	79.8	87.3	95.5	97.7	94.3	96.0	97.5	99.2	99.6	66.5

Note. PAR = pattern agreement; AAR = attribute agreement rate.

attribute profiles at different time points, and the accuracy increases with time. Figures 6 and 7 plot the Monte Carlo estimated expected values and RMSE of the slipping and guessing parameters in different simulation conditions. Figure 6 shows that the slipping parameters estimated by the proposed algorithm have small biases around 0.005 with the largest one equal to 0.027. For the guessing parameters, the biases are about −0.01 and with the largest bias equal to 0.079. Similarly, the small RMSEs for the slipping and guessing parameters across different conditions support the accuracy of the proposed algorithm.

In addition to the above simulation conditions, we designed another set of conditions to better validate our real data application in the following section. The sample size N is fixed as 351, reflecting the actual sample size in our experiment. The s and g parameters were simulated from a uniform distribution bounded by 0.05 and 0.6 and guarantee that $1 - s_j - g_j > 0$ to mimic the s and g parameters estimated from the real data application. The rest of the conditions for the learning model parameters and the examinees' true learning trajectories are simulated following the same way in our first set of simulation conditions. The full results are documented in the Online Supplemental Material (Figure S14–S15; Table S9–S10). In summary, the proposed model can have generally good recovery of the measurement model parameters (s and g) and classification results with such a relatively small sample size and moderate guessing and slipping parameters. The estimation of the learning model parameters $\boldsymbol{\lambda}$ is less accurate than the results when $s_j = g_j = 0.1$. In the future, we need to further investigate the model identifiability and consistency of the model parameters.

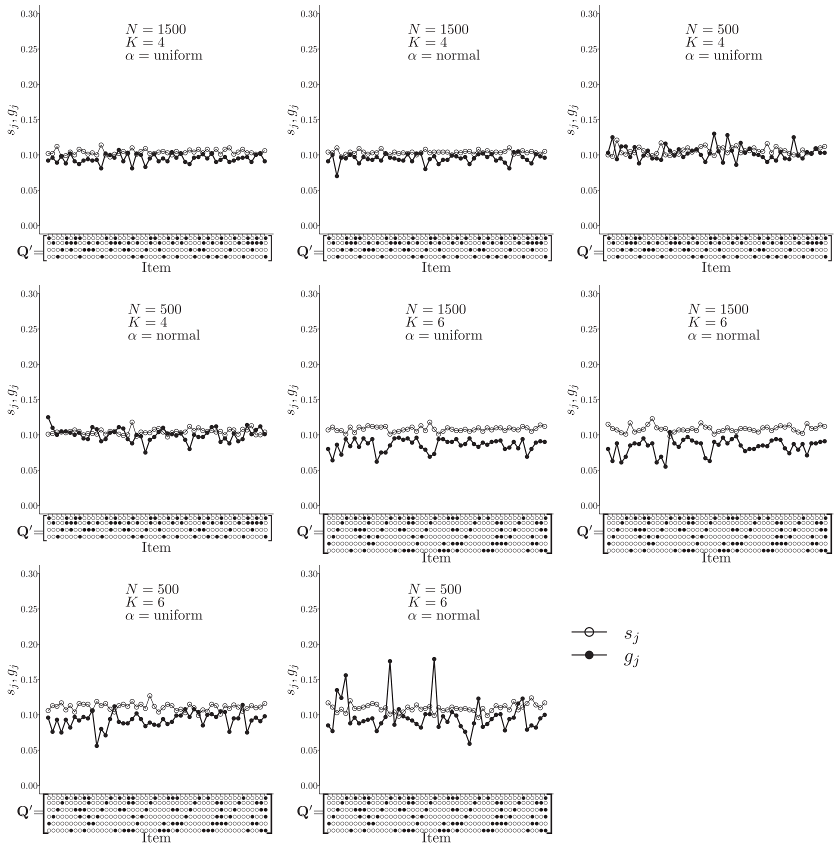

FIGURE 6. *Markov chain Monte Carlo expected slipping and guessing parameters under different simulation conditions. The true value for* s *and* g *were fixed as .1. The skills required for each item are listed in the* K × J $\boldsymbol{Q'}$ *matrix along the* x*-axis where* o = *"*0*" and* • = *"*1.*" Expected values were estimated from 50 replications.*

Data Analysis for the Spatial Reasoning Study

In this section, we report an application of the proposed learning model to 351 participants' responses to 50 questions collected from the designed spatial reasoning test. For the transition model from time point t to $t + 1$, the identified covariates include the continuous general learning ability θ, the learned attribute skills at a previous time point, the type of learning intervention (LV), and the practice each subject received related on skill k until this time. The transition model we consider is

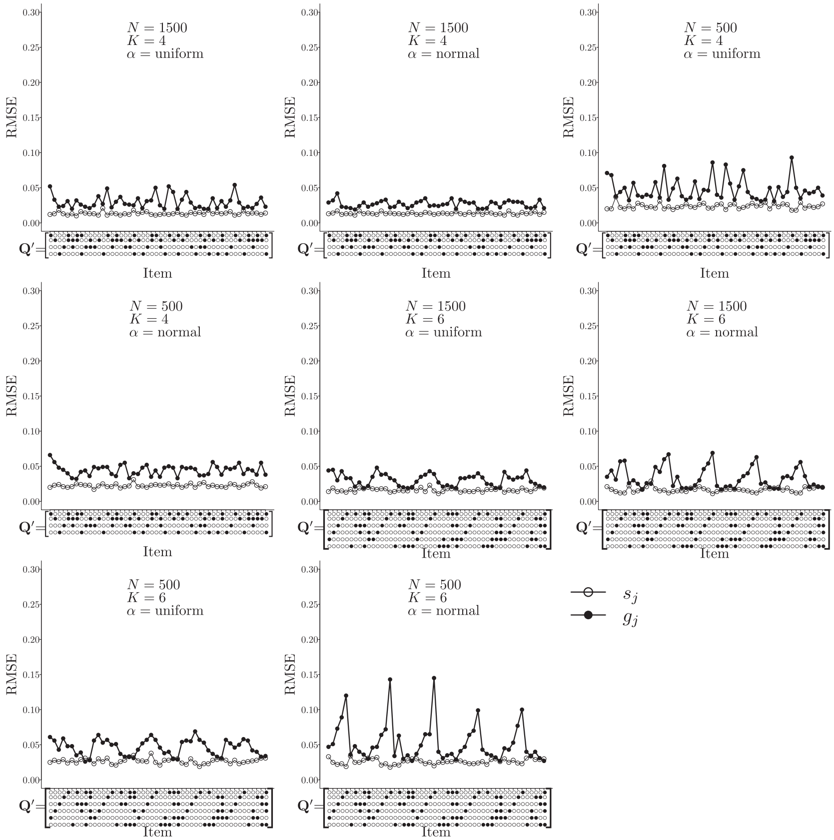

FIGURE 7. *Markov chain Monte Carlo root mean square error (RMSE) of slipping and guessing parameters under different simulation conditions. The true values for* s *and* g *were fixed as .1. The skills required for each item are listed in the* K × J $\boldsymbol{Q}'$ *matrix along the* x*-axis where* o = *"0" and* • = *"0." RMSE values were estimated from 50 replications.*

$$\operatorname{logit}\left[P\Big(\alpha_{ik}(t+1) = 1 \mid \alpha_{ik}(t) = 0, \boldsymbol{\alpha}_i(t)\Big)\right]$$
$$= \lambda_0 + \lambda_1\theta_i + \lambda_2\sum_{l\neq k}\alpha_{il}(t) + \lambda_3\sum_{m=1}^{t}\sum_{j=1}^{J_t} q_{jkm} + \lambda_4 LV_i, \tag{17}$$

where LV_i is a dummy variable with 0 indicating the first type of learning intervention and 1 indicating the second type of learning intervention. The number of items student i has practiced related to skill k is $\sum_{m=1}^{t}\sum_{j=1}^{J_t} q_{jkm}$, which means we put the same weight on each item ($c_{jk} = 1$).

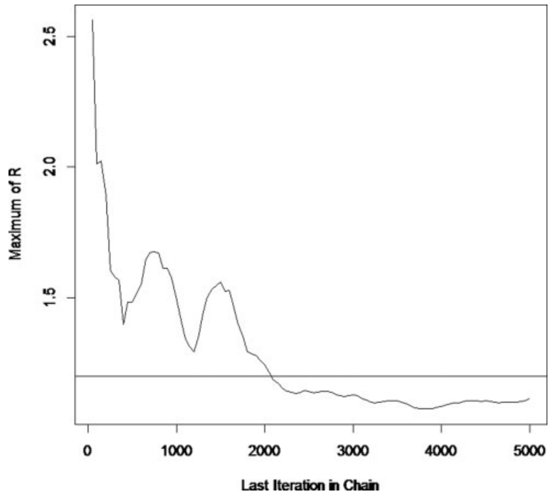

FIGURE 8. *Plot of the maximum Brooks–Gelman potential scale reduction factor,* $\hat{\mathrm{R}}$*, of the 105 parameters. Note the maximum* R *is based upon five independent chains with five different starting values and assesses convergences of the learning model parameters* λ *and deterministic input, noisy "and" gate model parameters* $\boldsymbol{\beta}$*. The plot is based upon chain lengths from 50 to 8,000 in increment of 50.*

There are a total of 105 parameters related to the leaning model parameters $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$ and the DINA model parameters $\boldsymbol{\beta} = (s_j, g_j)_{j=1}^{50}$. The same type of priors as those in the simulation studies were used for the parameters in the real data analysis as well. Figure 8 plots the maximum Brooks–Gelman $\hat{R}$ of the 105 parameters to provide information about the posterior convergence for $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$. Five independent chains with 10,000 iterations were implemented with random starting values. Specifically, starting values for λ_1,λ_2, λ_3, λ_4, s_j, and g_j were drawn from uniform distributions, and λ_0 and θ_i were drawn from normal distributions. The reduction factor $\hat{R}$ was calculated for each of the 105 parameters, and we obtained the maximum of these values to assess the convergence of $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ in a conservative way. The plot is based upon chain lengths from 50 to 5,000 in increment of 50. The posterior distribution was assumed to be stationary based on a rule of thumb that R should be below 1.1 or 1.2. Figure 8 indicates that the maximum $R < 1.1$ after 2,000 iterations. In the following analysis, we use a burn-in of 5,000 iterations to compute parameter estimates.

Table 5 documents the MCMC estimates of the learning model coefficients. First, the 95% credible intervals of the five parameters do not contain 0, indicating the significance of estimated parameters. The positive coefficient $\lambda_1 = 1.861$ for the general learning ability θ shows the significant variability among

TABLE 5.
Markov Chain Monte Carlo Parameters for Learning Model Parameters

Estimation	λ_0	λ_1	λ_2	λ_3	λ_4
Posterior mean	−3.268	1.861	.277	.055	.276
95% Credible interval	[−4.128, −2.608]	[0.834, 2.888]	[0.001, 0.553]	[0.018, 0.092]	[0.013, 0.539]

different participants' learning rates. The coefficient $\lambda_2 = 0.277$ reflects that the learned attributes during the test process can have a positive contribution to the mastery of the unmastered skills. The coefficient λ_3 relates to the effect of practice on student's learning. Note that the quantity $\sum_{m=1}^{t}\sum_{j=1}^{J_t} q_{jkm}$ measures the number of questions related to the skill k the participants have practiced. Controlling for other covariates, if the student practices five more questions on a skill, then the odds of mastering this skill will increase 30%. The coefficient λ_4 indicates there is a significant difference between the efficacy of the first type of learning intervention and that from the second type. Label the test with the first type of learning intervention and the one with the second type of learning intervention as LI_1 and LI_2, respectively. In the learning model, a dummy variable $\mathrm{LV} = 0,1$ is used to differentiate the two types of learning interventions. Fixing the other parameters, we can obtain the odds ratio for learning skill k using LI_1 and LI_2 as

$$\frac{\mathrm{odds}(\mathrm{LV}=1)}{\mathrm{odds}(\mathrm{LV}=0)} = \exp(0.276) = 1.318.$$

This means the odds of mastering skill k using LI_2 will be 31.8% higher than that of using LI_1, controlling for the other covariates. In other words, by showing the participants how the reference item is rotated to the correct position, as the additional function in LI_2, has a better effect on learning. Figure 9 documents the participants' opinions regarding the two learning interventions. The results are consistent with what we infer from the learning model. This indicates the proposed learning model can be used to assess the efficacy of different learning interventions.

Figure 10 reports the **Q** matrix and estimates of the DINA item parameters. The columns of the Q-matrix represent the four skills: (1) 90° x-axis, (2) 90° y-axis, (3) 180° x-axis, and (4) 180° y-axis. In order to guarantee the identifiability of all possible attribute patterns, each test block contains a complete sub-Q-matrix. Note that the guessing parameters range from 0.219 to 0.784, and the slipping parameters range from 0.016 to 0.573, demonstrating the heterogeneity in slipping and guessing parameters. One reason for the relatively large estimated guessing parameters might be due to the simple shapes for some items and

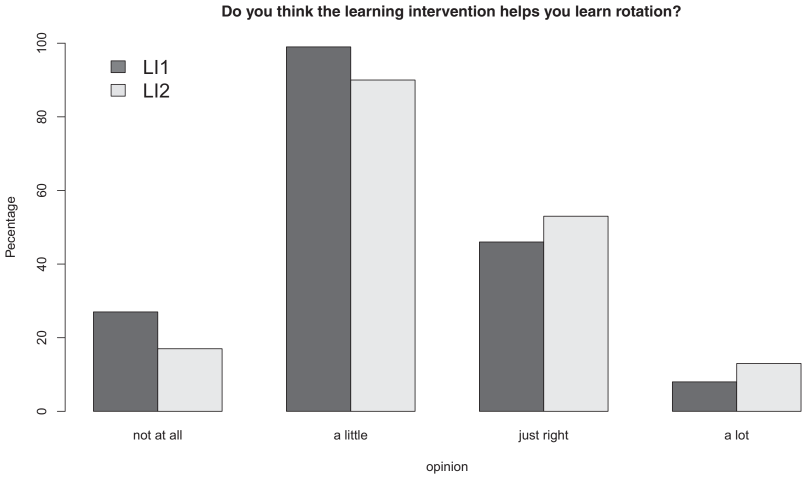

FIGURE 9. *Participants' opinions regarding the effectiveness of two learning interventions.*

distractors that can be easily eliminated from consideration. The 50 items were constructed based on a set of 30 PSVT-R items. Specifically, 30 items of the 50 items were identical to the previous 30 PSVT-R items, and the rest 20 items were developed based on the same principles as the original 30 items. A previous study (Culpepper, 2015) estimated the slipping and guessing parameters for the 30 items based on responses from a standard static cognitive diagnostic test. We compared the MCMC estimates of slipping and guessing parameters on the same set of 30 items with those in Culpepper (2015). The results are presented in Figure 11. The estimates of the slipping and guessing parameters from both studies are quite consistent.

Table 6 documents the estimated mastery rate for each of the four skills over time. We can observe an increase of mastery rates for each skill with time, which shows the effectiveness of the designed learning blocks in the test. Furthermore, Table 7 presents the frequencies and proportions of the number of mastered skills at each time point. The percentage of participants who were masters of none of the four attributes reduces 4.7% throughout the exam, and the percentage of participants who were masters of all four skills increases about 12.25% after practicing rotations using the designed learning interventions. Figure 12 presents the skill mastery rates over time for LI_1 and LI_2. The results are consistent with the estimated learning parameter λ_4 and indicate that LI_2 has better learning effectiveness than LI_1.

The reliability of the classification results based on the index proposed by J. Templin and Bradshaw (2013) at the five time points are reported in Table 8 . The classification reliabilities for all skills are higher than .85 and tend to increase

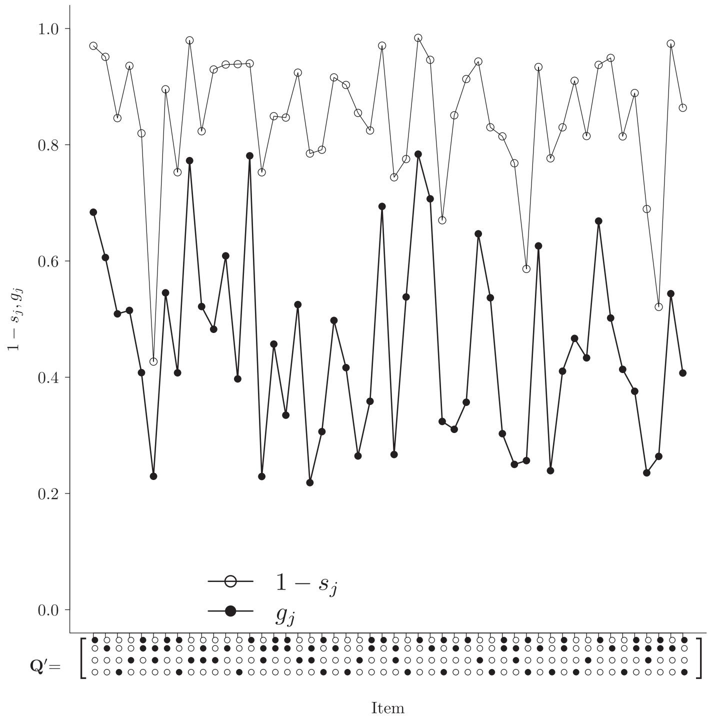

FIGURE 10. *Markov chain Monte Carlo estimate of 1 −* s_j *and* g_j. *The skills required for each item are listed in the* K × J ***Q′*** *matrix along the* x*-axis where* o = *“0” and* • = *“0”. Expected values were estimated from five independent chain with length 10,000 and burn in of 5,000.*

with the increase in time. The pairwise correlations of the four skills are around .57 to .85, indicating those skills are moderately correlated but clearly separable. In addition, we use the posterior predictive model checking (PPMC) method (Rubin, 1984) to assess model fit. Such a method has been applied to assess model fit in both IRT models and cognitive diagnostic models (Sinharay, 2006; Sinharay & Almond, 2007; Sinharay, Johnson, & Stern, 2006). The PPMC method compared the observed data with replicated data (data that are predicted by the model) using a number of diagnostic measures that are sensitive to model misfit. The details of this method and its application to psychometrics can be found in Sinharay, Johnson, and Stern (2006). We use a graphical display of the PPMC method by comparing the observed proportion correct with the predicted proportion correct on each item. Figure 13 presents the overall item fit plots for all items across different time points. The vertical axis represents the proportion correct. For any item, a point denotes the observed proportion correct, and a box represents the distribution of the replicated proportion correct for that item. The

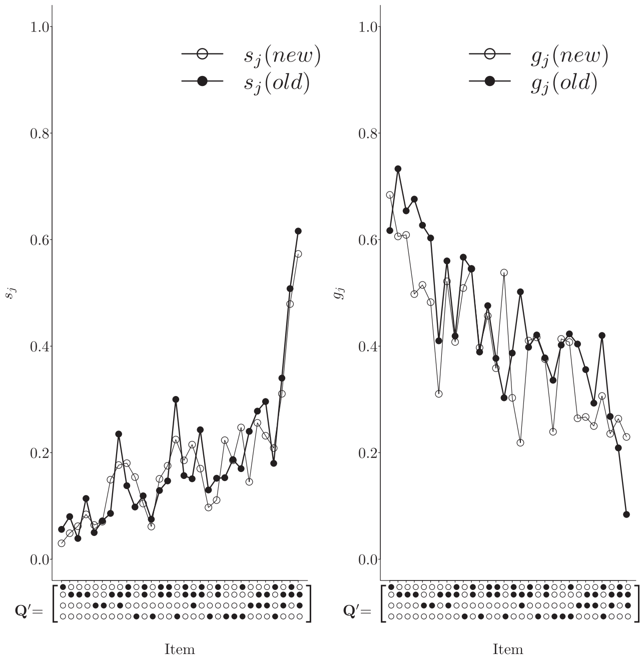

FIGURE 11. *Comparison of the Markov chain Monte Carlo (MCMC) estimates of* $\mathrm{s_j}$ *and* $\mathrm{g_j}$ *with those from Culpepper (2015) on the set of 30 common items.* (new) *represents the MCMC estimates from the current study,* (old) *represents the MCMC estimates from Culpepper.*

TABLE 6.
Skill Mastery Rate Over Time

Skill	Time Point 1 (%)	Time Point 2 (%)	Time Point 3 (%)	Time Point 4 (%)	Time Point 5 (%)
90° *x*-axis	76	80	81	83	86
90° *y*-axis	80	85	87	88	89
180° *x*-axis	71	76	78	80	82
180° *y*-axis	67	72	75	77	80

whisker of the box stretches to include the 5th and 95th percentiles of the empirical distribution, and a notch near the middle of the box denotes the median. We can observe that all the observed proportions fall within the 90% predictive intervals, and most of them are near the center of the replicated values. In order to assess the model fit in terms of learning, we investigate the observed proportion correct at different time points and the predicted ones for each item. Except for

TABLE 7.
Frequencies and Proportions of the Number of Mastered Skills at Each Time

Number of Mastered Skills	Time Point 1	Time Point 2	Time Point 3	Time Point 4	Time Point 5
0	34 (9.97%)	23 (6.55%)	20 (5.70%)	20 (5.70%)	20 (5.70%)
1	50 (14.25%)	38 (10.83%)	27 (7.69%)	26 (7.41%)	24 (6.84%)
2	39 (11.11%)	43 (12.25%)	48 (13.68%)	36 (10.26%)	16 (4.56%)
3	7 (1.99%)	17 (4.84%)	16 (4.56%)	21 (5.98%)	18 (5.13%)
4	220 (62.68%)	230 (65.53%)	240 (68.38%)	248 (70.66%)	263 (74.93%)

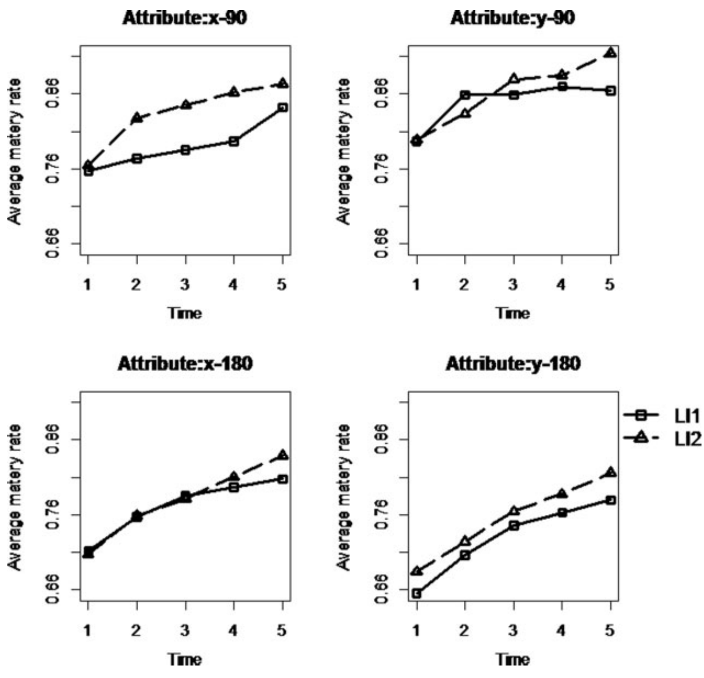

FIGURE 12. *Comparison of the skill mastery rate using two types of learning interventions.*

Item 33 and Item 35, which have the observed proportion correct at Time Point 5 lower or at the lower boundary of the predicted interval, for all other items, the observed proportion correct values fall into the 90% predictive intervals. The lack of fit for Item 33 and Item 35 at Time Point 5 may due to students' fatigue, making the observed value lower than the predicted value.

TABLE 8.
Attribute Classification Reliability

Skill	Time Point 1	Time Point 2	Time Point 3	Time Point 4	Time Point 5
90° x-axis	.854	.896	.919	.919	.906
90° y-axis	.865	.906	.928	.935	.925
180° x-axis	.906	.920	.935	.937	.930
180° y-axis	.910	.914	.916	.930	.906

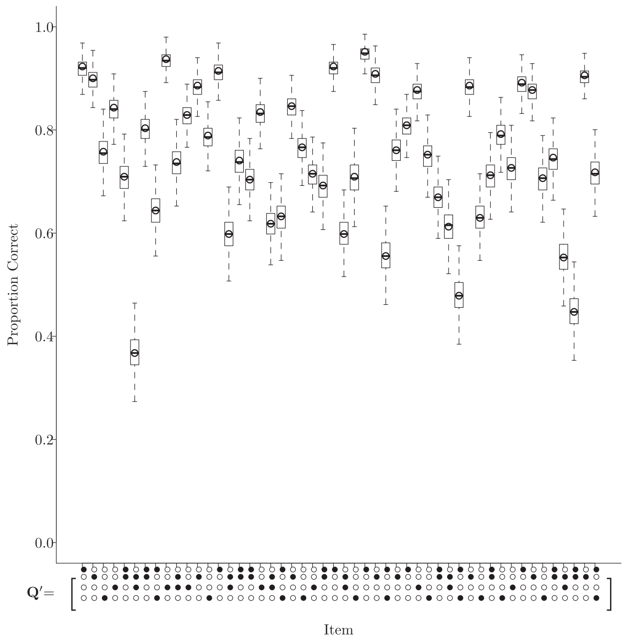

FIGURE 13. *The posterior predictive check based on proportion correct. Boxplot at each item is the distribution of the replicated proportion corrects. The circle at each item is the observed proportion correct.*

Conclusions and Discussions

A learning model that integrates a cognitive diagnostic model with a hidden Markov model using several latent and observed covariates to model transition probability is proposed in this study. The proposed model is applied to a computerized assessment with learning interventions to track the change of students' skills directly over time and evaluate the efficacy of two types of learning interventions. This new method can be applied in a variety of learning environments in which repeated measures are obtained at different time points to track a change in skills. The spatial reasoning example in this study included five blocks of items with interventions taking place between them, for a total training and testing time of roughly 1 hr. It may also be used for longer periods of instruction, say for a course lasting a semester, or even longitudinal studies lasting over a year. However, in the latter case, the model assumption that a student will never lose an attribute after having mastered it may need to be relaxed to allow for the possibility of forgetting over a long time period.

Compared with previous research on KT, the proposed model can track the change of multiple skills simultaneously through an appropriate CDM model with a prespecified Q-matrix. In addition, contrasting with previous work that combines CDMs with a latent transition model, the model proposed here allows for the use of both observable and latent covariates in modeling transition probabilities. The spatial reasoning example showed the benefits of practice, an enhanced intervention, and the value of knowing some of the attributes, on the probability of making a transition to a master of a fixed attribute. It also revealed significant heterogeneity in individual learning rates. Demographic variables and other observable variables can easily be considered as covariates with a straightforward implementation in the model.

There are several possible directions for future studies. One is to consider different types of CDMs as the measurement model. The current study used the DINA model that demonstrated very good fit. This can be partly attributed to the very clearly defined and fine-grained skills required for the spatial tasks. The DINA model makes a very plausible assumption that all required skills are required for a correct answer, apart from guessing. This can be a good approximation when the Q-matrix is correct and the items were designed to fit a particular Q-matrix. It also requires that the attributes be very specific and so fine-grained that they can be viewed as binary. Had more broadly defined attributes been studied with a less obvious Q-matrix, a more general CDM might have been required to compensate for misspecification.

An important area of research, both for learning models and for static models, is to determine how general the measurement model should be. In this initial work on learning models, we used the DINA model as an example. As recommended by an anonymous reviewer, in practice, however, several models should be fit and the relative fit be examined. This learning model framework does allow more general models, and future work will examine incorporating more general measurement models in the learning model proposed in this study. Secondly, different forms of the learning transition model could be considered. This might involve a link function other than the logistic. It is also possible that better ways of coding the practice effect or incorporating individual learning rate parameters could improve the model. Thirdly, the model identifiability conditions for the general model that allows for some covariates to have different influences on different skills need to be rigorously studied. Finally, more efficient computational methods may be developed when the goal is to train individuals on a large number of attributes. Increasing the number of attributes could lead to more difficult model fitting, and research in this area could be a worthwhile challenge to support future efforts to model, detect, and accelerate learning within the CDM framework.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partially supported by UIUC campus research board RB15139 and NSF grant SES-1632023.

References

Andersen, E. B. (1985). Estimating latent correlations between repeated testings. *Psychometrika*, *50*, 3–16.

Bradshaw, L., & Madison, M. (2016). Invariance properties for general diagnostic classification models. *International Journal of Testing*, *16*, 99–118

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

Cappé, O., Moulines, E., & Rydén, T. (2005). An information-theoretic perspective on order estimation. In O. Cappé, E. Moulines, & T. Rydén (Eds.), *Inference in hidden Markov models* (pp. 565–601). New York, NY: Springer.

Chen, Y., Culpepper, S. A., Wang, S., & Douglas, J. A. (in press). A hidden Markov model for learning trajectories in cognitive diagnosis with application to spatial rotation skills. *Applied Psychological Measurement*.

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, *110*, 850–866.

Chiu, C.-Y., & Köhn, H.-F. (2015). The reduced RUM as a logit model: Parameterization and constraints. *Psychometrika*, *81*, 350–370.

Chung, H., Walls, T. A., & Park, Y. (2007). A latent transition model with logistic regression. *Psychometrika*, *72*, 413–435.

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-adapted Interaction*, *4*, 253–278.

Culpepper, S. A. (2015). Bayesian estimation of the DINA model with Gibbs sampling. *Journal of Educational and Behavioral Statistics*, *40*, 454–476.

Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*, 287–307.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*, 179–199.

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

de la Torre, J., & Lee, Y. (2010) A note on the invariance of the DINA model parameters. *Journal of Educational Measurement*, *47*, 115–127

Eddelbuettel, D. (2013). *Seamless R and C++ integration with Rcpp*. New York, NY: Springer.

Embretson, S. E. (1991). Implications of a multidimensional latent trait model for measuring change. In L. Collins & J. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 184–197). Washington, DC: American Psychological Association Books.

González-Brenes, J., Huang, Y., & Brusilovsky, P. (2014). General features in knowledge tracing: Applications to multiple subskills, temporal item response theory, and expert knowledge. In M. Mavrikis & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 84–91). Pittsburgh, PA: University of Pittsburgh.

González-Brenes, J. P., & Mostow, J. (2013). What and when do students learn? Fully data-driven joint estimation of cognitive and student models. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 236–240). Memphis, TN: Educational Data Mining Society.

Hartz, S., & Roussos, L. (2008). The fusion model for skills diagnosis: Blending theory with practicality. *ETS Research Report Series*, *2008*, i–57.

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

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.

Kaya, Y., & Leite, W. L. (2017). Assessing change in latent skills across time with longitudinal cognitive diagnosis modeling an evaluation of model performance. *Educational and Psychological Measurement*, *77*, 369–388. doi:0013164416659314.

Khajah, M. M., Huang, Y., González-Brenes, J. P., Mozer, M. C., & Brusilovsky, P. (2014). Integrating knowledge tracing and item response theory: A tale of two frameworks. In M. Kravcik, O. C. Santos, & J. G. Boticario (Eds.), *Proceedings of Workshop on Personalization Approaches in Learning Environments (PALE 2014) at the 22th International Conference on User Modeling, Adaptation, and Personalization* (pp. 7–12). Pittsburgh, PA: University of Pittsburgh.

Li, F., Cohen, A., Bottge, B., & Templin, J. (2016). A latent transition analysis model for assessing change in cognitive skills. *Educational and Psychological Measurement*, *76*, 181–204.

Li, H., Hunter, C. V., & Lei, P.-W. (2015). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, *33*, 391–409.

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

Pardos, Z. A., & Heffernan, N. T. (2010). Modeling individualization in a Bayesian networks implementation of knowledge tracing. In P. De Bra, A. Kobsa, & D. Chin (Eds.), *User Modeling, Adaptation, and Personalization: 18th International Conference, UMAP 2010*, Big Island, HI, USA, June 20–24, 2010, (Vol. 6075, p. 255). Berlin, Germany: Springer Science & Business Media.

Ravand, H. (2015). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*, *1*, 18.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, *12*, 1151–1172.

Sinharay, S. (2006). Model diagnostics for Bayesian networks. *Journal of Educational and Behavioral Statistics*, *31*, 1–33.

Sinharay, S., & Almond, R. G. (2007). Assessing fit of cognitive diagnostic models a case study. *Educational and Psychological Measurement*, *67*, 239–257.

Sinharay, S., Johnson, M. S., & Stern, H. S. (2006). Posterior predictive assessment of item response theory models. *Applied Psychological Measurement*, *30*, 298–321.

Studer, C. (2012). *Incorporating learning over time into the cognitive assessment framework* (PhD thesis). Carnegie Mellon University, Pittsburgh, PA.

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational and Behavioral Statistics*, *10*, 55–73.

Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, *30*, 251–275.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*, 287.

Templin, J. L., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus. *Educational Measurement: Issues and Practice*, *32*, 37–50.

Wang, S., & Douglas, J. (2015). Consistency of nonparametric classification in cognitive diagnosis. *Psychometrika*, *80*, 85–100.

Xu, G., & Zhang, S. (2015). Identifiability of diagnostic classification models. *Psychometrika*, *81*, 625–649.

Xu, Y., & Mostow, J. (2012). Comparison of methods to trace multiple subskills: Is LRDBN best? In K. Yacef, O. Zaïane, H. Hershkovitz, M. Yudelson, & J. Stamper (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining*, (pp. 41–48). Chania, Greece.

Ye, S., Fellouris, G., Culpepper, S., & Douglas, J. (2016). Sequential detection of learning in cognitive diagnosis. *British Journal of Mathematical and Statistical Psychology*, *69*, 139158.

Yoon, S. Y. (2011). *Psychometric properties of the revised Purdue Spatial Visualization Tests: Visualization of rotations (The revised PSVT-R)* (Unpublished doctoral dissertation). Purdue University, West Lafayette.

Authors

SHIYU WANG is an assistant professor at the University of Georgia, 325N Aderhold Hall, Department of Educational Psychology, Athens, GA, 30602; email: swang44@uga.edu. Her research interests focus on theoretical and applied development in latent variable modeling, such as item response theory models and cognitive diagnostic model, and adaptive testing, such as computerized adaptive testing and multistage adaptive testing.

YAN YANG is a research assistant and PhD candidate at the University of Illinois at Urbana-Champaign, Champaigan, IL, 61820; email: yanyang5@illinois.edu. Her research interests are multidisciplinary, including medical robotics, physical activity and health, physical fitness measurement, item response theory, and measurements in kinesiology.

STEVEN ANDREW CULPEPPER is an associate professor at the University of Illinois at Urbana-Champaign, 725 South Wright Street, Champaign, IL 61820; email: sculpepp@illinois.edu. His research interests include statistical modeling in the social sciences, Bayesian models and computation, cognitive diagnosis, and alternative standardized testing formats.

JEFFREY A. DOUGLAS is a professor at the University of Illinois at Urbana-Champaign, 725 South Wright Street, Champaign, IL 61820; email: jeffdoug@illinois.edu. His research interests include multivariate latent variable methodology for educational and psychological testing with an emphasis on cognitive diagnosis and item response theory.

Manuscript received August 27, 2016
First revision received March 9, 2017
Second revision received May 16, 2017
Accepted June 8, 2017