Article

Educational and Psychological Measurement
1–24

DOI: 10.1177/0013164415588946
epm.sagepub.com

A Latent Transition Analysis Model for Assessing Change in Cognitive Skills

Feiming Li[1], Allan Cohen[2], Brian Bottge[3], and Jonathan Templin[4]

Abstract

Latent transition analysis (LTA) was initially developed to provide a means of measuring change in dynamic latent variables. In this article, we illustrate the use of a cognitive diagnostic model, the DINA model, as the measurement model in a LTA, thereby demonstrating a means of analyzing change in cognitive skills over time. An example is presented of an instructional treatment on a sample of seventh-grade students in several classrooms in a Midwestern school district. In the example, it is demonstrated how hypotheses could be framed and then tested regarding the form of the change in different groups within the population. Both manifest and latent groups also are defined and used to test additional hypotheses about change specific to particular subpopulations. Results suggest that the use of a DINA measurement model expands the utility of LTA to practical problems in educational measurement research.

Keywords

Cognitive diagnostic models, Latent transition analysis

[1]University of North Texas, Fort worth, TX, USA
[2]University of Georgia, Athens, GA, USA
[3]University of Kentucky, Lexington, KY, USA
[4]University of Kansas, Lawrence, KS, USA

Corresponding Author:
Feiming Li, University of North Texas Health Science Center, LIB416, 3500 Camp Bowie Blvd, Fort Worth, TX 76109, USA.
Email: feimingli@hotmail.com

Introduction

Learning means that change has occurred in some behavior and studying learning requires statistical models for studying that change. New instructional methods, new materials, and new policies all need to be studied to determine their short- and long-term effectiveness. Much early measurement of educational change is similar to that typified by Bereiter's (1963) efforts to examine differences in raw scores, such as between a pretest and a posttest. In that approach, raw score differences are assumed to convey the weight of the change in an attribute, such as learning. Although this appears to be a straightforward approach, it unfortunately also leads to some fundamentally unsound outcomes. The reliability of change measured in this way, for example, is composed of the combined unreliability of pretest and posttest scores. In addition, change scores have a spuriously negative relation to initial scores because of regression effects. Finally, change may not be measured on the same scale for persons at different initial score levels.

The use of latent variables as found in item response theory (IRT; Hambleton, Rogers, & Swaminathan, 1991; Lord, 1980; Lord & Novick, 1968) can be helpful in resolving some of these problems. Andersen (1985) proposed a multidimensional Rasch model, for example, for repeated testing of the same persons over several occasions. In that model, the item difficulties are assumed to be constant over occasions, but abilities change. Abilities on all occasions are correlated. Embretson (1991) presented a multidimensional IRT model to measure both initial ability and learning, and where learning is the increment in ability between any two successive occasions. Though these models are multidimensional, the multidimensionality refers not to different ability dimensions but rather to the same unidimensional ability measured at different times. Andersen's model and Embretson's model are both appropriate for evaluating the increment in unidimensional latent abilities over repeated measures. They are not as useful, however, for diagnosing the balance of change among several possibly related abilities or cognitive skills over multiple occasions. This is unfortunate as it this latter kind of information that is potentially most useful for educational or psychological diagnostic purposes, for example, to provide information to teachers to help them adjust their instructional emphases and strategies for optimizing student learning.

Several cognitive diagnostic models have been developed to evaluate examinees' status relative to mastery or nonmastery on each of a set of cognitive skills (DiBello, Stout, & Roussos, 1995; Hartz, 2002; Junker & Sijtsma, 2001). These models provide helpful, fine-grained information regarding individual or population-level learning weaknesses and strengths. What is missing, however, is a means for using these models for measuring change in mastery or nonmastery status. The main technical difficulty in this regard stems from the fact that mastery or nonmastery on each cognitive skill is a binary decision, and binary latent variables cannot be handled well by means of variance–covariance-based approaches (Collins & Wugalter, 1992). It is not possible, for example, to simply assume a multivariate normal distribution, such as is the case for continuous latent abilities over repeated measures as used by Andersen (1985) and Embretson (1991).

In the research described here, an existing approach for analyzing change, latent transition analysis (LTA; Collins & Wugalter, 1992), was combined with a cognitive diagnostic model and used to account for change in the latent binary variables measured by cognitive diagnostic models. LTA is typically used for detecting the probabilities that members of different latent groups in the data will remain in those groups or shift into other latent groups. This kind of analysis is an excellent way to study changes in development, such as transitions among the stages described in Piagetian theory. LTA can also be used to study transitions in attitudes, in personality patterns, in behaviors, and in learning. The LTA in combination with a cognitive diagnostic model, as described in this article, presents a new method that can accommodate a far wider range of research problems, accounting for transition statuses on multiple latent categorical variables. It will be possible, for example, to study mathematics learning over a wide range of different cognitive skills. It will further be possible to examine how multiple educational indicators react as a function of particular changes in educational instruction.

The Latent Transition Analysis Model

LTA was initially developed to study stage sequential change in particular types of latent variables called dynamic latent variables (Collins & Wugalter, 1992). Dynamic latent variables include characteristics such as attitudes and personality patterns that change over time. This technique has also been found to be useful for studying transitions subsequent to interventions, such as cessation in tobacco use and reduction of inappropriate behavior. LTA can be a useful method for investigating academic growth when latent variables are categorical (e.g., Boscardin, Muthén, Francis, & Baker, 2008; Compton, Fuchs, Fuchs, Elleman, & Gilbert, 2008; Trentacosta et al., 2011).

In this study, we describe the use of LTA combined with a cognitive diagnostic model to accommodate typically longer educational tests developed to measure multiple cognitive skills. This new class of models enables a researcher to investigate hypotheses about the effect(s) of an intervention (e.g., an instructional program or a change in educational policy) by testing models of changes in the transition probabilities prior to and following an intervention. We provide an example to illustrate the use of this model from a study of an instructional intervention.

Transitions. Transitions are expressed as probabilities indicating the likelihood of changing (i.e., transitioning) from one latent status to another. In the cognitive diagnostic models described below, two statuses, mastery or nonmastery, are estimated for each skill. When considered over two time points, this will yield four transition probabilities between the two occasions: the probability from nonmastery to mastery ($p_{m|n}$), from nonmastery to nonmastery ($p_{n|n}$), from mastery to nonmastery ($p_{n|m}$), and from mastery to mastery ($p_{m|m}$), where n and m stand for nonmastery and mastery, respectively.

Three groups of parameters are estimated in the LTA with a cognitive diagnostic measurement model. The first group of parameters is made up of the latent transition probabilities. This is particularly useful, because it offers the solution to a major type of research question that can now be addressed with this new model: How do the latent cognitive skills mastery statuses change from occasion to occasion? The second group of parameters estimates the mastery or nonmastery status for each individual at the first occasion of measurement. The third group of parameters estimates the relation between the latent statuses and the question. That is, it yields a probability of a correct or incorrect response to a given item for each of the different latent statuses. The second and third groups of parameters also provide the solution to another research question, namely, what is the individual mastery status on each cognitive skill on each occasion. The answer to this question can provide useful information about student or patient status at each point in time.

A simple example of the new model is provided here using a single manifest indicator and a single latent variable: Assume a single categorical manifest variable is measured on T separate occasions, and Y_t denotes the response given at occasion t (where $t = 1, \ldots, T$). There are D levels or categories for each Y_t, and y_t is a particular level or category (where $y_t = 1, \ldots, D$). A_t is a latent variable measured at time t, C is the number of categories of A_t, and a_t is a particular latent class at time t (where $a_t = 1, \ldots, C$). The probability of response pattern $\vec{Y} = (y_1, \ldots, y_t)$ is expressed as follows:

$$P(\vec{Y}) = \sum_{a_1 \ldots a_T = 1}^{C} P(A_1 = a_1) \prod_{t=2}^{T} P(A_t = a_t | A_{t-1} = a_{t-1}) \prod_{t=1}^{T} P(Y_t = y_t | A_t = a_t) , \qquad (1)$$

where $P(A_1 = a_1)$ is the probability of a certain status of the latent variable at the first occasion, $\prod_{t=2}^{T} P(A_t = a_t | A_{t-1} = a_{t-1})$ are the transition probabilities from previous occasions to current occasions, and $\prod_{t=1}^{T} P(Y_t = y_t | A_t = a_t)$ is the measurement model incorporating measurement error. Here, $\prod_{t=2}^{T} P(A_t = a_t | A_{t-1} = a_{t-1})$ measures the transition probability between two adjacent time points, the so-called lag-1 effect. Some studies (Marcoulides, Gottfried, Gottfried, & Oliver, 2008; Nylund, 2007) have introduced high-order effect (i.e., lag-2) to test the lasting and direct impact between nonadjacent time points when the longitudinal data include more than two time points.

Extending the single-indicator LTA model to multiple indicators is straightforward. In the cognitive diagnostic model, however, we have both multiple indicators (e.g., multiple items on a test) and also multiple latent variables (e.g., multiple cognitive skills) at each occasion. For simplicity, assume the latent variables (e.g., the cognitive skills) are independent and their growth transition probabilities are also independent. Let $\vec{A}_t$ denote all cognitive skills at occasion t and A_{kt} denote the cognitive skill k at occasion t with $k = 1, \ldots, K$. Then, the model in Equation (1) can be extended to

$$P(\vec{Y}) = \sum_{a_{kt}=1}^{C} \prod_{k=1}^{K} P(A_{k1} = a_{k1}) \prod_{t=2}^{T} \prod_{k=1}^{K} P(A_{kt} = a_{kt} | A_{k(t-1)} = a_{k(t-1)}) \prod_{t=1}^{T} P(Y_t = y_t | \vec{A}_t = \vec{a}_t), \tag{2}$$

where $\prod_{k=1}^{K} P(A_{k1} = a_{k1})$ is the probability of certain statuses of multiple cognitive skills at the first occasion, $\prod_{t=2}^{T} \prod_{k=1}^{K} P(A_{kt} = a_{kt} | A_{k(t-1)} = a_{k(t-1)})$ represents the transition probabilities for all cognitive skills from previous occasions to current occasions, and the measurement model $P(Y_t = y_t | \vec{A}_t = \vec{a}_t)$ is estimated by the DINA (Deterministic Inputs, Noisy ''AND'' Gate) model (Haertel, 1989; Junker & Sijtsma, 2001; Macready & Dayton, 1977). The DINA model is a cognitive diagnostic model for estimating mastery or nonmastery statuses on multiple cognitive skills based on the raw item responses. This model is discussed below. Parameter invariance is assumed for the DINA measurement model, that is, $P(Y_t = y_t | \vec{A}_t)$ is assumed to be time-homogeneous for $1 \leq t \leq T$. In this study, we calibrate responses from each administration simultaneously by the DINA model and assume the item parameters to be time-homogeneous, that is, estimated as equal for each occasion. Each person's skill mastery status, however, is allowed to change over occasions. In addition, transition probabilities $P(A_{kt} = a_{kt} | A_{k(t-1)} = a_{k(t-1)})$ are assumed to differ for each skill.

DINA Model

The DINA model describes the probability of a correct response as predicted by examinees' skill profiles and the item parameters. In this model, the latent variable is estimated from a vector of 0s and 1s for each examinee, indicating whether or not examinee i answered item j correctly. To indicate which items would load on which skills, a Q-matrix was constructed typically based on the judgment of content experts. The elements of a Q-matrix are specified as $q_{jk} = 1$ when item j measures skill k or 0 otherwise. The mastery or nonmastery state is specified for an examinee for each skill in the DINA model as $a_{ik} = 1$ or 0 indicating whether or not examinee i has mastered skill k.

In the DINA model, each item divides the population into two classes, those who master all required skills for that item and those who miss at least one required skill for that item. Let ξ_{ij} denote whether examinee i has mastered all required attributes for item j,

$$\xi_{ij} = \prod a_{ik}^{q_{jk}}. \tag{3}$$

As in the signal detection model (Green & Swets, 1966), the ξ_{ij} are detected from noisy observations Y_{ij} by two error probabilities, a ''slipping'' parameter, s_j, indicating the false negative rate for item j, and a ''guessing'' parameter, g_j, indicating the false positive rate for item j. s_j is the probability of missing item j for someone classified as mastering all required skills,

$$s_j = P(Y_{ij} = 0|\xi_{ij} = 1), \tag{4}$$

and g_j is the probability of a correct response for someone classified as lacking at least one required skill,

$$g_j = P(Y_{ij} = 1|\xi_{ij} = 0). \tag{5}$$

Therefore, given examinee parameter ξ_{ij} and item parameters s_j and g_j, the probability of a correct response can be written as

$$P(Y_{ij} = 1|a) = (1 - s_j)^{\xi_{ij}} g_j^{(1-\xi_{ij})}. \tag{6}$$

Assuming local independence and independence among examinees, the joint likelihood for all responses under the DINA model is

$$L(\mathrm{s},\ \mathrm{g};\ \mathrm{a}) = \prod\prod P(Y_{ij} = 1|\xi_{ij}, s_j, g_j). \tag{7}$$

Example: Change in Mastery Status in Mathematics Achievement Following an Instructional Intervention

Here we present an example to illustrate the application of LTA-DINA, the LTA with a DINA measurement model, for assessing change subsequent to an instructional intervention. The example presents a multi-wave experiment designed to assess the effects of an instructional treatment called Enhanced Anchored Instruction (EAI; Bottge, Heinrichs, Chan, Mehta, & Watson, 2003) on mathematics achievement of middle schools with learning disabilities (LD) and without learning disabilities (NLD).

There were two EAI treatments in the larger study. One treatment was presented in the first semester of the academic year followed by the second treatment in the second semester of the same academic year. Results for the second of the EAI instructional treatments, Fraction of the Cost (FOC), are reported here. The impact of the EAI intervention on student mathematics achievement was assessed by examining the different transition probabilities following each of the two instructional treatments and by examining the different transition probabilities among different examinee groups.

Procedures

Data. Students in the study were drawn from six mathematics classrooms in a school district in a small upper Midwestern town. The sample consisted of 50 males and 59 females, all in the seventh grade. Nine of the students were diagnosed with LD. The remainder of the students were NLD.

Design of Study. The FOC test (described below) was administered four times, at Weeks 1, 4, 19, and 24 of the academic year. Two instructional treatments were administered to these same students during the school year: A 13-day EAI instructional treatment, called Kim's Komet (KK), was taught between Weeks 1 and 4, and an 11-day EAI instructional treatment called FOC was administered between Weeks 19 and 24. Between these two instructional treatments, teachers followed their regular math curriculum. Four administrations of the FOC test were given. No feedback about correct answers was given to students about their performance in an effort to avoid possible memory effects. As results of this study presented in the sequel suggest, this strategy was somewhat successful.

Both KK and FOC instruction used video-based instructions. The KK video portrayed two girls competing in pentathlon events. The focus of the KK instruction was on developing students' informal understanding of pre-algebra concepts, including linear functions, line of best fit, variables, rate of change (slope), and reliability and measurement error. FOC instruction was designed to simulate mathematics problem solving and was situated in a task requiring students to figure out how they could afford to buy materials for building a skateboard ramp. The FOC video included three middle grades students, a student with LD, an average-achieving student, and a high-achieving student. Previous research suggested that the video format was particularly useful for students with reading difficulties because, in contrast to text-based presentation, it allowed them to access and engage in the solving of meaningful problems along with their classmates (Bottge, Rueda, Serlin, Hung, & Kwon, 2007). After the KK instruction and prior to the FOC instruction, the mathematics teachers followed their regular curriculum to teach units on concepts related to geometry and proportional reasoning. Specifically, the FOC test was given before and after each of the three instructional sequences in the following order: first FOC test administration, KK instruction, second FOC test administration, regular instruction, third FOC test administration, FOC instruction, and fourth FOC test administration. The focus of this example is to illustrate estimation of the effects of each instruction unit on students' learning. As a result, the study focused on lag-1 transition probabilities as these reflect the effects of each instructional sequence.

Measures. The FOC test consisted of 23 short-answer items designed to assess the effects of the FOC instruction. Items were scored dichotomously as either correct or incorrect. The test was administered at each of four time points. Items on the test were constructed to measure students' ability to follow a schematic diagram, take measurements, compute lengths, and determine costs involved in constructing the skateboard ramp. The reliability of this test was previously reported by Bottge et al. (2007) as .80. The reliability based on four administrations in this study was .78, .79, .79, and .78, respectively.

Construction of the Q-Matrix. Construction of the Q-matrix (shown in Table 1) is an important step for cognitive diagnosis modeling as mis-specification will provide

Table 1. Q-Matrix for Fraction of Cost Test.

Items	Number & Operation	Measurement	Problem Solving	Representation
1	1	0	0	0
2	1	0	0	0
3	1	1	0	0
4	1	1	0	0
5	1	1	0	0
6	1	1	0	0
7	1	1	0	0
8	1	1	0	0
9	0	1	0	0
10	0	1	0	1
11	1	1	1	1
12	1	1	1	1
13	1	1	1	1
14	1	1	1	1
15	1	1	1	1
16	0	0	1	0
17	1	1	1	1
18	0	1	0	1
19	1	0	0	1
20	1	0	0	0
21	1	0	0	0
22	1	0	0	0
23	1	0	1	0

misleading results. For the FOC test, four cognitive skills were measured: Number & Operation, Measurement, Problem Solving, and Representation. A mathematics education content expert identified the four skills and indicated which of these four skills were measured by each item on the FOC test. The results of that analysis were used to create a 23×4 Q-matrix with elements q_{jk} that indicated whether skill k was required to solve item j.

Research Questions and Model-Based Research Hypotheses

In this section, we describe a hypothesis testing framework that enables us to determine (1) whether skill mastery statuses changed across the four test administrations and whether the changes subsequent to each instruction were similar or different and (2) whether different groups of examinees had different transition probabilities. We begin by assuming that the four skills are independent and that skill growth is also independent. We further assume that the transition probabilities are different for each skill.

With regard to Question 1, we establish a basis of comparison by first assuming that the population is homogeneous with respect to the transition probability matrix.

Table 2. Transition Probability Matrix for a Two-Category Latent Class Transition Over Four Time Points.

Test Wave 1 → Test Wave 2 (KK instruction)	Test Wave 2 → Test Wave 3 (Regular instruction)	Test Wave 3 → Test Wave 4 (FOC instruction)
$p^{21}_{n\|n} p^{21}_{m\|n}$	$p^{32}_{n\|n} p^{32}_{m\|n}$	$p^{43}_{n\|n} p^{43}_{m\|n}$
$p^{21}_{n\|m} p^{21}_{m\|m}$	$p^{32}_{n\|m} p^{32}_{m\|m}$	$p^{43}_{n\|m} p^{43}_{m\|m}$

Table 2 presents the transition probability matrix for a two-category latent class transition over four occasions of measurements. In the table, m indicates mastery, and n indicates nonmastery. As an example, consider the transition from first to second test wave (i.e., from pre- to post-KK Instruction). The transition probability matrix is composed of four components: $p^{21}_{n|n}$ (remaining in nonmastery status), $p^{21}_{m|m}$ (remaining in mastery status), $p^{21}_{m|n}$ (transitioning from nonmastery to mastery status), and $p^{21}_{n|m}$ (transitioning from mastery to nonmastery status). The rows of each transition matrix are conditional probabilities and, therefore, sum to 1 (e.g., $p^{21}_{n|n}+p^{21}_{m|n}=1$, $p^{21}_{n|m}+p^{21}_{m|m}=1$). Thus, for each transition matrix, only two transition probabilities need be estimated.

Three transition probability matrices were constructed to reflect the effects of KK-instruction (Test Wave 1 to Test Wave 2), regular instruction (Test Wave 2 to Test Wave 3), and FOC-instruction (Test Wave 3 to Test Wave 4). Four different models of these three transition probabilities were tested to determine whether effects of the three different instructional sequences were approximately equivalent or not:

Model A: In this model, the transition probability matrix for each transition is assumed to not be equal to the others, thereby indicating that the effects of all three instructional sequences are different.

$$\begin{vmatrix} p^{21}_{n|n} & p^{21}_{m|n} \\ p^{21}_{n|m} & p^{21}_{m|m} \end{vmatrix} \neq \begin{vmatrix} p^{32}_{n|n} & p^{32}_{m|n} \\ p^{32}_{n|m} & p^{32}_{m|m} \end{vmatrix} \neq \begin{vmatrix} p^{43}_{n|n} & p^{43}_{m|n} \\ p^{43}_{n|m} & p^{43}_{m|m} \end{vmatrix}$$

Model B: In this model, the two transition probability matrices prior to the FOC instruction were assumed to be equal. The transition following FOC instruction, however, were assumed to be different. That is, the effects of KK instruction and the regular math instruction were assumed to be the same, but both differed from the effect of FOC instruction.

$$\begin{vmatrix} p^{21}_{n|n} & p^{21}_{m|n} \\ p^{21}_{n|m} & p^{21}_{m|m} \end{vmatrix} = \begin{vmatrix} p^{32}_{n|n} & p^{32}_{m|n} \\ p^{32}_{n|m} & p^{32}_{m|m} \end{vmatrix} \neq \begin{vmatrix} p^{43}_{n|n} & p^{43}_{m|n} \\ p^{43}_{n|m} & p^{43}_{m|m} \end{vmatrix}$$

Model C: In this model, the transition probability matrix from pre- to post-KK instruction is assumed to be equal to the transition probability matrix from pre- to post-FOC instruction, but not equal to that from pre- to postregular instruction. That is, the two EAI instructional treatments, that is, KK and FOC, were assumed to affect the transition probabilities similarly, but both differed from the effects of regular mathematics instruction.

$$\begin{vmatrix} p^{21}_{n|n} & p^{21}_{m|n} \\ p^{21}_{n|m} & p^{21}_{m|m} \end{vmatrix} = \begin{vmatrix} p^{43}_{n|n} & p^{43}_{m|n} \\ p^{43}_{n|m} & p^{43}_{m|m} \end{vmatrix} \neq \begin{vmatrix} p^{32}_{n|n} & p^{32}_{m|n} \\ p^{32}_{n|m} & p^{32}_{m|m} \end{vmatrix}$$

Model D: In this model, all transition probability matrices are assumed to be equal. That is, all instructional sequences are assumed to have the same effect on students' mastery status.

$$\begin{vmatrix} p^{21}_{n|n} & p^{21}_{m|n} \\ p^{21}_{n|m} & p^{21}_{m|m} \end{vmatrix} = \begin{vmatrix} p^{32}_{n|n} & p^{32}_{m|n} \\ p^{32}_{n|m} & p^{32}_{m|m} \end{vmatrix} = \begin{vmatrix} p^{43}_{n|n} & p^{43}_{m|n} \\ p^{43}_{n|m} & p^{43}_{m|m} \end{vmatrix}$$

With respect to Question 2, that is, whether different groups of examinees had different transition probabilities, the constraint was removed on the population homogeneity assumption, such that different populations were allowed to have different transition probability matrices. The most interesting groups for this study would be the students with LD or NLD. Unfortunately, the sample of students with learning disability was very small (i.e., $N = 9$) and would produce poor precision in estimates of transition probabilities for that group. Therefore, for purposes of illustration, two alternative grouping variables were chosen. Gender was chosen as a manifest group indicator; a latent group indicator reflecting different ability levels was chosen from a previous study of same data. Cohen (2006) reported two latent classes using a mixture Rasch model (Rost, 1990) analysis of these data. The members of Latent Class 1 were significantly lower in ability ($N = 62$, $M = -4.36$, $SD = 1.08$) than members of Latent Class 2 ($N = 47$, $M = 0.18$, $SD = 0.90$) ($p < .01$). Class 1 was labeled as the low-ability group and Class 2 as the high-ability group. The nine students with LD were all classified in the low-ability group. Each of Model A, B, C, and D was extended to test group heterogeneity (female vs. male; and low-ability group vs. high-ability group) by estimating different transition probabilities for the different groups.

Parameter Estimation

A Markov chain Monte Carlo (MCMC) algorithm employing Gibbs sampling was used to estimate the model parameters. This algorithm was implemented in the WinBUGS software (Spiegelhalter, Thomas, & Best, 2003) and used to simulate a Markov chain in which values representing parameters of the model are repeatedly sampled from their full conditional posterior distributions over a large number of iterations. The MCMC algorithm used here samples values in each iteration for each of the parameters in the model conditional on those parameters already estimated. In this study, the model parameters were sampled from their full posterior distributions conditional on the already sampled parameters and examinees' skill mastery parameters.

Prior distributions were specified to derive the posterior distribution for each parameter. For the priors of the two item parameters g_j and s_j, Junker and Sijtsma (2001) used a uniform distribution and de la Torre and Douglas (2004) used a beta distributions. In this study, we used beta (1, 1), which is equal to a uniform (0, 1). For mastery status estimates, a Bernoulli distribution was specified with the mastery probability following a uniform prior (0, 1). The transition probability parameters also were specified by a uniform prior (0, 1).

In MCMC estimation, some information from the initial iterations is discarded as a burn-in. The remaining iterations are based on a chain that is assumed to have converged to its stationary distribution. Estimates of sampled parameters are then calculated from these final iterations. The WinBUGS software provides some indices that help determine whether the chain has converged to a stationary distribution. The length of a suitable burn-in for the example was determined by analyzing convergence statistics provided by WinBUGS. Based on results of the convergence analysis, the burn-in length for all parameters was set at 4,000. Estimates of model parameters were based on the means of the sampled values from *10,000* iterations subsequent to burn-in.

Results

Item Parameter Estimates in DINA Model

In this study, the same test was administered four times. We assumed that memory did not play a significant role, as the test was a performance assessment in which students were asked to show how they would perform each of the measuring, cutting, and assembling tasks involved in building the skateboard ramp, and since students were not given feedback about their correct or incorrect answers after each administration. In addition, we assumed item parameter invariance for the DINA model across four administrations, although population mastery proportions were allowed to change. Further, item parameters were estimated jointly across four administrations. Mastery proportions for each skill, however, were allowed to be estimated freely across administrations.

The model-based hypotheses above deal with transition probabilities, and as a result, the item parameter estimates were not affected much. The item parameter estimates from each hypothesized model were relatively similar to one another, so here we only present them for Model A (see Table 3). In Table 3, the first two columns

Table 3. Item Parameter Estimates for Model A.

Items	g_j (*SE*)	s_j (*SE*)	$\frac{(1-s_j)/s_j}{g_j/(1-g_j)}$
1	0.336 (0.053)	0.301 (0.026)	4.584
2	0.454 (0.053)	0.223 (0.023)	4.204
3	0.041 (0.017)	0.317 (0.030)	51.024
4	0.074 (0.024)	0.115 (0.021)	96.153
5	0.074 (0.026)	0.035 (0.012)	345.981
6	0.015 (0.010)	0.410 (0.032)	92.561
7	0.084 (0.025)	0.060 (0.016)	172.161
8	0.113 (0.026)	0.103 (0.021)	68.292
9	0.507 (0.052)	0.149 (0.022)	5.534
10	0.239 (0.040)	0.141 (0.038)	19.302
11	0.004 (0.004)	0.529 (0.080)	207.728
12	0.004 (0.003)	0.450 (0.084)	349.669
13	0.003 (0.003)	0.659 (0.076)	160.467
14	0.005 (0.004)	0.371 (0.082)	358.650
15	0.003 (0.002)	0.371 (0.085)	670.044
16	0.049 (0.014)	0.789 (0.046)	5.172
17	0.030 (0.009)	0.607 (0.077)	20.701
18	0.334 (0.031)	0.480 (0.047)	2.169
19	0.011 (0.009)	0.695 (0.041)	40.384
20	0.050 (0.028)	0.471 (0.029)	21.242
21	0.011 (0.011)	0.675 (0.026)	44.059
22	0.017 (0.016)	0.508 (0.029)	55.969
23	0.016 (0.007)	0.870 (0.025)	9.097

provide estimates of item parameters g_j and s_j, respectively. The third column provides estimates of the diagnostic quality of each item. This estimate was obtained as

$$\frac{(1 - s_j)/s_j}{g_j/(1 - g_j)},$$

which is the odds ratio between responding positively conditional on $\xi_{ij} = 1$ and responding positively conditional on $\xi_{ij} = 0$. The item with the largest odds ratio is considered to be the most diagnostic in terms of distinguishing between the two latent classes, where the two latent classes are defined as one in which members have mastered all skills required by that item (i.e., $\xi_{ij} = 1$) and a second in which members have not mastered at least one skill required by that item (i.e., $\xi_{ij} = 0$).

As can be seen in Table 3, no item had an odds ratio lower than 2 and most were relatively large. Moreover, monotonicity, which is $1 - s > g$ held for all items (Junker & Sijtsma, 2001). This is interpreted to mean that the FOC items were effective at discriminating between examinees in the two latent classes. It also indicates that the entries in the Q-matrix sufficiently identified the requisite skills needed for the items on the test. The high odds ratios, furthermore, were due primarily to relatively small guessing parameters. This follows from the fact that all 23 items from the FOC tests were short answer items and, therefore, the likelihood of guessing was lower.

Table 4. Frequency and Proportion of Mastery of Each Skill at Each Time Point.

Skills	Time Point 1	Time Point 2	Time Point 3	Time Point 4
Number & Operation	51 (46.79%)	78 (71.56%)	85 (77.98%)	109 (100.00%)
Measurement	55 (50.46%)	77 (70.64%)	74 (67.89%)	104 (95.41%)
Problem Solving	2 (1.83%)	2 (1.83%)	3 (2.75%)	36 (30.28%)
Representation	31 (28.44%)	18 (16.51%)	27 (24.78%)	103 (94.50%)

Mastery Probabilities Across Time Points

Table 4 provides frequencies and proportions of examinees who mastered each of the cognitive skills at each time point. The mastery proportions of Number & Operation and Measurement increased following first KK instruction (which occurred between Test Wave 1 and 2), but those for Problem Solving and Representation did not. All four skills improved substantially at Test Wave 4, following FOC instruction: 100% mastered Number & Operation, and approximately 95% mastered Measurement and Representation at the fourth administration. The proportion mastering Problem Solving, the most difficult of the four skills, increased to 30.28% from an initial value of 1.83%. The pattern of change in mastery proportions across the four time points suggested FOC instruction markedly improved the mastery of the four cognitive skills. In contrast, regular instruction (between Test Wave 2 and 3) had little effect on the four skills. It also appears that KK instruction had some effect on Number & Operation and Measurement, but little effect on Problem Solving and Representation.

The patterns of mastery status across the four test administrations provide an indication of the effectiveness of the FOC instruction. Table 5 presents frequencies and probabilities of the mastery patterns observed in the FOC data for each of the four test waves. A pattern indicating mastery for all four skills at a single test administration is represented in Table 5 as (1, 1, 1, 1). Similarly, a pattern indicating nonmastery of the first two skills and mastery of the second two skills is represented as (0, 0, 1, 1). The four skills can potentially generate 2^4 mastery status patterns where a pattern is indicated by a 0 for nonmastery and a 1 for mastery. In real data, however, differential difficulties of particular skills can sometimes result in some patterns not being observed, as was the case in this example. In Table 5, patterns (0, 0, 1, 0), (0, 0, 1, 1), (0, 1, 1, 0), (1, 0, 1, 0), and (0, 1, 1, 1) are missing. Furthermore, only 1 or 2 examinees exhibited patterns (1, 0, 1, 1) or (1, 1, 1, 0).

Examination of the observed patterns clearly indicates that no examinee mastered the third skill, Problem Solving, without first mastering the other three skills. The mastery patterns suggest that mastery of problem solving does not occur in the absence of mastery of the other skills. From Table 4, it seems clear that Problem Solving was the most difficult to master, followed by Representation, Number & Operation, and finally Measurement. As a result, the patterns (1, 1, 0, 0), (1, 1, 0, 1), (1, 0, 0, 0), and (0, 0, 0, 0) were the most frequently observed over the measures.

Table 5. Frequencies and Proportions of Mastery Status Patterns Over Four Skills at Each Time Point.

Pattern	Time Point 1	Time Point 2	Time Point 3	Time Point 4
(0, 0, 0, 0)	37 (33.94%)	12 (11.01%)	12 (11.01%)	0 (0%)
(0, 0, 0, 1)	4 (3.67%)	0 (0%)	0 (0%)	0 (0%)
(0, 1, 0, 0)	7 (6.42%)	18 (16.51%)	5 (4.59%)	0 (0%)
(0, 1, 0, 1)	10 (9.17%)	1 (0.92%)	7 (6.42%)	0 (0%)
(1, 0, 0, 0)	13 (11.93%)	17 (15.60%)	21 (19.27%)	0 (0%)
(1, 0, 0, 1)	0 (0%)	3 (2.75%)	2 (1.83%)	4 (3.67%)
(1, 0, 1, 1)	0 (0%)	0 (0%)	0 (0%)	1 (.92%)
(1, 1, 0, 0)	20 (18.35%)	44 (40.37%)	44 (40.37%)	5 (4.59%)
(1, 1, 0, 1)	16 (14.68%)	12 (11.01%)	15 (13.76%)	64 (58.72%)
(1, 1, 1, 0)	1 (0.92%)	0 (0%)	0 (0%)	1 (.92%)
(1, 1, 1, 1)	1 (0.92%)	2 (1.83%)	3 (2.75%)	34 (31.19%)

Following FOC instruction, mastery of skills improved such that lower frequencies were observed for patterns with zero or one skill mastered, such as (0, 0, 0, 0), (1, 0, 0, 0), (0, 1, 0, 0), and (0, 0, 0, 1). In addition, following instruction, patterns with more skills mastered, such as (1, 1, 1, 1), increased in frequency. These results provide support for the effectiveness of the FOC instruction.

Model Comparisons

Models B and C were nested in Model A, and Model D was nested in Models A, B, C; therefore, a likelihood ratio test could be used to compare the fit of each model to the data. However, inspection of the posterior mean of the deviances from WinBUGS indicated that those for the augmented models were slightly larger than for the compact models. As an example, Model A had a higher deviance (Deviance_A = 6,590) than Model D (Deviance_D = 6,580). This situation is somewhat unusual. Spiegelhalter (2006) suggested that this outcome occurs because of the addition of a covariate that is either uncorrelated with or less correlated with the dependent variable. As a result, the posterior mean deviance should not be used as a measure of the fit to the data. In this study, therefore, we used the Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002) instead of likelihood ratio tests.

DIC is composed of a Bayesian measure of fit called the posterior mean of deviance $\bar{D}$ and a penalty for model complexity, p_D:

$$\text{DIC} = \bar{D} + p_D, \tag{8}$$

where $p_D = \bar{D} - D(\hat{\theta})$, and $D(\hat{\theta})$ is the deviance of the posterior model. The model with the smallest DIC is selected as the model that would best predict a replicate data set of the same structure (Spiegelhalter et al., 2002).

Table 6. Model Comparison.

Assumptions	Models	$\bar{D}$	$D(\hat{\theta})$	p_D	DIC
Homogeneity	Model A	6,590	6,356	234	6,824
	Model B	6,589	6,357	232	6,821
	Model C	6,620	6,395	225	6,845
	Model D	6,580	6,331	249	6,829
Gender heterogeneity	Model A	6,593	6,350	243	6,836
	Model B	6,589	6,333	256	6,845
	Model C	6,598	6,349	249	6,847
	Model D	6,584	6,325	259	6,843
Ability heterogeneity	Model A	6,594	6,362	232	6,826
	Model B	6,593	6,373	220	6,813
	Model C	6,604	6,365	239	6,843
	Model D	6,589	6,336	253	6,842

To answer the first research question, we compared DICs from Models A, B, C, and D (see Table 6). Model B was a better fit than the other three models, although it was not substantially different than Model A. Recall that Model B was the hypothesis that FOC instruction affected the transition differently from either the KK instruction or the regular classroom instruction, and both KK instruction and regular instruction affected the transition similarly. In other words, the transition probabilities were equivalent before FOC instruction but changed following FOC instruction.

This transition pattern actually was also observed in the transition probability matrices based on Model A (see Table 7). The observed transition probabilities for each skill from Test Wave 1 to 2 and from Test Wave 2 to 3 were quite similar, but clearly different than those from Test Wave 3 to 4. From Test Wave 3 to 4, the transition probabilities from nonmastery to mastery status (shown in bold) became substantially larger for each skill. This was particularly noticeable for the harder skills of Problem Solving and Representation, suggesting that the FOC instruction had stronger effects on fractions computation than the other two types of instruction.

To answer the second research question about population homogeneity or heterogeneity, that is, whether different groups of examinees had different transition probabilities, we compared models A, B, C, D under the homogeneity assumption to that under the heterogeneity assumption based on gender or initial ability level.

The DIC values of Models A, B, C, D based on gender heterogeneity were all higher than those based on homogeneity assumption, suggesting no gender difference on the pattern of change across four test administrations. Table 8 also indicates that the transition probability matrices for different gender were quite similar.

The comparison between ability heterogeneity assumption and homogeneity assumption found: (1) under either assumption, Model B appears best fit compared with other models; (2) Model B with ability heterogeneity assumption had lower DIC than that with the homogeneity assumption. Under both assumptions, results

Table 7. Transition Probability Matrices for Model A.

Skills	$t_1 \rightarrow t_2$	$t_2 \rightarrow t_3$	$t_3 \rightarrow t_4$
Skill 1: Number & Operation	$\begin{pmatrix} .55 & .45 \\ .03 & .97 \end{pmatrix}$	$\begin{pmatrix} .66 & .34 \\ .06 & .94 \end{pmatrix}$	$\begin{pmatrix} .05 & .95 \\ .01 & .99 \end{pmatrix}$
Skill 2: Measurement	$\begin{pmatrix} .71 & .29 \\ .03 & .97 \end{pmatrix}$	$\begin{pmatrix} .82 & .18 \\ .05 & .95 \end{pmatrix}$	$\begin{pmatrix} .12 & .88 \\ .03 & .97 \end{pmatrix}$
Skill 3: Problem Solving	$\begin{pmatrix} .93 & .07 \\ .87 & .13 \end{pmatrix}$	$\begin{pmatrix} .90 & .10 \\ .54 & .46 \end{pmatrix}$	$\begin{pmatrix} .65 & .35 \\ .51 & .49 \end{pmatrix}$
Skill 4: Representation	$\begin{pmatrix} .83 & .17 \\ .54 & .46 \end{pmatrix}$	$\begin{pmatrix} .74 & .26 \\ .12 & .88 \end{pmatrix}$	$\begin{pmatrix} .07 & .93 \\ .12 & .88 \end{pmatrix}$

Table 8. Transition Probability Matrices for Males and Females.

		Transitions		
Gender	Skills	$t_1 \rightarrow t_2$	$t_2 \rightarrow t_3$	$t_3 \rightarrow t_4$
Males	Number & Operation	$\begin{pmatrix} .49 & .51 \\ .07 & .93 \end{pmatrix}$	$\begin{pmatrix} .51 & .49 \\ .12 & .88 \end{pmatrix}$	$\begin{pmatrix} .12 & .88 \\ .03 & .97 \end{pmatrix}$
	Measurement	$\begin{pmatrix} .67 & .33 \\ .13 & .87 \end{pmatrix}$	$\begin{pmatrix} .54 & .46 \\ .13 & .87 \end{pmatrix}$	$\begin{pmatrix} .11 & .89 \\ .04 & .96 \end{pmatrix}$
	Problem Solving	$\begin{pmatrix} .85 & .15 \\ .82 & .18 \end{pmatrix}$	$\begin{pmatrix} .88 & .12 \\ .60 & .40 \end{pmatrix}$	$\begin{pmatrix} .63 & .37 \\ .52 & .48 \end{pmatrix}$
	Representation	$\begin{pmatrix} .73 & .27 \\ .62 & .38 \end{pmatrix}$	$\begin{pmatrix} .55 & .45 \\ .48 & .52 \end{pmatrix}$	$\begin{pmatrix} .15 & .85 \\ .17 & .83 \end{pmatrix}$
Females	Number & Operation	$\begin{pmatrix} .55 & .45 \\ .06 & .94 \end{pmatrix}$	$\begin{pmatrix} .55 & .45 \\ .09 & .91 \end{pmatrix}$	$\begin{pmatrix} .08 & .92 \\ .03 & .97 \end{pmatrix}$
	Measurement	$\begin{pmatrix} .60 & .40 \\ .06 & .94 \end{pmatrix}$	$\begin{pmatrix} .77 & .23 \\ .12 & .88 \end{pmatrix}$	$\begin{pmatrix} .15 & .85 \\ .06 & .94 \end{pmatrix}$
	Problem Solving	$\begin{pmatrix} .89 & .11 \\ .84 & .16 \end{pmatrix}$	$\begin{pmatrix} .84 & .16 \\ .54 & .46 \end{pmatrix}$	$\begin{pmatrix} .64 & .36 \\ .51 & .49 \end{pmatrix}$
	Representation	$\begin{pmatrix} .78 & .22 \\ .63 & .37 \end{pmatrix}$	$\begin{pmatrix} .77 & .23 \\ .55 & .45 \end{pmatrix}$	$\begin{pmatrix} .09 & .91 \\ .21 & .79 \end{pmatrix}$

suggested that effects of the FOC instruction were different than that of either the KK or regular instruction, although some differences were observed on the transition probabilities between the two latent groups. The most substantial difference was that the higher ability group had higher transition probabilities from nonmastery to mastery status from Test Wave 1 to 2 in Number & Operation (.76 vs. .38) and in Measurement (.72 vs. .46; see Table 9). This suggests that the high-ability latent group improved their learning on Number & Operation and on Measurement at the very beginning, following the KK instruction.

Table 9. Transition Probability Matrices for Two Latent Classes.

Latent class	Skills	Transitions $t_1 \rightarrow t_2$	$t_2 \rightarrow t_3$	$t_3 \rightarrow t_4$
Low-ability group	Number & Operation	$\begin{pmatrix} .62 & .38 \\ .10 & .90 \end{pmatrix}$	$\begin{pmatrix} .57 & .43 \\ .16 & .84 \end{pmatrix}$	$\begin{pmatrix} .06 & .94 \\ .03 & .97 \end{pmatrix}$
	Measurement	$\begin{pmatrix} .54 & .46 \\ .35 & .65 \end{pmatrix}$	$\begin{pmatrix} .67 & .33 \\ .30 & .70 \end{pmatrix}$	$\begin{pmatrix} .11 & .89 \\ .09 & .91 \end{pmatrix}$
	Problem Solving	$\begin{pmatrix} .83 & .17 \\ .81 & .19 \end{pmatrix}$	$\begin{pmatrix} .86 & .14 \\ .56 & .44 \end{pmatrix}$	$\begin{pmatrix} .61 & .39 \\ .52 & .48 \end{pmatrix}$
	Representation	$\begin{pmatrix} .76 & .24 \\ .71 & .29 \end{pmatrix}$	$\begin{pmatrix} .51 & .49 \\ .49 & .51 \end{pmatrix}$	$\begin{pmatrix} .15 & .85 \\ .15 & .85 \end{pmatrix}$
High-ability group	Number & Operation	$\begin{pmatrix} .24 & .76 \\ .07 & .93 \end{pmatrix}$	$\begin{pmatrix} .42 & .58 \\ .07 & .93 \end{pmatrix}$	$\begin{pmatrix} .18 & .82 \\ .02 & .98 \end{pmatrix}$
	Measurement	$\begin{pmatrix} .28 & .72 \\ .11 & .89 \end{pmatrix}$	$\begin{pmatrix} .41 & .59 \\ .08 & .92 \end{pmatrix}$	$\begin{pmatrix} .17 & .83 \\ .02 & .98 \end{pmatrix}$
	Problem Solving	$\begin{pmatrix} .90 & .10 \\ .80 & .20 \end{pmatrix}$	$\begin{pmatrix} .85 & .15 \\ .55 & .45 \end{pmatrix}$	$\begin{pmatrix} .68 & .32 \\ .49 & .51 \end{pmatrix}$
	Representation	$\begin{pmatrix} .57 & .43 \\ .51 & .49 \end{pmatrix}$	$\begin{pmatrix} .81 & .19 \\ .54 & .46 \end{pmatrix}$	$\begin{pmatrix} .08 & .92 \\ .21 & .79 \end{pmatrix}$

Table 10. Mastery Proportions in Each Skill at Each Time Point.

Skills	Latent classes	Time Point 1	Time Point 2	Time Point 3	Time Point 4
Number & Operation	Low-ability group	17.70%	61.30%	69.40%	100.00%
	High-ability group	85.10%	85.10%	89.40%	100.00%
Measurement	Low-ability group	21.00%	61.30%	51.60%	91.90%
	High-ability group	89.40%	83.00%	89.40%	100.00%
Problem Solving	Low-ability group	0.00%	1.60%	3.20%	33.90%
	High-ability group	4.30%	2.10%	2.10%	31.90%
Representation	Low-ability group	17.70%	14.50%	24.20%	93.50%
	High-ability group	42.60%	19.10%	25.50%	95.70%

Table 10 and Figures 1 to 4 further illustrate the differences in the mastery probability profiles across measures for the two latent classes. For Problem Solving and Representation, both latent classes had similar patterns of change in mastery proportions. Furthermore, as can be seen in Figures 3 and 4, these proportions did not increase much until after the FOC instruction.

The main differences between the two classes still occurred in Number & Operation and Measurement (see Figures 1 and 2): The mastery proportions of low-ability group increased sharply compared with those of high-ability group. This is

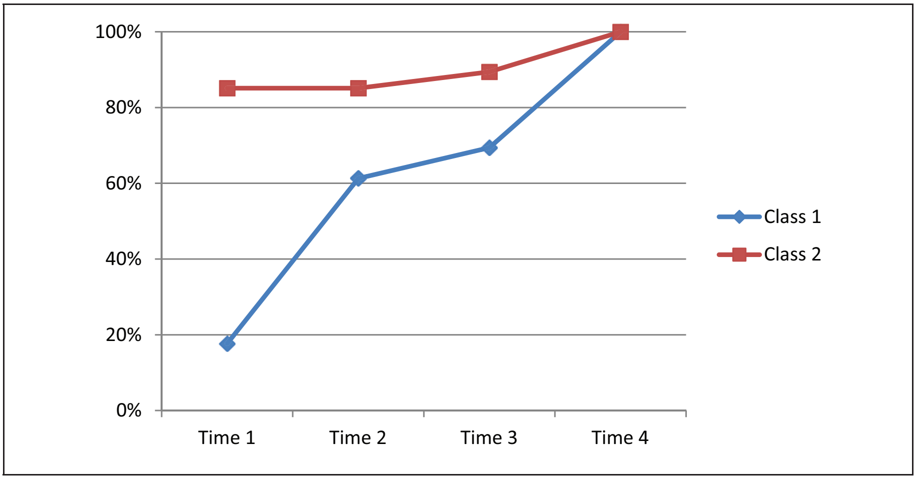

Figure 1. Profile of mastery proportions over four time points for Number & Operation.

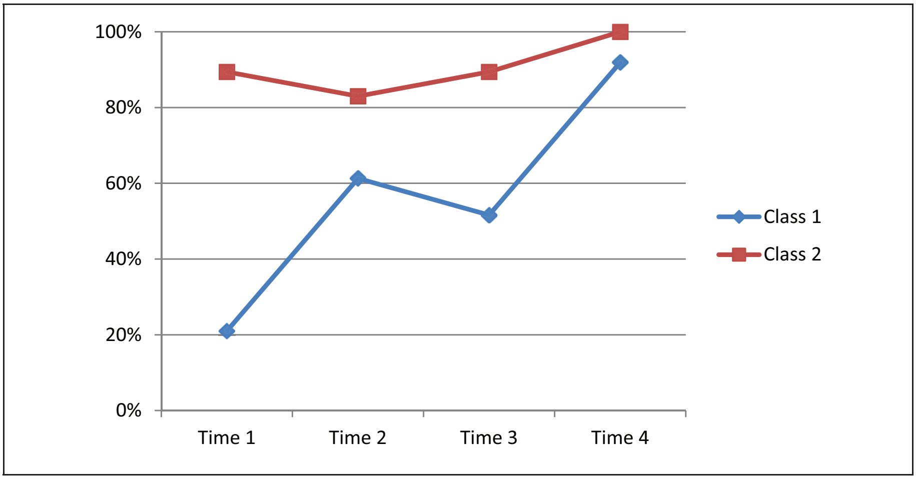

Figure 2. Profile of mastery proportions over four time points for Measurement.

reasonable given that Class 2, the higher ability group, had higher initial mastery probabilities in these two skills and, therefore, little room for improvement compared with the lower ability group.

As indicated above, the higher ability class had a higher transition probability from nonmastery to mastery in Number & Operation and in Measurement from Time 1 to Time 2 (see Table 9). One might then ask why was the increase in mastery proportion lower as shown in Figures 1 and 2? Actually, these two facts are not

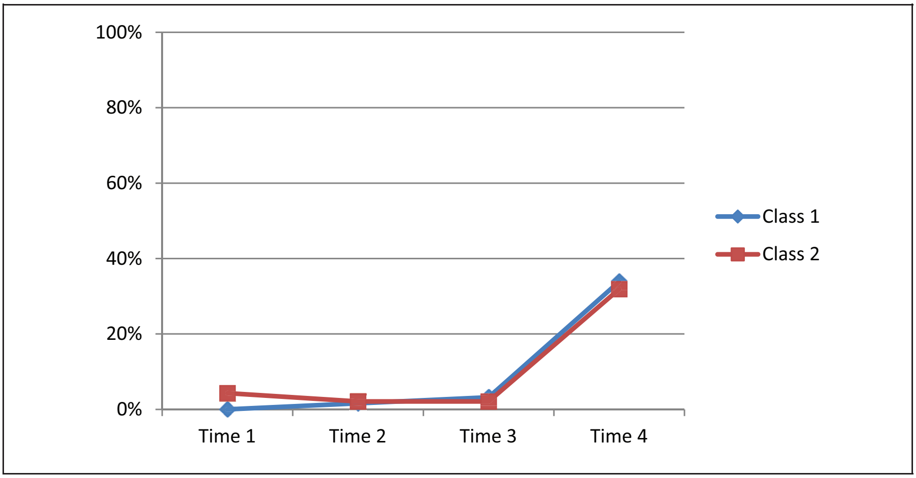

Figure 3. Profile of mastery proportions over four time points for Problem Solving.

necessarily incompatible. The change in mastery proportions is not only dependent on the transition probability from nonmastery to mastery status, but it is also dependent on the transition probability from mastery to mastery status as well as the proportions of nonmastery and mastery status at the previous time point. Consider the transition from Time 1 to Time 2 as example. This can be expressed as

$$p_m^2 - p_m^1 = (p_{m|n}^{21} * p_n^1 + p_{m|m}^{21} * p_m^1) - p_m^1. \tag{9}$$

Thus, even though the higher ability class had higher transition probabilities from nonmastery to mastery status ($p_{m|n}^{21}$) in the two skills from Time 1 to Time 2, it also had lower nonmastery proportions (p_n^1) and higher mastery proportions (p_m^1) at Time 1; therefore, its increase in mastery proportion ($p_m^2 - p_m^1$) from Time 1 to Time 2 could be lower than that of lower ability class.

In essence, the main ability difference between the two latent classes only occurred in two easier skills, in which the lower ability class had significantly lower initial mastery level. As is evident in the graphs in Figures 1 and 2, however, the lower ability class actually learned as well as the higher ability class following the FOC the instruction. They also had higher rates of increase on the Number & Operation and Measurement skills because they had more room to improve before reaching mastery. In contrast, the two latent groups did not show differences in either initial or final mastery levels for the two harder skills. The similarity in growth pattern of the two latent classes (as shown in Figures 3 and 4) indicates they both did not change much until after the FOC component of the EAI intervention. This indicated that both latent groups benefited from the FOC intervention. Overall, the results suggest that the EAI intervention was at least equally and possibly even more effective for members of the lower ability latent class. The fact that the nine LD

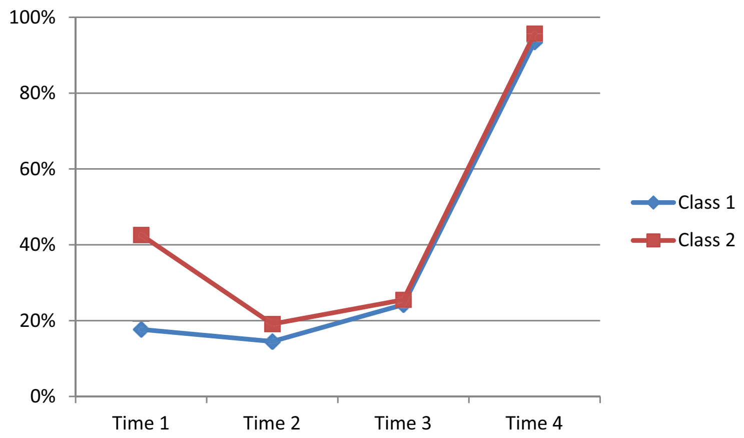

Figure 4. Profile of mastery proportions over four time points for Representation.

students were all classified into the low-ability latent class might have contributed to the heterogeneity in the transition between the latent classes.

Discussion

LTA was initially developed to study stage sequential change over time in particular types of latent variables called dynamic latent variables (Collins & Wugalter, 1992). Applications to educational problems have been lacking in part because of the categorical nature of the latent variables considered. In this article, we extend the LTA to an LTA-DINA by including a DINA cognitive diagnostic measurement model and demonstrate its use in the analysis of effects of an instructional intervention. The combination of these two models provides several advantages compared with other growth modeling methods.

The use of the DINA model permits one to make relatively fine-grained assessments of students' responses to test items. Including this model in an analysis of change provides a means of analyzing different latent components to determine where the change might occur for different groups of examinees. Individuals with the same raw scores or even the same ability estimates can easily have different skills mastery profiles. Similarly, groups may differ in terms of raw scores or ability estimates, but may have common patterns of growth for some skills. This was the case for the two latent classes in the example. The transition probabilities estimated by the LTA-DINA were similar for the two latent classes in some skills, even though the classes differed in ability.

Differences in proportions of mastery of different cognitive skills also seemed to reflect differential difficulties in learning of the four cognitive skills. For example, Problem Solving was the most challenging of the four cognitive skills, even for the high-ability members of Class 2.

The DINA model is capable of diagnosing individual- or group-level weaknesses and strengths in terms of mastery and nonmastery status on a set of latent cognitive skills. It is also capable of diagnosing whether each item was an effective measure of the cognitive skills measured on the test. Furthermore, information provided by the model can be used to assess whether or not the Q-matrix was appropriately defined. Higher guessing parameters or higher slipping parameters serve to lower the diagnostic quantity index as indicated in the odds ratio

$$\frac{(1-s_j)/s_j}{g_j/(1-g_j)},$$

or nonmonotonicity $(1-s<=g)$. Deleting or revising items of low diagnostic quality or non-monotonicity, or refining the Q-matrix may be used to help improve the quality of information obtained from the test.

LTA-DINA also provides a transition probability matrix instead of a single quantity to index change. LTA-DINA permits determining not only the proportions moving from nonmastery to mastery status but also the proportions moving from mastery to nonmastery, staying at nonmastery or staying at mastery statuses. Successful instruction should have (1) high proportions of students moving from nonmastery to mastery status and (2) low proportions moving from mastery to nonmastery status. This was most evident in the transition matrices for Number & Operations, Measurement, and Representation. For Problem Solving, the transitions from nonmastery to mastery status were not as large as for the other three cognitive skills, and the transition from mastery to nonmastery status were much larger than for the other three cognitive skills. The reason for this appears to be in the differential difficulty of the four cognitive skills. Problem Solving was clearly more difficult and, therefore, less likely to show change in mastery status for this sample. This specific information was evident in the proportions mastering Problem Solving in the total sample and in the latent classes.

An important component of this use of an LTA-DINA model was the capability for testing the likelihood of different assumptions about the nature of change at the level of cognitive skills. That is, it was possible to test different hypotheses about the form of the transition probabilities for specific groups of examinees. As the models in this example were all nested, it was possible to use a likelihood ratio test to determine best fit to the data. It should be possible to test other, non-nested models, possibly using measures of model fit such as AIC (Akaike, 1974) or BIC (Schwarz, 1978). Including the DINA model in an LTA-DINA provides an analysis of changes that focuses on patterns of mastery of each of the cognitive skills measured by the test. As such, the LTA-DINA model provides richer information about instructional effects at both an individual and group level.

The sample size in the example was small, quite likely affecting the precision of estimates, although the diagnostic quantity indices indicated the items were effective for the DINA model as specified in the Q-matrix. The independence assumption among skills in the example is actually tenuous as skills are always correlated to some extent. Furthermore, examinees in this study took the same test four times with the likely result that some form of memory effects may have been present.

MCMC has been shown to be useful for estimation of complex IRT models (Patz & Junker, 1999). It is also possible to estimate the LTA-DINA model using maximum likelihood estimation. The commercially available software Mplus (Muthén & Muthén, 1998-2012) does provide an option to use maximum likelihood estimation and also includes an option to estimate LTA. Templin (2006) has shown how the DINA model can be estimated using this software. In the future, exploring how to implement LTA with DINA as measurement model in Mplus will be potentially possible and valuable.

Declaration of Conflicting Interests

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

Funding

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

References

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

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

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 3-20). Madison: University of Wisconsin Press.

Boscardin, C. K., Muthén, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology,100*, 192-208.

Bottge, B. A., Heinrichs, M., Chan, S. Y., Mehta, Z. D., & Watson, E. (2003). Effects of video-based and applied problems on the procedural Math skills of average- and low-achieving adolescents. *Journal of Special Education Technology*, *18*(2), 5-22.

Bottge, B. A., Rueda, E., Serlin, R. C., Hung, Y. H., & Kwon, J. M. (2007). Shrinking achievement difference with anchored math problems: Challenges and possibilities. *Journal of Special Education,41*, 31-49.

Cohen, A. S. (2006). *Analysis of cognitive strategies used in enhanced anchored instruction in a technology rich learning environment* (Technical report).

Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research, 27*, 131-157.

Compton, D. L., Fuchs, D., Fuchs, L. S., Elleman, A. M., & Gilbert, J. K. (2008). Tracking children who fly below the radar: Latent transition modeling of students with late-emerging reading disability. *Learning and Individual Differences, 18*, 329-337.

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

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 361-390). Hillsdale, NJ: Lawrence Erlbaum.

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula.

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

Hambleton, R. K., Rogers, H. J., & Swaminathan, H. (1991). *Fundamentals of item response theory*. Thousand Oaks, CA: Sage.

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

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

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence-Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

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

Marcoulides, G. A., Gottfried, A. E., Gottfried, A. W., & Oliver, P. (2008). Latent transition analysis of academic intrinsic motivation from childhood through adolescence. *Educational Research and Evaluation, 14*, 411-427.

Muthén, L., & Muthén, B. (1998-2012). *Mplus user's guide*. Version *7*. Los Angeles, CA: Muthén & Muthén.

Nylund, K. (2007). *Latent transition analysis: Modeling extensions and an application to peer victimization* (Doctoral dissertation). University of California, Los Angeles.

Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics, 24*, 146-178.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271-282.

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

Spiegelhalter, D. J. (2006, February). *Some DIC slides*. Retrieved from http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/dicpage.shtml#q17

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*, 583-616.

Spiegelhalter, D. J., Thomas, A., & Best, N. (2003). WinBUGS (version 1.4) [computer program]. Cambridge, England: MRC Biostatistics Unit, Institute of Public Health.

Templin, J. (2006). CDM: Cognitive diagnosis modeling with Mplus [Computer software]. (Author owned: Retrievable on request)

Trentacosta, C. J., Criss, M. M., Shaw, D. S., Lacourse, E., Hyde, L. W., & Dishion, T. J. (2011). Antecedents and outcomes of joint trajectories of mother-son conflict and warmth during middle childhood and adolescence. *Child Development, 82*, 1676-1690.