Article

Educational and Psychological Measurement
2017, Vol. 77(3) 369–388

DOI: 10.1177/0013164416659314
journals.sagepub.com/home/epm

Assessing Change in Latent Skills Across Time With Longitudinal Cognitive Diagnosis Modeling: An Evaluation of Model Performance

Yasemin Kaya[1] and Walter L. Leite[2]

Abstract

Cognitive diagnosis models are diagnostic models used to classify respondents into homogenous groups based on multiple categorical latent variables representing the measured cognitive attributes. This study aims to present longitudinal models for cognitive diagnosis modeling, which can be applied to repeated measurements in order to monitor attribute stability of individuals and to account for respondent dependence. Models based on combining latent transition analysis modeling and the DINA and DINO cognitive diagnosis models were developed and then evaluated through a Monte Carlo simulation study. The study results indicate that the proposed models provide adequate convergence and correct classification rates.

Keywords

cognitive diagnosis models, latent transition analysis, diagnostic classification models, latent class analysis, longitudinal assessment

Cognitive diagnosis models (CDMs), alternatively called as diagnostic classification models (e.g., Tatsuoka, 1983; Templin & Henson, 2006), are models used to classify

[1]Ataturk University, Erzurum, Turkey
[2]University of Florida, Gainesville, FL, USA

Corresponding Author:
Yasemin Kaya, Kazim Karabekir Egitim Fakultesi, Ataturk Universitesi, Yoncalik Kampusu, Yakutiye, Erzurum, 25240, Turkey.
E-mail: yasemin.kaya@atauni.edu.tr

respondents into homogenous groups based on multiple categorical latent variables representing the measured cognitive attributes (Rupp, Templin, & Henson, 2010). CDMs are a group of models used in development and implementation of diagnostic assessments. In other words, CDMs provide classification of respondents into mastery profiles of measured attributes that respondents need to acquire in order to respond to the assessments correctly (Rupp et al., 2010).

CDMs are commonly used for analyzing observed categorical response data collected through diagnostic assessments in education and social sciences. In education, there are two domains where CDMs can be meaningfully applied (Rupp et al., 2010): (a) cognitive diagnosis assessments where CDMs are used to provide exploratory information for instruction and learning and (b) standards-based assessments and classifications. CDMs have recently received increasing attention of researchers in education because of their potential use for diagnosing individual mental states and weaknesses in learning skills. However, the majority applications of CDMs have focused on analyzing a single data collection, without paying attention to the subsequent stages of students' mastered skills or the development of students over time.

Latent transition analysis (LTA) is a longitudinal application of the latent class analysis (LCA) model, which aims to identify whether there is a change between latent classes over time (Collins & Lanza, 2010). While LCA is performed on a set of variables collected at one time to identify latent classes that represent the response patterns in the data, LTA is applied to measurements repeated over time to follow the move between the identified latent classes (Collins & Lanza, 2010).

To address the scarcity of methods for cognitive diagnosis modeling for multiple waves of measurement, this study presents a longitudinal CDM that can be applied to repeated measurements in order to monitor transition of individuals between cognitive classes and to observe gain of necessary attributes throughout time in a curricular setting. We present and evaluate models based on extending the LTA model to cognitive diagnosis in order to explain changes of the attribute profiles of the students over time. Through the development of a model combining the LTA and CDMs, we address within-individual change in follow-up measurements of learning, as well as moving and static groups in between cognitive classes.

Cognitive Diagnosis Models

Cognitive diagnostic assessments in education have been developed to assess cognitive processes behind learning and to provide formative information about examinees. In more technical terms, educational cognitive diagnosis is the assessment that is used to identify the cognitive attributes that an examinee needs to master in order to achieve a particular domain, and to classify such examinees into diagnostic groups based on the attributes they possess (Rupp et al., 2010). Cognitive diagnosis provides a detailed profile of mastered skills or attributes for an examinee rather than giving a general score. Furthermore, each respondent is classified into one cognitive class in which all the individuals will have the same pattern of attribute mastery. Before using cognitive diagnostic assessments, it is first necessary to identify the latent variables that are measured by a particular assessment. These latent variables have received diverse labels in the literature, such as skills, attributes, and latent traits. In the current study, we used the term *attributes* to refer these latent variables.

A Q-matrix is an essential piece of cognitive diagnostic models because it defines the connection between items and attributes. More specifically, a Q-matrix is an item by attribute table that specifies whether each attribute is measured by each item through binary entries of 1s or 0s (Rupp et al., 2010). Let q_{ia} symbolize an element of a Q-matrix in which i and a are the number of rows and columns, respectively, representing the items and the attributes. A value of 1 represents that item i measures attribute a. The creation of a Q-matrix is usually made by educational experts based on theoretical judgments, but methods for automated construction of Q-matrices have been proposed (e.g., Barnes, 2011). Any misspecification of the Q-matrix patterns can cause error in the classification of examinees to diagnostic groups. To minimize the classification error caused by false specification of the Q-matrix and to improve and evaluate Q-matrix specification methods, several studies have been conducted (e.g., Chiu, 2013; Close, 2012; J. Liu, Xu, & Ying, 2012; Xu, 2013) about the Q-matrix specification and validation procedures and the performance of the CDMs under different Q-matrix specification conditions. Because evaluation of Q-matrix specifications are not the scope our research, in this study, it is assumed that the Q-matrix has been correctly specified.

Different CDMs exist in the literature, which can be classified as compensatory or noncompensatory. Noncompensatory CDMs assume that the lack of one latent variable, which is an attribute in our context, cannot be compensated by another latent variable. Common examples of noncompensatory models are the deterministic-input noisy-and-gate (DINA); the noisy-input deterministic-and-gate; and noncompensatory reparameterized unified model (de la Torre & Douglas, 2004; DiBello, Stout, & Roussos, 1995; Haertel, 1989; Hartz, 2002; Junker & Sijtsma, 2001). On the other hand, compensatory CDMs presume that lack of one latent variable can be compensated by the presence of another latent variable. Examples of commonly used compensatory models are the deterministic input, noisy-or-gate model (DINO); the noisy input, deterministic-or-gate model; and the compensatory reparameterized unified model (Hartz, 2002; Templin, 2004; Templin & Henson, 2006). In this section, a brief review of the CDMs mentioned above will be provided. Our study particularly focuses on DINA and DINO models, because of their common use in current methodological and empirical studies (e.g., de la Torre, 2008; Henson & Douglas, 2005; von Davier, 2014).

The DINA Model

The DINA model (de la Torre & Douglas, 2004; Haertel, 1989; Junker & Sijtsma, 2001; Macready & Dayton, 1977) is a noncompensatory CDM that assumes that lack of one attribute cannot be reimbursed by the existence of another attribute. DINA works with a conjunctive condensation rule, which means that in order to have a high probability of responding an item correctly, an individual needs to master all the attributes required by that item. The main limitation of the DINA model is that it does not make a distinction between respondents who did not master only one or more than one attribute.

The DINA model estimates the probability of a correct response to item i for all the respondents in latent class c as follows:

$$\pi_{ic} = P(X_{ic} = 1|\xi_{ic}) = (1 - s_i)^{\xi_{ic}} g_i^{(1-\xi_{ic})} \tag{1}$$

where π_{ic} is the probability of correct response, x_{ic} is the observed response, ξ_{ic} is the attribute mastery indicator, and s_i and g_i are, respectively, the slipping and the guessing parameters. The slipping parameter, s_i, is defined as the probability of responding an item incorrectly for a respondent who has mastered all the required attribute:

$$s_i = P(X_{ic} = 0|\xi_{ic} = 1) \tag{2}$$

The guessing parameter, g_i, is the probability of responding an item correctly for a respondent who has not mastered at least one required attribute:

$$g_i = P(X_{ic} = 1|\xi_{ic} = 0) \tag{3}$$

If a respondent masters all the required attributes, $\xi_{ic} = 1$, the probability of responding the item correctly is equal to the probability of not slipping for the item, $1 - s_i$. On the other hand, if the respondent fails to master at least one of the required attributes, $\xi_{ic} = 0$, the probability of responding the item correctly drops to the probability of guessing for the item, g_i. The DINA model order-constrains the slipping and guessing parameters: $1 - s_i$ is assumed to be greater than g_i; thus, the probability of responding an item correctly is guaranteed to be always higher for the respondents who mastered all the measured attributes than the respondents who lacked at least one of the measured attributes, regardless of the magnitudes of slipping and guessing parameters (Rupp et al., 2010).

The attribute mastery indicator is formulated as follows:

$$\xi_{ic} = \prod_{a=1}^{A} \alpha_{ca}^{q_{ia}} \tag{4}$$

where A is the total number of attributes measured, and q_{ia} indicates whether attribute a is measured by item i. The possible values that q_{ia} takes are 0 or 1. The other indicator α_{ca} identifies whether the respondent in latent class c mastered attribute a, which takes values of 0 or 1 as well. Since the attribute mastery indicator, ξ_{ic}, is created through multiplication of each alpha for every measured attribute, lack of a single measured attribute would cause the value of ξ_{ic} to be 0.

The DINO Model

The deterministic input, noisy-or-gate model, known as DINO, is a compensatory CDM (Templin, 2004; Templin & Henson, 2006) because it assumes that lack of one measured attribute can be compensated by another attribute. More specifically, mastery of at least one attribute compensates deficit of all the other measured attributes. Similarly to the DINA model, the slipping and the guessing parameters are estimated at the item level. The DINO model works with a disjunctive condensation rule in which the presence of at least one measured attribute guaranties a high probability of endorsing an item (Rupp et al., 2010).

DINO model estimates the probability of a correct response for item i in latent class c as follows:

$$\pi_{ic} = P(X_{ic} = 1 | \omega_{ic}) = (1 - s_i)^{\omega_{ic}} g_i^{(1-\omega_{ic})} \tag{5}$$

where π_{ic} is the probability of correct response, x_{ic} is the observed response, ω_{ic} is the latent response variable, and s_i and g_i are, respectively, the slipping and the guessing parameters (Rupp et al., 2010). The latent response variable ω_{ic} in the DINO model above is defined as follows:

$$\omega_{ic} = 1 - \prod_{a=1}^{A} (1 - \alpha_{ca})^{q_{ia}} \tag{6}$$

where q_{ia} specifies whether attribute a is measured by item i. q_{ia} takes the binary values of 0 or 1. Likewise, α_{ca} indicates whether the respondent in latent class c mastered attribute a, which takes values of 0 or 1 as well. In case that attribute a is not measured by item i, q_{ia} would take a value of 0, and consequently the value of $1 - \alpha_{ca}$ would not matter. On the other hand, if attribute a is measured by item i, q_{ia} would take a value of 1, and accordingly $1 - \alpha_{ca}$ counts for the final value that ω_{ic} takes. If the respondent in latent class c masters attribute a, α_{ca} takes a value of 1, and thereby $1 - \alpha_{ca}$ would be 0. However, if the respondent in latent class c does not master attribute a, $1 - \alpha_{ca}$ is 1. Because the occurrence of $\omega_{ic} = 1$ depends on existence of at least one 0 in the multiplication term, mastering at least one attribute greatly increases the probability of endorsing the item. The DINO model is useful when only one attribute is required to be mastered among more than one attribute (Rupp et al., 2010). The slipping, s_i, and the guessing, g_i, parameters of the DINO model are defined in the same way as in the DINA model.

Cognitive Diagnostic Assessment With Latent Class Analysis

The DINA and DINO models can be viewed as constrained LCA models. The constraints are placed on class probabilities, and the models are structured by specified Q-matrices. Thus, in this section, before extending these CDMs to longitudinal data, the unrestricted LCA model will be briefly summarized. The LCA model identifies unobserved groups (i.e., latent classes) based on observed categorical indicators, which are assumed to have multinomial distributions. The LCA model is (Collins & Lanza, 2010)

$$P(Y=y)=\sum_{c=1}^{C}\gamma_c\prod_{j=1}^{J}\prod_{r_j=1}^{R_j}\rho_{j,r_j|c}^{I(y_j=r_j)} \tag{7}$$

where $I(y_j - r_j)$ is an indicator function that is equal to 1 when the response is r_j, and equal to 0 otherwise. In LCA, two different parameters are estimated: γ_c is the latent class prevalence, which is the probability of being in a latent class and sum to 1; $\rho_{j,r_j|c}$, which is the probability of response r_j to observed variable j, conditional on membership in latent class c (Collins & Lanza, 2010) and also sum to 1. LCA is most frequently used as an exploratory method where the latent classes are unknown, and identified through model comparisons using likelihood ratio tests and information indices (Nylund, Asparouhov, & Muthén, 2007). Confirmatory latent class analysis can be performed by adding equality constraints (i.e., parameters are set to equality across classes), deterministic constraints (i.e., set the conditional probability of responding to an item at zero or one for a certain class), or inequality constraints (i.e., specify an order of size of conditional probabilities of responding to an item between classes; Finch & Bronk, 2011). The DINA and DINO models can be specified as a LCA with deterministic constraints. These constraints are detailed by Rupp et al. (2010).

Latent Transition Analysis Model

The LTA model, also known as Mixture Latent Markov model (Collins & Wugalter, 1992; Hagenaars, 1990; Poulsen, 1982; Van de Pol & Langeheine, 1990), is a latent variable model similar to the LCA model. However, while LCA is used with cross-sectional data, the LTA is used with longitudinal data to investigate whether any change occurred in between latent classes across time. The latent transition model with only two time points is

$$P(Y=y)=\sum_{s_1=1}^{S}\sum_{s_2=1}^{S}\delta_{s_1}\tau_{s_2|s_1}\prod_{t=1}^{2}\prod_{j=1}^{J}\prod_{r_{j,t}=1}^{R_j}\rho_{j,r_{j,t}|s_t}^{I(y_{j,t}=r_{j,t})} \tag{8}$$

where

$$\sum_{s_t=1}^{S}\delta_{s_t}=1 \tag{9}$$

$$\sum_{r_{j,t}=1}^{R_j}\rho_{j,r_{j,t}|s_t}=1 \tag{10}$$

and

$$\sum_{r_{j,t}=1}^{S} \rho_{j,r_{j,t}|s_t} = 1 \tag{11}$$

where S is the number of latent statuses, T is number of time points, J is number of items, $r_{j,t}$ indicates response category, R_j is number of response categories, and s_t is the latent status at time t. $I(y_j = r_j)$ is an indicator function that is equal to 1 when the response is r_j, and 0 otherwise. LTA models estimate three different sets of parameters: δ_{s_t} is the latent status prevalence, which is the probability of membership in latent status s at time t; $\rho_{j,r_{j,t}|s_t}$ is the item-response probability; $\tau_{s_{t+1}|s_t}$ is the transition probability, which is the probability of a transition to latent status s at time $t+1$ conditional on membership in latent status s at time t (Collins & Lanza, 2010). Latent status prevalences and item response probabilities are also estimated by LCA models.

Latent Status Prevalences. Latent status prevalences are the counterparts of latent class prevalences in the LCA model. Latent status prevalences are estimated for each time point, and one value is estimated for each latent status. Thus, for example, for a model with two time points and four estimated latent statuses, eight prevalences are estimated in total as showing the proportion of people on each latent status. For each time point, latent status prevalences across the latent statuses sum to 1. Hence, by looking at latent status prevalences, we can see increases and decreases on prevalences between time points, but we cannot see patterns of change and how individuals moved between statuses.

Item Response Probabilities. Item response probabilities represent probabilities of individuals being in each latent status who gave the correct or a particular response to each item. Because LTA is most commonly used as an exploratory method, the labelling of latent statuses is performed by examining item response probabilities. There is an item-response probability for each item–status combination. Thus, the number of cells in an item-response probabilities table is the number of items times the number of latent statuses.

For model identification, and interpretability of estimates, a restriction is placed on item-response probability estimation in LTA: Item-response probabilities are constrained to be equal across different time points. This corresponds to assuming measurement invariance in LTA and allows the meaning of latent statuses to remain the same across different time points. Consequently, in LTA, a single item-response probability matrix is estimated for all the time points and probabilities.

Transition Probabilities. Transition probabilities are critical for the current study, because they allow identification of whether any change occurs between latent classes across time. For a model with two time points, transition probability allows the creation of a table with a cross-classification of the number of latent statuses in Time 1 by the number of latent statuses in Time 2. Given a certain number of time points T, an LTA analysis estimates $T - 1$ transition probability matrices. For each individual, latent status membership is mutually exclusive and exhausted. In other words, individuals can be a member of only one latent status at each time point.

Longitudinal Cognitive Diagnostic Assessment

The main objective of the current study is to propose longitudinal DINA and DINO models to classify examinees with respect to mastery of attributes at multiple measurement waves, and follow individual stability or change in mastery of attributes across time. The proposed models classify examinees into latent statuses through consecutive measurements and estimate interclass transition probabilities of examinees from Time 1 to Time T.

The longitudinal DINA model estimates the probability of obtaining a particular vector of responses as a function of the probabilities of membership in each cognitive class at Time t, the probabilities of transitioning to a cognitive class at Time $t+1$ from the cognitive class membership at Time 1, and the probabilities of observing each response at each time point conditional on cognitive class membership c. The longitudinal DINA combines the LTA model and the DINA model as follows:

$$P(Y=y)=\sum_{c_1=1}^{C}\cdots\sum_{c_T=1}^{C}\delta_{s_1}\tau_{c_2|c_1}\cdots\tau_{c_T|c_{T-1}}\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{r_{i,t}=1}^{R_i}\left[(1-s_i)^{\xi_{ic_t}}g_i^{(1-\xi_{ic_t})}\right]^{I(y_{i,t}=r_{i,t})} \tag{12}$$

where R_i is the number of response categories of item i, s_i and g_i, respectively, represent slipping and guessing parameters of item i, and $\mathrm{I}(y_{j,t}=r_{j,t})$ is an indicator function which is equal to 1 when the response to item i at time t is r_j, and is equal to 0 otherwise. Another required component of the longitudinal DINA model is the latent response variable ξ_{ic_t} (see Equation 4) that is defined for item i, and cognitive class c at time t.

Similarly to the longitudinal DINA model, the longitudinal DINO model also estimates the probability of observing a particular vector of responses. The model is formulated as follows:

$$P(Y=y)=\sum_{c_1=1}^{C}\cdots\sum_{c_T=1}^{C}\delta_{c_1}\tau_{c_2|c_1}\cdots\tau_{c_T|c_{T-1}}\prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{r_{i,t}=1}^{R_i}\left[(1-s_i)^{\omega_{ic_t}}g_i^{(1-\omega_{ic_t})}\right]^{I(y_{i,t}=r_{i,t})} \tag{13}$$

where ω_{ic_t} is the latent response variable (see Equation 6) of the longitudinal DINO model, which differentiates it from the longitudinal DINA model.

Similarly to the LTA, three types of parameters are produced as a result of longitudinal DINA and DINO model estimations: item-response probabilities, latent status prevalences, and transition probabilities. In addition to these three parameters, posterior probabilities (expected a posteriori) are estimated to determine classes for each individual in each time point. For the practical use of longitudinal DINA and DINO models, the parameters of interest are posterior probabilities and transition probabilities. Posterior probabilities include individual information and are used to evaluate each student's performance based on their estimated class group. On the other hand, transition probabilities illustrate the probabilities of remaining in the same class or moving between classes from time t to time $t + 1$ for the respondents who are in the same cognitive class, which gives us an overall idea of the success of a treatment applied in between two measurements. In the next section, we present an evaluation of the performance of the longitudinal DINA and the DINO models through a Monte Carlo simulation study of conditions that are typical of applications of cognitive diagnostic modeling. The conditions examined reflect manipulations of the correlation between attributes, sample size, and slipping and guessing parameters.

Method

Data Generation

To evaluate the application of the proposed model, a simulation study was implemented with $T = 2$ time points, $I = 30$ items, $A = 3$ attributes, and $N = 500$ or 1,000 examinees. To analyze attribute stability of the examinees between two hypothetical data collections, two data sets were simulated for the same examinees representing the two time points. The Q-matrix was generated including items with one, two, and three attribute requirements. Table 1 presents the Q-matrix for 30 items and 3 attributes, and Table 2 presents all the possible simulated attribute profiles by showing the items sharing the same attribute profile. As shown in the both tables, 15 out of 30 items (50%) involve a single attribute, 11 items (37%) involve two attributes, and 4 items (13%) involve three attributes.

Two correlation conditions between the attribute pairs were developed. First, the attributes were simulated with a population correlation of zero, and then a correlation of 0.5 was simulated between the attributes to represent a moderate correlation level. These correlation conditions are similar to examples in the literature (Henson & Douglas, 2005; Henson, Templin, & Douglas, 2007). Also, the proportion of examinees mastering an attribute, p_k, was set to 0.50.

The number of possible attribute profiles are defined through the formula of $2^A = 2^B = 8$, with A representing the number of attributes. Of 8 possible attribute profiles, one is 000, indicating that the item does not measure any of the present attributes. Thus, we did not include the profile of 000 in this simulation study. The other 7 possible profiles were included in the design. The item-to-profile table is presented in Table 2. We tried to equally distribute the number of items to each attribute profile.

Following previous studies (de la Torre & Douglas, 2004; de la Torre & Lee, 2010; J. Liu et al., 2012), two sample sizes were simulated: 500 and 1,000. True alphas, α_{ca}, which are the binary indicators of whether the examinee in latent class c mastered the measured attribute or not, were simulated from a probit model with a multivariate normal vector with mean of zero and correlation between the attributes of 0 or 0.5.

Table 1. Simulated Q-Matrix for 30 Items and 3 Attributes.

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

Note. 1 = item measures attribute; 0 = item does not measure attribute.

Table 2. Simulated Attribute Profiles.

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

Finally, the slipping and the guessing parameters were varied to represent high and low discriminating items. Two different conditions were simulated for the slipping and guessing parameters: In the first condition, the slipping and guessing parameters were randomly assigned from a uniform distribution with a maximum of 0.2 and a minimum of 0. In the second condition, the ranges of the slipping and guessing parameters were determined between 0.2 and 0.4. These conditions have been frequently simulated in previous studies (e.g., de la Torre & Lee, 2010; Henson & Douglas, 2005; Y. Liu, Douglas, & Henson, 2009; Rupp & Templin, 2008).

The examinee responses were generated based on longitudinal DINA and DINO models. The R.3.1.0 0 software (R Development Core Team, 2012) and CDM package (Robitzsch, Kiefer, George, & Uenlue, 2015) were used to simulate data. To investigate the stability of the student classes from Time 1 to Time 2 in a longitudinal setting, measurements across two time points were simulated. Seven classes were defined for both time points. A thousand data sets were generated and analyzed for each condition. In total, 16 conditions were simulated based on a crossed design ($2 \times 2 \times 2 \times 2$) with the following factors and levels: number of sample sizes (500 or 1,000), magnitude of correlation between the attributes (0 or 0.5), size of slipping and guessing parameters (between 0 and 0.2, or 0.2 and 0.4), and the type of CDM (longitudinal DINA or DINO).

Data Analysis

The Mplus software (Muthén & Muthén, 2013) was used to implement longitudinal CDM analyses by modifying LCA and LTA applications of the software, following the recommendations of Rupp et al. (2010). Convergence rates of each condition over a thousand iterations were also obtained. The simulated 1,000 data sets per condition were analyzed, and the average of the correct classification rates (CCRs) and the marginal correct classification rates (MCCRs) were obtained over a 1,000 iterations. The CCRs for each time point were calculated by comparing the true and the estimated classes of the examinees for a 1,000 iterations and dividing the number of correctly classified iterations by a 1,000. Correct classifications of both time points together were also calculated. The MCCRs for each attribute were individually computed for each time point. The longitudinal DINA and DINO models were compared in terms of the correct classification of the examinees to latent classes.

Results

In this section, we present results from the Monte Carlo simulation study designed to evaluate the performance of the longitudinal DINA and DINO models to estimate diagnostic classes of the simulated examinees across two repeated measurements. The convergence rates of the longitudinal DINA model were over 0.99 for all the conditions, while the convergence rate for the longitudinal DINO model was over 0.93 for all the conditions. None of the manipulated conditions of sample size, size of slipping and guessing parameters, and correlations had any noticeable effect on the convergence rates for either model. The results of the longitudinal DINA and DINO models with respect to CCR, correct transition rate (CTR), and MCCR are presented separately in the two sections below.

Results for the Longitudinal DINA Model

Table 3 presents the MCCR for each attribute in Time 1 for the longitudinal DINA model. The two slipping and guessing parameter conditions are codded as *1* representing small slipping and guessing values simulated from a normal distribution with a maximum of 0.2 and a minimum of 0, and 2 representing large slipping and guessing condition, which is distributed with a maximum of 0.4 and a minimum of 0.2. The results indicated that sample size did not have an effect on the MCCRs of any of the attributes when the slipping and guessing parameters are small. Nonetheless, a slight increase of the MCCR was observed as the sample size increased from 500 to 1,000 for the large slipping and guessing conditions. Even though we randomly simulated the attribute parameters, the second attribute's MCCR values were lower than both the first and the third attributes. This could only be explained by the random variation of the parameters. Table 3 also shows that higher slipping and guessing parameters produced lower MCCR values. While the MCCRs were as high as 0.998 for the first slipping and the guessing condition, the highest values were around 0.896 for the second condition. For low slipping and guessing parameters, correlations between the attributes did not affect the MCCRs. However, for higher slipping and guessing parameters, correlation of 0.5 provided higher MCCRs than correlation of 0 did.

Since the parameters of Time 1 and Time 2 were randomly simulated, the MCCRs for Time 2, which is presented in Table 4, were similar to the MCCRs of Time 1. The CCR values were around 0.99 for high slipping and guessing parameters, and ranged from 0.856 to 0.895 for low slipping and guessing parameters.

Table 5 presents the overall CCRs for each time point, and the CTRs of the individuals between these two time points. As expected, CCRs showed similar patterns for the two time points because of the random simulation of the parameters. For the low slipping and guessing parameters, obtained CCR values were over 0.98 in both time points, which represents a nearly perfect classification rate. On the other hand, CCRs were unacceptably low, ranging from 0.68 to 0.72, in the conditions where slipping and guessing parameters were between 0.2 and 0.4.

The effect of sample size on the classification rates interacted with the level of slipping and guessing parameters. When the slipping and guessing parameters were between 0 and 0.2, sample size did not show an effect on the CCRs. However, when we increased slipping and guessing parameters to the range of 0.2 to 0.4, we obtained higher CCRs for the data with 1,000 sample size than for those with 500 samples. Likewise, an interaction was observed between the correlation size and the level of slipping and guessing parameters on the CCR values. For low slipping and guessing parameters, correlations between the attributes did not affect the correct classification

Table 3. Marginal Correct Classification Rates for Each Attribute in Time 1 for DINA Model.

Sample	S and G	Correlation	MCCR in Attribute 1	MCCR in Attribute 2	MCCR in Attribute 3
500	0-0.2	0	0.997	0.995	0.997
500	0-0.2	0.5	0.997	0.995	0.997
1,000	0-0.2	0	0.998	0.995	0.997
1,000	0-0.2	0.5	0.998	0.995	0.997
500	0.2-0.4	0	0.874	0.855	0.873
500	0.2-0.4	0.5	0.893	0.877	0.891
1,000	0.2-0.4	0	0.879	0.860	0.876
1,000	0.2-0.4	0.5	0.896	0.883	0.896

Note. DINA = deterministic input, noisy-and-gate model; S = slipping parameter; G = guessing parameter; MCCR = marginal correct classification rate.

Table 4. Marginal Correct Classification Rates for Each Attribute in Time 2 for DINA Model.

Sample	S and G	Correlation	MCCR in Attribute 1	MCCR in Attribute 2	MCCR in Attribute 3
500	0-0.2	0	0.997	0.995	0.997
500	0-0.2	0.5	0.997	0.995	0.997
1,000	0-0.2	0	0.997	0.995	0.998
1,000	0-0.2	0.5	0.997	0.995	0.997
500	0.2-0.4	0	0.872	0.856	0.873
500	0.2-0.4	0.5	0.890	0.877	0.892
1,000	0.2-0.4	0	0.876	0.859	0.877
1,000	0.2-0.4	0.5	0.894	0.882	0.895

Note. DINA = deterministic input, noisy-and-gate model; S = slipping parameter; G = guessing parameter; MCCR = marginal correct classification rate.

of the examinees; however, as slipping and guessing parameters increased, correlation of 0.5 provided higher CCRs than no correlation did. CTRs of the examinees from Time 1 to Time 2 followed a similar pattern as the CCR values for each time point, but the values were distinctively lower than CCRs.

Results for the Longitudinal DINO Model

The simulation results for the longitudinal DINO model showed that, similarly to the longitudinal DINA model, the CCRs and the MCCRs were at adequate levels. Table

Table 5. Correct Classification Rates for the Two Time Points and the Correct Transition Rates Between the Two Time Points for DINA Model.

Sample	S and G	Correlation	CCR in Time 1	CCR in Time 2	CTR between Time 1 and Time 2
500	0-0.2	0	0.990	0.990	0.980
500	0-0.2	0.5	0.989	0.989	0.978
1,000	0-0.2	0	0.990	0.990	0.980
1,000	0-0.2	0.5	0.990	0.989	0.979
500	0.2-0.4	0	0.683	0.680	0.462
500	0.2-0.4	0.5	0.723	0.721	0.521
1,000	0.2-0.4	0	0.692	0.690	0.476
1,000	0.2-0.4	0.5	0.733	0.731	0.536

Note. DINA = deterministic input, noisy-and-gate model; S = slipping parameter; G = guessing parameter; CCR = correct classification rate; CTR = correct transition rate.

Table 6. Marginal Correct Classification Rates for Each Attribute in Time 1 for DINO Model.

Sample	S and G	Correlation	MCCR in Attribute 1	MCCR in Attribute 2	MCCR in Attribute 3
500	0-0.2	0	0.998	0.995	0.997
500	0-0.2	0.5	0.997	0.995	0.997
1,000	0-0.2	0	0.998	0.995	0.997
1,000	0-0.2	0.5	0.997	0.995	0.997
500	0.2-0.4	0	0.875	0.857	0.873
500	0.2-0.4	0.5	0.894	0.879	0.892
1,000	0.2-0.4	0	0.879	0.861	0.877
1,000	0.2-0.4	0.5	0.897	0.882	0.894

Note. DINO = deterministic input, noisy-or-gate model; S = slipping parameter; G = guessing parameter; MCCR = marginal correct classification rate.

6 illustrates the MCCRs for each attribute in Time 1 with the longitudinal DINO model.

According to Table 6, the MCCRs of the longitudinal DINO model for Time 1 were at or over the level of 0.85 for each condition. Because each attribute was simulated randomly, as expected the results did not differ between the three attributes. Overall results indicated that the effect of sample size interacted with the magnitude of slipping and guessing parameters. This finding corresponds to the results that were obtained from the longitudinal DINA model. For low slipping and guessing condition, MCCRs did not differ between 500 and 1,000 sample sizes. However, as the slipping and guessing parameters increased, MCCRs showed slightly higher values

Table 7. Marginal Correct Classification Rates for Each Attribute in Time 2 for DINO Model.

Sample	S and G	Correlation	MCCR in Attribute 1	MCCR in Attribute 2	MCCR in Attribute 3
500	0-0.2	0	0.998	0.995	0.997
500	10-0.2	0.5	0.997	0.995	0.997
1,000	0-0.2	0	0.998	0.995	0.998
1,000	0-0.2	0.5	0.997	0.995	0.997
500	0.2-0.4	0	0.873	0.854	0.873
500	20.2-0.4	0.5	0.892	0.877	0.892
1,000	0.2-0.4	0	0.879	0.860	0.878
1,000	0.2-0.4	0.5	0.896	0.881	0.895

Note. DINO = deterministic input, noisy-or-gate model; S = slipping parameter; G = guessing parameter; MCCR = marginal correct classification rate.

for the larger sample sizes. Moreover, the effect of the correlation between the attributes was observed for larger slipping and guessing conditions, but not for the smaller values. For the conditions with high slipping and guessing parameters, the correlation of 0 between the attributes provided higher MCCR values than the correlation of 0.5 did.

Table 7 presents MCCRs for each attribute in Time 2 for the longitudinal DINO model. The results for Time 2 were slightly different from the results of Time 1, but still showed similar patterns. Because the parameters were randomly and independently simulated for each time point, this difference could only be explained by randomness. The MCCRs for Time 2 were at high levels for the all attributes. The results did not vary between the three attributes, or between the two sample sizes. In terms of the correlation between the attributes, the MCCRs were higher for the correlation of 0.5 than the MCCRs for the correlation of 0. The smallest MCCR was 0.854.

Table 8 shows CCRs for the two time points, and CTRs between these two time points for the longitudinal DINO model. The CCRs of both time points and the CTRs between these points were at adequate levels. The highest and lowest CCRs were 0.991 and 0.722, respectively. Even though the CCRs for the longitudinal DINO model differed between some conditions, the results followed the same pattern and were roughly similar for the two time points. Sample size did not show a visible effect on CCRs, but increasing the correlation between the attitudes positively affected the CCRs when slipping and guessing parameters were relatively high, which corresponds to the findings of the longitudinal DINA model presented before. Data sets with 0.5 correlation between the attitudes produced higher CCRs than the data sets with 0 correlation. This effect was not observed when the slipping and the guessing parameters were small. Finally, the CTRs between Time 1 and Time 2 for the DINO model were at adequate levels when slipping and guessing parameters were small. However, a significant reduction was observed for the conditions with

Table 8. Correct Classification Rates for the Two Time Points and the Correct Transition Rates Between the Two Time Points for DINO Model.

Sample	S and G	Correlation	CCR in Time 1	CCR in Time 2	CTR between Time 1 and Time 2
500	0-0.2	0	0.990	0.991	0.981
500	10-0.2	0.5	0.989	0.990	0.979
1,000	0-0.2	0	0.990	0.991	0.981
1,000	0-0.2	0.5	0.990	0.990	0.980
500	0.2-0.4	0	0.685	0.681	0.465
500	20.2-0.4	0.5	0.725	0.722	0.524
1,000	0.2-0.4	0	0.693	0.693	0.480
1,000	0.2-0.4	0.5	0.732	0.731	0.535

Note. DINO = deterministic input, noisy-or-gate model; S = slipping parameter; G = guessing parameter; CCR = correct classification rate; CTR = correct transition rate.

low slipping and guessing parameters. The values ranged from 0.732 to 0.465, which indicated a poor classification.

Discussion and Conclusion

Longitudinal versions of the DINA and the DINO models for cognitive diagnostic assessment across multiple waves were evaluated in this study. The results indicated that the longitudinal DINA and the DINO models under the simulated conditions provided high convergence rates and accurate examinee classifications. Because of the similarity in performance between the two models, in this section, the effective use of the proposed models will be evaluated together, and an overall conclusion will be provided.

The study results suggest that size of slipping and guessing parameters is an important factor for obtaining correct classification of individuals into diagnostic groups in longitudinal cognitive diagnosis. As the slipping and guessing parameters decrease, CCRs substantially increase. The findings of this study suggest that the size of slipping and guessing parameters would be problematic for longitudinal DINA and DINO models if larger than 0.2. In this study, we aimed to examine realistic levels of the slipping and guessing conditions by including values at a range of 0 to 0.4. Levels higher than 0.4 are rarely encountered and also not desired in real applications. Furthermore, the effects of sample size and the correlation between the attributes on the correct classification of the individuals to the cognitive groups are also affected by the size of the slipping and guessing parameters. When slipping and guessing parameters are low, sample size or correlation between attributes do not have an effect on the classification rates. However, the effects of correlation and sample size appear as slipping and guessing parameters increase.

Correlation between attributes indicates that the attributes are associated with each other, and the mastery of an attribute provides information about the other associated attribute as well. The study findings show that, under small slipping and guessing conditions, existence of a correlation between attributes provides better classification accuracy. This finding agrees with Henson et al.'s (2007) findings from another cognitive diagnosis study. In their study, researchers argue that as the correlation increases, mastery of an attribute implies a greater chance of mastery of the other attributes that are correlated with this attribute. In this study, we only evaluated the effects of a zero and a moderate correlation to demonstrate the overall effects of the association between the attributes. The study indicated that existence of a correlation between the attributes helps improve CCRs of examinees to diagnostic groups; however, further research is needed to clarify this effect.

For the conditions with high slipping and guessing parameters, sample size has a clear effect on the accuracy of classification of the longitudinal cognitive diagnosis. As indicated in the Results section, for groups with higher sample sizes, chances of classifying the examinees into correct classes increase. Moreover, existence of a correlation between attributes provides higher CCRs than lack of correlation, when slipping and guessing parameters are fairly high. de la Torre and Douglas (2004) state that high slipping and guessing parameters are indicators of poor fit, and that high slipping and guessing parameters are caused either by incorrect identification of attributes, or misspecification of Q-matrix. For correct classification, researchers should make sure that the attributes and the Q-matrices are correctly identified by experts before deciding on the application of a longitudinal CDM.

The overall conclusion of this study is that the CCRs of the longitudinal DINA and DINO models do not differ significantly. Both models work well to classify individuals into cognitive groups when multiple examinations occurred. These are not competing models, but they are appropriate for different situations. The DINA is a model that works with a conjunctive condensation rule, and lack of one attribute substantially drops the probability of responding an item correctly. However, the DINO model works with a disjunctive condensation rule in which the presence of at least one measured attribute can produce a high probability of endorsing an item. The same rules apply to the longitudinal DINA and DINO models. Thus, researchers' decisions on which model to apply to their assessment situation should depend on the nature of the required diagnosis. For instance, if not possessing even only one of the required skills would be enough to be lead to inadequate performance on an assessment, the longitudinal DINA model would be the appropriate model for the analysis.

Our study is related to a recent study by Li, Cohen, Bottge, and Templin (2016) where they illustrated the use of LTA for assessing change in cognitive diagnosis through the DINA model on a real data application, and their findings support the results of current study. Our recommendation to researchers is to use longitudinal CDMs only in situations where small slipping and guessing parameters are expected. Although we did not examine sample sizes below 500, we expect the longitudinal DINA and DINO models not to perform well with smaller sample sizes when slipping and guessing parameters are larger than 0.2. The study results also show that nonzero correlations between attributes promote increases in CCRs.

Researchers may be interested in adding grouping variables or continuous covariates to the longitudinal CDM models as predictors. Including grouping variables allows testing of measurement invariance, equality of item response probabilities, and also equality of latent status prevalences at Time 1 across groups (Collins & Lanza, 2010). In addition, covariates can predict latent status prevalences and transition probabilities in the model. For instance, in a cognitive diagnostic analysis about an online course, a continuous variable providing information about how many video lectures the student completed or a binary variable displaying whether the student finished the homework can be added as predictors of the latent cognitive classes. Predictors of class membership can be added to the longitudinal CDM models via multinomial logistic regression. There has been recent research comparing one-step methods and three-step methods of adding predictors of class membership to LCA models (e.g., Asparouhov & Muthén, 2014; Collier & Leite, 2016). One future avenue of research is to compare one-step versus three-step methods for adding covariates in longitudinal CDM models.

Some limitations exist in the design of the Monte Carlo simulation study. First, we only simulated data for two measurement waves, so it is unclear to what extent our results generalize to higher numbers of waves. Also, we simulated data allowing attribute levels to drop between Time 1 and Time 2. However, the data could have been simulated assuming that the achievement level would not drop from a higher class to a lower class level, which could provide different results. In a real scenario in which achievement level is measured, latent classes can be ordered from low to high, and students can be allowed to move to only higher classes, with the assumption that learning levels of students always tend to increase rather than to decrease. Furthermore, because of the coding and labeling complexity of Mplus program, only three attributes were simulated in this study. The number of attributes is expected to be higher in some assessment scenarios. Data with higher number of attributes need to be evaluated in the future studies. Future studies could also investigate extensions of the longitudinal DINA and DINO models to detect treatment effects on change in mastery of time. Finally, the effect of Q-matrix and model misspecifications on the longitudinal DINA and DINO models are areas of interest for the future research.

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

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling*, *21*, 329-341.

Barnes, T. M. (2011). Novel derivation and application of skill matrices: The q matrix method. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. D. Baker (Eds.), *Handbook on educational data mining* (pp. 159-172). Boca Raton, FL: CRC Press.

Chiu, C. Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement*, *37*, 598-618.

Close, C. N. (2012). *An exploratory technique for finding the Q-matrix for the DINA model in cognitive diagnostic assessment: Combining theory with data* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, MN.

Collier, Z. K., & Leite, W. L. (2016, April). *Testing the effectiveness of three-step approaches with auxiliary variables in latent class and profile analysis*. Paper presented at the annual meeting of the American Education Research Association, Washington, DC.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences*. Hoboken, NJ: John Wiley.

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

de la Torre, J. (2008). An empirically-based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, *45*, 343-362.

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

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

DiBello, L. V., Stout, W., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-389). Hillsdale, NJ: Erlbaum.

Finch, W. H., & Bronk, K. C. (2011). Conducting confirmatory latent class analysis using Mplus. *Structural Equation Modeling*, *18*, 132-151.

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

Hagenaars, J. A. (1990). *Categorical longitudinal data: Log linear analysis of panel, trend and cohort data*. Newbury Park, 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, Champaign, IL.

Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement*, *29*, 262-277.

Henson, R., Templin, J., & Douglas, J. (2007). Using efficient model based sum-scores for conducting skills diagnoses. *Journal of Educational Measurement*, *44*, 361-376.

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.

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.

Liu, Y., Douglas, J. A., & Henson, R. A. (2009). Testing person fit in cognitive diagnosis. *Applied Psychological Measurement*, *33*, 579-598.

Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement*, *36*, 548-564.

Macready, G. B., & Dayton, C. M. (1977). Use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, *2*, 99-120.

Muthén, L. K., & Muthén, B. O. (2013). *Mplus* (Version 7.0). Los Angeles, CA: Muthén & Muthén.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*, 535-569.

Poulsen, C. A. (1982). *Latent structure analysis with choice modeling applications*. Aarhus, Denmark: Aarhus School of Business Administration and Economics.

R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2015). *CDM: Cognitive diagnosis modeling* (R package version 4.2-12). Retrieved from http://CRAN.R-project.org/package=CDM

Rupp, A., & Templin, J. (2008). The effects of Q-matrix misspecification on parameter estimates and misclassification rates in the DINA model. *Educational and Psychological Measurement*, *68*, 78-98.

Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York, NY: Guilford.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*, 345-354.

Templin, J. L. (2004). *Generalized linear mixed proficiency models for cognitive diagnosis* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Urbana-Champaign, IL.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*, 287-305.

Van de Pol, F., & Langeheine, R. (1990). Mixed Markov latent class models. *Sociological Methodology*, *20*, 213-247.

von Davier, M. (2014). The DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology*, *67*, 49-71.

Xu, G. (2013). *Statistical inference for diagnostic classification models* (Unpublished doctoral dissertation). Columbia University, New York City, NY.