Article

Explanatory Cognitive Diagnostic Models: Incorporating Latent and Observed Predictors

Yoon Soo Park[1], Kuan Xing[1], and Young-Sun Lee[2]

Abstract

Large-scale educational testing data often contain vast amounts of variables associated with information pertaining to test takers, schools, or access to educational resources—information that can help *explain* relationships between test taker performance and their learning environment. This study examines approaches to incorporate latent and observed explanatory variables as predictors for cognitive diagnostic models (CDMs). Methods to specify and simultaneously estimate observed and latent variables (estimated using item response theory) as predictors affecting attribute mastery were examined. Real-world data analyses were conducted to demonstrate the application using large-scale international testing data. Simulation studies were conducted to examine the recovery and classification for simultaneously estimating multiple latent (using dichotomous and polytomous items as indicators for the latent construct) and observed predictors for varying sample sizes and number of attributes. Results showed stable parameter recovery and consistency in attribute classification. Implications for latent predictors and attribute specifications are discussed.

Keywords

cognitive diagnostic models, explanatory models, item response theory, covariates

Introduction

Scores and survey data from large-scale educational assessments can be a rich and untapped resource for *explaining* relationships among fine-grained skills and student background variables. In analyzing such educational data, focus can be placed on one of two areas: (a) improving the precision around test taker performance from an underlying latent construct (i.e., studies on *descriptive measurement*) or (b) examining the relationship between item response data and other variables, targeting factors that explain associations with test taker performance (De Boeck & Wilson, 2004). This study focuses on the latter tradition of explanatory psychometric

[1]The University of Illinois at Chicago, USA
[2]Columbia University, New York City, NY, USA

Corresponding Author:
Yoon Soo Park, Department of Medical Education, The University of Illinois at Chicago, 808 South Wood Street, 986 CMET (MC 591), Chicago, IL 60612-7309, USA.
Email: yspark2@uic.edu

models that aim to investigate the relationship between test taker performance and explanatory variables.

Cognitive diagnostic models (CDMs) refer to a class of psychometric models that classify examinees into skill profiles that indicate their mastery in fine-grained skills or attributes (Rupp, Templin, & Henson, 2010; von Davier, 2005, 2008). Mastery of skills may be expressed by dichotomous or polytomous categories. Most CDMs are semiparametric models that use a latent class model for specifying skills (discrete latent variables), while including skill-based restrictions through parametric item response functions, in a manner similar to latent trait models such as item response theory (IRT; see DiBello, Roussos, & Stout, 2007). While CDMs and their applications to large-scale data have expanded rapidly, these studies center on improving the descriptive measurement of test takers—that is, describing the performance of test takers at a more fine-grained level. On the contrary, methods for examining relationships between explanatory variables and CDMs have generally been limited. Moreover, despite large amounts of data collected, methods to integrate explanatory variables to CDMs have not been fully studied. For example, techniques for estimating latent variables (e.g., measure of self-confidence in mathematics) using items as indicators that can then be simultaneously fit as explanatory variables predicting skill profiles of examinees have yet to be explicitly investigated. Moreover, incorporating both latent and observed predictors (e.g., access to educational resources such as calculator, computer, and Internet) has not been examined. As such, a framework to incorporate explanatory variables into CDMs requires attention.

Explanatory variables that serve as predictors, in the sense of traditional regression models, can take various forms—observed variables (continuous or categorical) or latent variables (continuous or categorical). To date, only a limited scope of explanatory factors for observed covariates (continuous or discrete) has been examined within the CDM framework through a latent regression modeling approach (Dayton & Macready, 1988; Huang & Bandeen-Roche, 2004; Park & Lee, 2014). On the contrary, explanatory IRT models, incorporating both observed and latent variables, have been widely studied and used in various applications (De Boeck & Wilson, 2004). The development of a CDM framework that incorporates both observed and latent variables would be useful to expand prior work on covariate-only CDMs (using only observed variables) and to advance explanatory models targeting CDMs.

The purpose of this study is to incorporate the explanatory framework based on latent regression that combines measurement and structural components for CDMs, resembling a structural equation model (SEM) for CDMs. In this approach, observed and latent predictors can be either continuous or categorical, and can be specified within an explanatory CDM framework. Latent explanatory factors can be estimated using IRT and simultaneously specified as predictors in CDMs. The deterministic input, noisy, ''and'' gate (DINA; Junker & Sijtsma, 2001) model is used to demonstrate the framework, where predictors can be specified at the item, attribute, or higher order latent trait levels of the CDM, and extensions of this approach could be applied to other CDM families and generalizations of DINA models (de la Torre, 2011; von Davier, 2008).

This study is divided into two parts: A real-world data analysis was conducted to demonstrate the utility and flexibility of the approach using the Trends in International Mathematics and Science Study (TIMSS) data. Simulation studies were conducted to examine the stability of parameter recovery under varying sample sizes, number of attributes, and types of explanatory factors (observed discrete and latent continuous variables based on dichotomous and polytomous items). Findings from this study can provide implications on the use of explanatory CDMs for applied researchers and can inform the development of more flexible explanatory latent variable models.

CDMs

CDMs were developed to provide more targeted information in the form of skill profiles that resolves limitations of IRT models (de la Torre, 2009). Various forms of CDMs have been developed in the measurement literature, where patterns of attribute mastery have been represented as latent classes (e.g., Haertel, 1989; Henson, Templin, & Willse, 2009), including polytomous levels of mastery (von Davier, 2005, 2008). Holistically, these models cover a variety of situations (i.e., types of construct, response, and dimensionality). Generalizations of CDMs have also been recently proposed, which include the log-linear cognitive diagnosis model (LCDM; Henson et al., 2009), general diagnostic model (GDM; von Davier, 2008), and the generalized DINA model (G-DINA; de la Torre, 2011).

Reparameterized DINA (RDINA) Model

This study focuses on the DINA model, which estimates the probability that examinee i answers item j correctly, given that all required skills (i.e., attributes, $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)'$) have been mastered. The estimated binary latent variable $\eta_{ij} = \sum_{k=1}^{K} \alpha_{ik}^{q_{jk}}$ classifies whether examinee i has mastered all required attributes for item j, as specified in the Q-matrix, where $q_{jk} = 1$ indicates that attribute k is required for solving item j, and $q_{jk} = 0$ otherwise.

   This study uses the RDINA (DeCarlo, 2011) to apply the explanatory framework, as it facilitates incorporating predictors based on a latent regression approach. The RDINA takes the logit of the traditional DINA model:

$$\text{logit } p\left(Y_{ij} = 1 | \eta_{ij}\right) = f_j + d_j \eta_{ij}. \tag{1}$$

The f_j parameter indicates the log odds of the guessing (g_j) parameter, and the d_j parameter provides a measure of how well the item can discriminate an examinee with or without the mastery of required skills. The guessing and slip parameters used in the DINA model can be recovered by exponentiating the RDINA parameters (DeCarlo, 2011): $g_j = \exp(f_j)/[1 + \exp(f_j)]$ and $s_j = 1 - \exp(f_j + d_j)/[1 + \exp(f_j + d_j)]$.

Covariate Extension to the RDINA

When an observed covariate, $\underline{Z}$, is introduced into a latent class model, an examinee's response probability becomes the following (DeCarlo, 2005; Park & Lee, 2014):

$$p(Y_{i1}, Y_{i2}, \ldots, Y_{iJ} | \underline{Z}) = \sum_{\underline{\alpha}} p(\underline{\alpha} | \underline{Z}) \prod_j p\left(Y_{ij} | \underline{\alpha}, \underline{Z}\right). \tag{2}$$

Equation 2 represents the response probability conditioning on covariate $\underline{Z}$, where the effect of the observed variable on the response probabilities, $p(Y_{ij} | \alpha, \underline{Z})$, or on the attribute probability, $p(\alpha | \underline{Z})$, can be separated into two parts: The observed variable affecting the response probability is shown in Equation 3, and the observed variable affecting the attributes (assuming independence) is shown in Equation 4:

$$\text{logit } p\left(Y_{ij} | \underline{\alpha}, \underline{Z}\right) = f_j + d_j \eta_{ij} + l_j \underline{Z}, \tag{3}$$

$$\text{logit } p(\alpha_k | \underline{Z}) = b_k + h_k \underline{Z}. \tag{4}$$

When the covariate Z is conditioned on the response probability, the magnitude of the guessing and slip rates for item j shifts by l_j. Moreover, when the covariate is conditioned on the attribute probability, attribute difficulty (b_k) shifts by h_k for each unit of Z. Other covariate approaches to CDMs have been proposed, including discrete forms and multiple-group extensions (von Davier, Xu, & Carstensen, 2011; Xu & von Davier, 2008). Unobserved covariates can be examined using mixtures of diagnostic models, as presented in von Davier (2007). In this study, the explanatory CDM framework is proposed by extending the covariate RDINA (Park & Lee, 2014), as it can be generalized to specify different parameterizations.

Explanatory CDMs: Incorporating Observed and Latent Predictors

This study is motivated from explanatory IRT models that include latent or observed explanatory factors (see De Boeck & Wilson, 2004) and is extended here for latent class models that subsume CDMs. To date, explanatory variables have been limited to observed categorical or observed continuous variables, where a covariate ($\underline{Z}$) is specified as a direct predictor of attribute or item-level responses. The explanatory CDM is generalized to include *both* observed and latent explanatory variables, as graphically illustrated in Figure 1 (for details on graphical representations of CDMs, refer to Rupp et al., 2010). In this approach, a latent variable (simultaneously estimated with items as indicators of the latent variable) is specified as a predictor, in addition to the observed covariate. The figures represent the explanatory CDM with predictors specified at the attribute (top: attribute-level explanatory RDINA) and item (bottom: item-level explanatory RDINA) levels. The right side of the figures represents the CDM, and the left side represents the explanatory variables.

The shaded area indicates the explanatory latent variable (ξ) that is simultaneously estimated using item response data from X_M (for M items) as indicators. Prior covariate-based methods in Park and Lee's (2014) study used only observed variables ($\underline{Z}$) as covariates. With respect to the illustration in Figure 1, they represent a subset of the explanatory CDM approach proposed in this study, which includes only the observed variables as predictors (excluding the shaded area with latent variable ξ). As such, the conditional response probability of the latent class model becomes modified from Equation 2 and is represented in Equation 5:

$$p\left(Y_{i1}, Y_{i2}, \ldots, Y_{iJ} | \underline{Z}, \underline{\xi}\right) = \sum_{\alpha} p\left(\underline{\alpha} | \underline{Z}, \underline{\xi}\right) \prod_j p\left(Y_{ij} | \underline{\alpha}, \underline{Z}, \underline{\xi}\right). \tag{5}$$

The structural components of the CDM that links the explanatory variables (latent and/or continuous) are as follows:

$$\text{logit } p\left(Y_{ij} | \underline{\alpha}, \underline{Z}, \underline{\xi}\right) = f_j + d_j \eta_{ij} + l_j \underline{Z} + l_j' \underline{\xi}, \tag{6}$$

$$\text{logit } p\left(\underline{\alpha}_k, \underline{Z}, \underline{\xi}\right) = b_k + h_k \underline{Z} + h_k' \underline{\xi}. \tag{7}$$

Equations 6 and 7 show the regression parameters for item and attribute levels, respectively. In Figure 1, the structural parameters with a prime ($'$) denote the explanatory regression effects of the latent explanatory variable. The measurement model associated with the latent explanatory variable (ξ) is estimated using IRT. For dichotomous item response X_M, the two-parameter logistic item response theory (2PL-IRT) model is used (see Equation 8), where parameters a_m and b_m represent discrimination and difficulty parameters for item m. For ordinal item response of N categories, the graded response model (GRM; Samejima, 1969) can be used (see Equation 9), where a_m is the discrimination parameter and b_{nm} is the category threshold parameter.
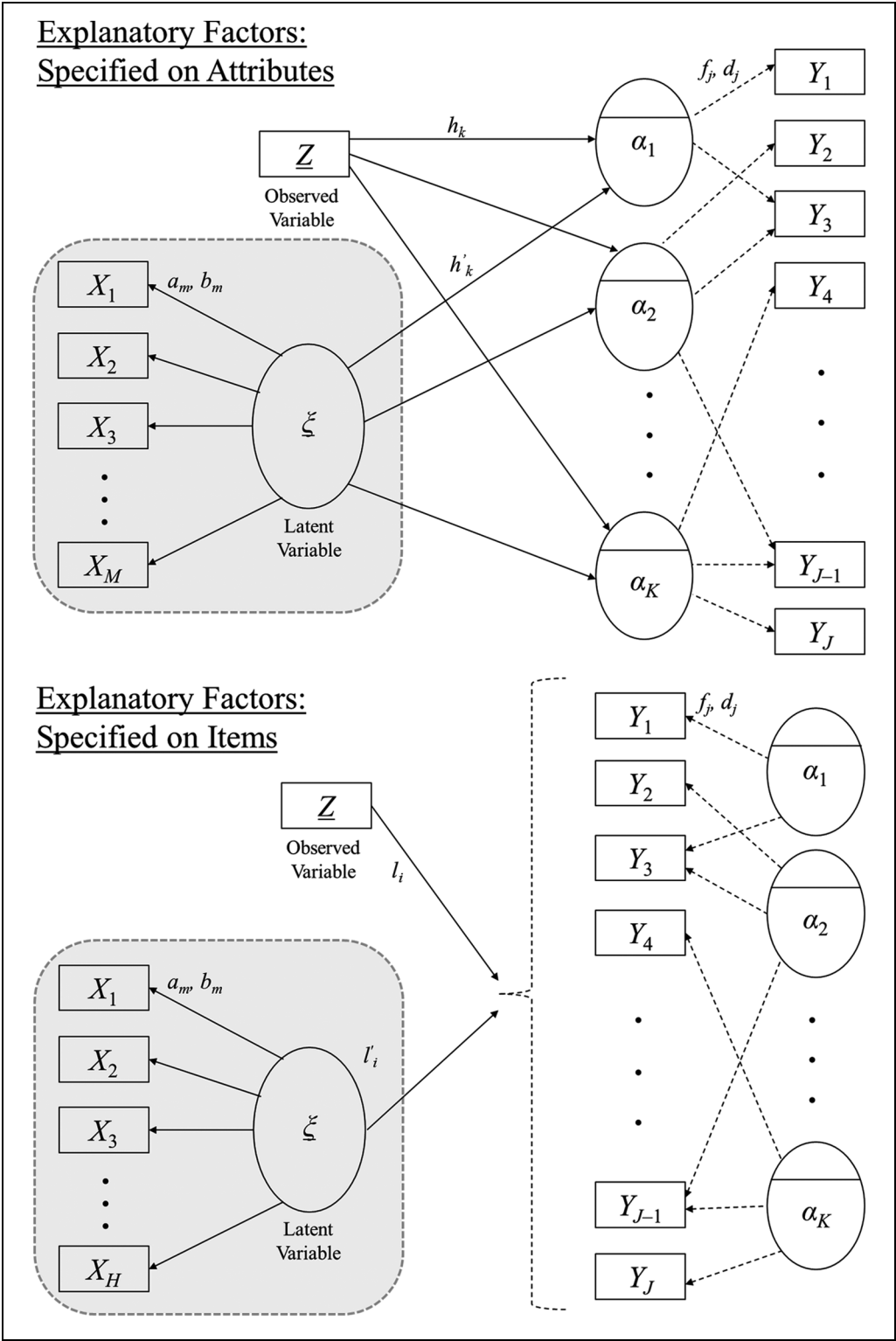
Figure 1. Graphical representation of explanatory CDM: explanatory factors (observed or latent) affecting attributes (top) and items (bottom).
Note. CDM = cognitive diagnostic model.

However, other variant IRT models can also be adopted based on theoretical motivation. For identification, the mean and variance of the ξ is fixed to (0, 1).

$$\text{logit } p(X_{im} = 1|\xi_i) = a_m(\xi_i - b_m), \tag{8}$$

$$\log \left[\frac{p(X_{im} \geq n|\xi_i)}{p(X_{im} < n|\xi_i)} \right] = a_m(\xi_i - b_{nm}). \tag{9}$$

There are several advantages to using the explanatory RDINA approach. The covariate model (Park & Lee, 2014) only includes the structural component, meaning that only observed covariates can be specified. If a latent variable (ξ) is included as a predictor, a two-step approach is required, where IRT is used to estimate the latent variable first; then, the estimated latent variable is used as an observed variable in the covariate model, forming the second step. However, the explanatory framework allows simultaneous estimation of latent factors as predictors within a single model (rather than a two-step process). The focus of this article is on the attribute- and item-level explanatory RDINA to examine factors that explain mastery of attributes and item responses.

Study I: Real-World Data Analysis Using TIMSS

Method

Real-world data analysis was conducted to demonstrate the application of the explanatory CDM presented in Figure 1. The 2007 fourth-grade mathematics TIMSS data were used, following the five-attribute Q-matrix specification in Park and Lee's (2014) study for 25 items from Booklet 4 ($M = 13.26$, $SD = 6.19$; see Table 1). In these data, examinees from Armenia ($n = 294$), Hong Kong ($n = 270$), Italy ($n = 321$), Morocco ($n = 265$), United States ($n = 565$), Massachusetts ($n = 127$), and Minnesota ($n = 133$) were aggregated, totaling 1,975 examinees.

Three predictors were used as explanatory variables:

1. Confidence: four-item self-reported measure of confidence in mathematics rated on a 4-point scale, ranging from 1 = "Disagree," 2 = "Agree a little," 3 = "Somewhat agree," and 4 = "Agree a lot." The items are as follows: (1) I usually do well in mathematics, (2) Mathematics is easier for me than for many of my classmates, (3) I am good at mathematics, and (4) I learn things quickly in mathematics (Cronbach's α = .70; $M = 2.02$, $SD = 0.71$).
2. Affect: three-item self-reported measure affect in mathematics, reported as "Yes = 1" or "No = 0." The items are as follows: (1) I enjoy learning mathematics, (2) Mathematics is fun, and (3) I like mathematics (Cronbach's α = .80; $M = 0.76$, $SD = 0.34$).
3. Calculator: self-reported status of calculator ownership (82% own calculators).

The explanatory variables were selected based on their association (correlation) with the mathematics score ($r_{\text{calculator}} = .14$, $r_{\text{affect}} = .13$, and $r_{\text{confidence}} = .33$, all ps < .001).

The attribute-level explanatory RDINA was estimated with "confidence" fit using GRM, "affect" fit using 2PL-IRT, and "calculator" fit as an observed dichotomous variable. To compare the attribute-level effects of the predictors, a covariate RDINA model was also fit, by separately estimating "confidence" and "affect" via IRT and using the predicted values as covariates (a two-step approach). The purpose of this comparison was to examine differences in estimates for simultaneous estimation (explanatory RDINA) versus a two-step approach

Table 1. RDINA Parameters.

Item	Attributes	f_j		g_j	d_j		s_j
1	α_1	0.21	(0.08)	0.55	1.55	(0.12)	0.15
2	α_2	−1.59	(0.29)	0.17	2.53	(0.31)	0.28
3	α_1,α_2	−1.44	(0.09)	0.19	2.15	(0.12)	0.33
4	α_1,α_2	−0.61	(0.08)	0.35	2.50	(0.13)	0.13
5	α_1,α_2	−0.75	(0.08)	0.32	2.54	(0.13)	0.14
6	α_3,α_4	0.64	(0.09)	0.66	2.45	(0.21)	0.04
7	α_3,α_4	−0.48	(0.08)	0.38	1.86	(0.12)	0.20
8	α_1,α_4	−0.91	(0.08)	0.29	1.80	(0.11)	0.29
9	α_1	−1.16	(0.13)	0.24	3.10	(0.16)	0.13
10	α_3,α_4	−1.58	(0.14)	0.17	2.63	(0.16)	0.26
11	α_1,α_3	−1.18	(0.09)	0.23	0.82	(0.11)	0.59
12	α_1,α_5	−0.68	(0.08)	0.34	2.40	(0.13)	0.15
13	α_1,α_5	−0.80	(0.08)	0.31	2.34	(0.12)	0.18
14	$\alpha_1,\alpha_2,\alpha_5$	−1.32	(0.08)	0.21	1.65	(0.11)	0.42
15	α_1	−0.45	(0.08)	0.39	2.61	(0.14)	0.10
16	α_1	−1.18	(0.09)	0.24	2.65	(0.13)	0.19
17	α_1,α_2	−2.14	(0.12)	0.11	2.12	(0.14)	0.50
18	α_1,α_2	−0.84	(0.08)	0.30	2.32	(0.12)	0.19
19	α_1,α_5	−1.78	(0.11)	0.14	2.51	(0.13)	0.32
20	$\alpha_1,\alpha_2,\alpha_5$	−0.50	(0.07)	0.38	1.87	(0.12)	0.20
21	α_1,α_2	−3.61	(0.24)	0.03	2.18	(0.26)	0.81
22	α_3,α_4	−0.80	(0.09)	0.31	1.11	(0.11)	0.42
23	α_1	−1.64	(0.12)	0.16	3.20	(0.14)	0.17
24	α_4	−1.14	(0.10)	0.24	1.58	(0.12)	0.39
25	α_1,α_5	−1.05	(0.09)	0.26	3.26	(0.15)	0.10

Note. Attributes are based on Q-matrix from Lee et al. (2011): α_1: Whole numbers (number domain), α_2: Fractions and number patterns (number domain), α_3: Lines and angles (geometric shapes and measures domain), α_4: Dimensions and locations (geometric shapes and measures domain), and α_5: Data display (data domain). Values in parentheses are standard errors. Data taken from TIMSS 2007 Fourth-Grade Mathematics Block 4. RDINA = reparameterized deterministic input, noisy, "and" gate.

(covariate RDINA). In both approaches, the variance of ''confidence'' and ''affect'' was fixed to 1.0 for identification.

Models were fit using Latent GOLD 5.0 (Vermunt & Magidson, 2013), with 100 sets of starting values. Estimation began with the expectation–maximization (EM) algorithm, followed by the Newton–Raphson (NR) to obtain maximum-likelihood (ML) or posterior mode (PM) estimates, to avoid boundary estimation issues. The Jacobian matrix was examined to be of full rank for local identification (Huang & Bandeen-Roche, 2004).

Results

The attribute-level model converged (tolerance criteria = 1.0×10^{-8}), with the following estimates: –2 log likelihood (–2LL) = 77,572.76, Akaike information criterion (AIC) = 77,756.75, Bayesian information criterion (BIC) = 78,270.88; 92 parameters.

RDINA item parameters. Table 1 shows the RDINA parameters. The corresponding guessing and slip parameters are also presented in the table. The mean guessing and slip parameter estimates were .28 and .27, respectively.

IRT parameters. IRT parameters for "confidence" and "affect" are presented in Table 2, where confidence was fit using GRM and affect was fit using 2PL-IRT. The item discrimination for confidence ranged between 1.26 and 1.54; for affect, item discrimination ranged between 1.28 and 3.44.

Attribute parameters. The attribute-level explanatory RDINA parameters indicate significant effects as predictors shifting attribute difficulty (b_k). Table 3 shows the results. To compare the simultaneous estimation provided by the explanatory RDINA and the two-step covariate RDINA models, parameter estimates are presented together.

Results show relatively large differences in the estimates for all parameters. In particular, the differences between the explanatory and the covariate models were larger when the predictor was based on a latent variable ("affect" and "confidence"). For b_k, the largest difference was in Attribute 2 (fractions and number patterns), with a difference of .98 units. For h_{k1} (calculator) and for h'_{k3} (confidence), the overall trends in the estimates remained the same, while the estimates themselves varied. For Attribute 1, h'_{13} (confidence) differed between the two models by as much as 6.13 points. However, for h'_{k2} (affect), the result for Attributes 2 (fractions and number patterns) and 4 (dimensions and locations) differed significantly between the explanatory and covariate models.

Owning a calculator and having higher self-confidence in mathematics had significant effects for Attributes 1 (whole numbers), 2 (fractions and number patterns), and 4 (dimensions and locations). On the contrary, higher "affect" had negative effects on attribute mastery for Attributes 1, 2, and 4. With respect to differences between models, the latent measure of affect, as estimated using the explanatory model, was significant for Attributes 2 (fraction and number patterns) and 4 (dimensions and location), whereas it was not significant for the covariate model, indicating differential attribute difficulty due to the inclusion of the measurement model, as opposed to only a structural component to the model.

Classification and latent class sizes. Classification (P_c) estimates based on posterior probabilities ranged between .86 and .98, with similar estimates between the explanatory and covariate RDINA models. Latent class sizes were similar, except for Attribute 2 (fractions and number patterns), which was .62 for the explanatory and .82 for the covariate RDINA models. For full comparison of classification and latent class sizes, see Online Appendix Table A1.

Item-level explanatory RDINA. The corresponding item-level explanatory RDINA model was also fit to examine convergence and trends in parameter estimates. For the item-level explanatory RDINA model, three predictors were specified on the items; GRM and 2PL were used to fit "confidence" and "affect," respectively. The model converged (tolerance criteria = 1.0×10^{-8}), with the following estimates: $-2LL = 76,132.24$, AIC $= 76,420.23$, BIC $= 77,224.95$; 144 parameters. P_c ranged between .77 and .92, and was lower for all attributes, when compared with the attribute-level explanatory RDINA model. The three predictors had significant effects on all 25 items (except for Item 11, which did not have a significant effect for calculator, $p = .400$). Complete results for the item-level explanatory RDINA can be obtained from the corresponding author.

Study II: Simulation Study

Method

Simulation studies were conducted to examine consistency in parameter recovery and classification using the TIMSS real-world results as generating population values. Five main conditions were examined:

Table 2. Explanatory Measurement Model Parameters: GRM and 2PL-IRT Models.

Construct	Item	Step (b_m)		Discrimination (a_m)		R^2
Confidence: Polytomous (graded response)	1. I usually do well in mathematics	b_1	−4.47 (0.18)	1.50	(0.09)	.73
		b_2	−2.42 (0.10)			
		b_3	0.49 (0.07)			
	2. Mathematics is easier for me than for many of my classmates	b_1	−2.35 (0.09)	1.32	(0.08)	
		b_2	−0.73 (0.06)			
		b_3	0.78 (0.06)			
	3. I am good at mathematics	b_1	−2.67 (0.11)	1.54	(0.09)	
		b_2	−1.05 (0.07)			
		b_3	0.32 (0.07)			
	4. I learn things quickly in mathematics	b_1	−3.15 (0.12)	1.26	(0.08)	
		b_2	−1.34 (0.07)			
		b_3	0.85 (0.06)			
Affect: Dichotomous (2 PL)	1. I enjoy learning mathematics	b_1	−1.30 (0.09)	2.53	(0.22)	.65
	2. Mathematics is fun	b_2	−0.64 (0.04)	1.28	(0.10)	
	3. I like mathematics	b_3	−1.44 (0.15)	3.44	(0.45)	

Note. Items measuring "confidence" are polytomous, scored using a 4-point scale ("Disagree,""Agree a little,""Somewhat agree,""Agree a lot"). GRM was used to fit the model; variance for ξ was fixed to 1 for identification. Items measuring "affect" is dichotomous ("No,""Yes"). The 2PL was used to fit the model; variance for ξ was fixed to 1 for identification. Values in parentheses are standard errors. R^2 is the proportion of reduction in error between total error and predicted error (see Vermunt & Magidson, 2013). GRM = graded response model; 2PL–IRT = two-parameter logistic item response.

Table 3. Attribute Difficulty (b_k) and Predictors (h_{kq}).

Attribute	Parameter	Explanatory RDINA		Covariate RDINA	
		Estimate	p value	Estimate	p value
1. Whole numbers	b_1	−1.26 (0.48)		−0.71 (0.13)	
	h_{11} (calculator)	3.82 (0.80)	<.001	1.36 (0.14)	<.001
	h'_{12} (affect)	−5.30 (1.55)	.001	−0.27 (0.08)	.001
	h'_{13} (confidence)	6.73 (2.15)	.002	0.60 (0.07)	<.001
2. Fractions and	b_2	−0.20 (0.54)		0.78 (0.25)	
number patterns	h_{21} (calculator)	1.74 (0.58)	.003	1.05 (0.24)	<.001
	h'_{22} (affect)	−2.21 (0.52)	<.001	−0.08 (0.14)	.600
	h'_{23} (confidence)	2.96 (0.71)	<.001	0.60 (0.14)	<.001
3. Lines and angles	b_3	5.65 (2.23)		5.34 (2.72)	
	h_{31} (calculator)	−1.07 (2.08)	.610	−0.70 (2.84)	.810
	h'_{32} (affect)	0.95 (0.82)	.250	0.24 (0.84)	.770
	h'_{33} (confidence)	0.41 (0.89)	.640	0.68 (0.86)	.430
4. Dimensions and locations	b_4	−1.38 (0.44)		−0.63 (0.15)	
	h_{41} (calculator)	3.25 (0.50)	<.001	1.55 (0.15)	<.001
	h'_{42} (affect)	−3.45 (0.44)	<.001	−0.14 (0.09)	.100
	h'_{43} (confidence)	4.17 (0.54)	<.001	0.47 (0.08)	<.001
5. Data display	b_5	1.41 (0.53)		2.00 (0.46)	
	h_{51} (calculator)	0.38 (0.45)	.400	0.09 (0.47)	.850
	h'_{52} (affect)	−0.52 (0.51)	.310	−0.23 (0.21)	.280
	h'_{53} (confidence)	0.60 (0.46)	.190	0.28 (0.18)	.110

Note. "Explanatory RDINA" used item-level affect and confidence measures as indicators (using GRM and 2PL) to predict α_k. "Covariate RDINA" used estimated covariates to predict α_k. "Calculator" is dichotomous (1 = "Yes," 0 = "No"); "affect" and "confidence" are continuous ($M = 0$, $SD = 1$). Values in parentheses are standard errors. Given that h'_{k2} (affect) estimates were negative, a separate analysis treating the three affect items to directly regress on the attributes was conducted. Results showed that Item 3 ("I like mathematics") had negative effects on attributes for all five attributes, while the effects for the two remaining items were either positive or negative depending on the attribute. GRM = graded response model; RDINA = reparameterized deterministic input, noisy, "and" gate.

1. Explanatory RDINA with one latent predictor (three dichotomous items fit using 2PL),
2. Explanatory RDINA with one latent predictor (four polytomous items fit using GRM),
3. Explanatory RDINA with one latent (GRM) and one observed dichotomous predictor,
4. Explanatory RDINA with two latent predictors (GRM and 2PL) and one observed dichotomous predictor, and
5. Covariate RDINA with three observed predictors (two continuous and one dichotomous).

The fifth condition was added to compare the parameter recovery and classification of the explanatory model that includes both latent and observed predictors (full model: Condition 4) with a covariate-based model that only has observed predictors (two continuous and one dichotomous observed predictors). Following the real-world data, the 25-item response data were generated. The choice to use three and four items for the dichotomous and polytomous (4-point scale) items was motivated from the real-world data, as most self-reported scales in TIMSS had similar structures.

Two sample size conditions (1,000 and 2,000 examinees) and attribute sizes (three and five attributes) were examined, resulting in a total of 20 conditions (= five model conditions × two sample sizes × two attribute specifications). The sample sizes were based on numbers of

examinees often used for large-scale data analysis, including the TIMSS; the number of attributes and Q-matrix specification follow from the TIMSS example. For the three-attribute simulation, attributes α_1 and α_2 were combined to form the "Number" attribute; attributes α_3 and α_4 were combined to form the "Geometry" attribute; and attribute α_5 was kept to form the "Data" attribute (see Table 1).

Data were generated using Stata 14 and fit using Latent GOLD 5.0. One hundred replications were fit for each condition. Parameter estimates were summarized and compared with population data using three measures: (a) bias, (b) % Bias, and (c) mean squared error (MSE). Here, $\text{Bias}(x) = 1/N \sum_{n=1}^{N} [\hat{e}_n(x) - e(x)]$, % Bias $= |\text{Bias}(x) / e(x)| \times 100\%$, and $\text{MSE}(x) = 1/N \sum_{n=1}^{N} [\hat{e}_n(x) - e(x)]^2$, where x is an arbitrary parameter, $e(x)$ is the generating (true) parameter value, and $\hat{e}_n(x)$ is the nth replicate estimate of parameter x among 100 replications. Similar to the real-world data analysis, P_c was calculated to examine the accuracy in the recovery of latent classes using the maximum posterior probability for each attribute. Both ML and PM estimations were used to avoid boundary estimation problems. Model identification was checked to ensure that the Jacobian matrix was of full rank.

Results

Condition 1 (using 2PL to estimate explanatory latent variable). Table 4 shows the parameter recovery results by sample size and by number of attributes.

For the three- and five-attribute conditions, the RDINA item parameters had % Bias less than or equal to 2.0%. Generally, recovery was better for the d_j parameter than f_j. Parameter recovery for the 2PL had % Bias less than or equal to 2.2%. The greatest variability was at the attribute level. The b_k parameter % Bias was less than 2.5% and 6.5% for the three- and five-attribute conditions. The % Bias was largest for the h'_k parameters, which was on average less than 10.5% for the three-attribute condition and less than 21.9% for the five-attribute condition.

Condition 2 (using GRM to estimate explanatory latent variable). The item-level trends in the explanatory CDM using items as indicators to predict attributes using GRM were similar to Condition 1, with % Bias less than 2.2% (see Table 4). The GRM parameter estimates had % Bias less than 1.3%. The largest bias was at the attribute level for the five-attribute condition (three-attribute condition had % Bias less than or equal to 2.0%), with % Bias of 6.3% and 4.7% for sample sizes of 1,000 and 2,000, respectively, for attribute difficulty.

Condition 3 (one observed covariate and one explanatory latent variable fit using GRM). Condition 3 examines estimation for one latent predictor (estimated using GRM) and one observed dichotomous predictor (see Table 5). Overall, the recovery for the three-attribute condition had % Bias less than 3.5% across both sample size conditions and models. For the five-attribute condition, the RDINA item and GRM parameters all had % Bias less than 1.7%. However, at the attribute level, the attribute difficulty parameter had % Bias of 15.2% and 10.2% for the three- and five-attribute conditions, respectively. The % Bias for h_{k1} (observed) and for h'_{k3} (latent) were as high as 16.8% and 10.3%, respectively.

Condition 4 (one observed covariate and two latent variables fit using GRM and 2PL). In Condition 4, two latent variables were fit using GRM and 2PL, in addition to an observed dichotomous variable. The % Bias for RDINA item and GRM was less than 2.3%. For the three-attribute condition, the % Bias for 2PL parameters was also less than 2.0%. For the five-attribute condition, the % Bias for 2PL discrimination parameters was as large as 29.0%. For the three-attribute condition, the % Bias for attribute difficulty were 12.1% and 8.3% for the 1,000 and 2,000

Table 4. Explanatory RDINA With IRT: Parameter Recovery (Conditions 1 and 2).

Condition	Attribute	Model	Parameter	$n = 1,000$			$n = 2,000$		
				Bias	% Bias	MSE	Bias	% Bias	MSE
1. Explanatory RDINA with 2PL	3	RDINA attribute	b_k	0.000	2.4	0.006	−0.005	2.0	0.003
			h'_k	0.008	10.4	0.030	−0.007	9.8	0.016
		RDINA item	f_j	−0.001	1.0	0.022	−0.005	1.0	0.010
			d_j	0.005	0.9	0.039	0.005	0.7	0.018
		2PL	b_m	−0.014	1.0	0.032	−0.013	1.2	0.019
			a_m	0.059	2.2	0.316	0.011	0.7	0.169
	5	RDINA attribute	b_k	0.106	6.5	0.096	0.092	5.0	0.062
			h_k	−0.042	21.8	0.111	−0.025	20.6	0.100
		RDINA item	f_j	0.003	2.0	0.046	−0.002	1.3	0.033
			d_j	0.005	1.3	0.063	0.010	0.8	0.043
		2PL	b_m	−0.007	0.7	0.026	−0.008	0.6	0.014
			a_m	0.030	1.8	0.266	0.038	1.4	0.161
2. Explanatory RDINA with GRM	3	RDINA attribute	b_k	0.004	1.8	0.006	−0.006	0.9	0.004
			h'_k	0.001	2.0	0.035	0.007	1.7	0.015
		RDINA item	f_j	−0.002	1.5	0.020	0.000	1.2	0.010
			d_j	0.008	1.0	0.039	0.002	0.7	0.018
		GRM	b_{nm}	0.007	0.6	0.026	0.004	0.4	0.013
			a_m	0.014	1.0	0.018	0.006	0.7	0.010
	5	RDINA attribute	b_k	0.111	6.3	0.087	0.082	4.7	0.062
			h_k	−0.005	1.6	0.117	−0.005	1.1	0.108
		RDINA item	f_j	−0.016	2.1	0.052	−0.015	1.7	0.038
			d_j	0.028	1.2	0.071	0.020	0.9	0.047
		GRM	b_{nm}	0.019	0.9	0.027	0.012	0.5	0.013
			a_m	0.016	1.2	0.017	0.013	1.2	0.008

Note. IRT = item response theory; MSE = mean squared error; 2PL = two-parameter logistic; GRM = graded response model; RDINA = reparameterized deterministic input, noisy, "and" gate.

Table 5. Explanatory RDINA With IRT and Observed Covariate: Parameter Recovery (Conditions 3 and 4).

Condition	Attribute	Model	Parameter	n = 1,000			n = 2,000		
				Bias	% Bias	MSE	Bias	% Bias	MSE
3. Explanatory RDINA with GRM and observed covariate	3	RDINA attribute	b_k	0.006	2.9	0.034	−0.013	3.4	0.018
			h'_{k1} (observed)	0.031	3.4	0.169	−0.015	2.2	0.077
			h'_{k3} (latent)	0.013	2.5	0.030	0.002	0.5	0.014
		RDINA item	f_j	−0.004	1.4	0.022	0.001	0.8	0.010
			d_j	0.011	0.8	0.039	−0.002	0.4	0.018
		GRM	b_{nm}	0.004	0.6	0.030	0.003	0.2	0.014
			a_m	−0.002	0.9	0.019	−0.001	0.5	0.010
	5	RDINA attribute	b_k	0.163	15.2	0.073	0.122	10.2	0.041
			h'_{k1} (observed)	0.214	16.6	0.438	0.213	16.8	0.373
			h'_{k3} (latent)	−0.045	10.3	0.155	−0.027	6.0	0.109
		RDINA item	f_j	−0.010	1.6	0.048	−0.008	1.2	0.034
			d_j	0.019	1.2	0.067	0.012	0.7	0.043
		GRM	b_{nm}	0.014	0.7	0.026	0.005	3	0.013
			a_m	0.001	0.7	0.018	−0.007	0.5	0.008
4. Explanatory RDINA with GRM, 2PL, and observed covariate	3	RDINA attribute	b_k	0.022	12.1	0.132	−0.011	8.3	0.074
			h'_{k1} (observed)	−0.005	6.4	0.769	−0.022	3.4	0.375
			h'_{k2} (Latent 1: 2PL)	0.129	6.1	0.535	0.013	3.9	0.311
			h'_{k3} (Latent 2: GRM)	−0.095	5.7	0.526	−0.022	3.1	0.323
		RDINA item	f_j	0.002	2.1	0.018	−0.007	2.2	0.009
			d_j	0.008	1.0	0.033	0.013	0.8	0.017
		2PL	b_j	−0.010	1.2	0.021	−0.008	1.6	0.010
			a_j	0.050	1.9	0.173	0.019	1.9	0.091
		GRM	b_{nm}	0.011	0.9	0.026	0.005	0.3	0.011
			a_m	0.015	1.0	0.016	0.007	0.5	0.007
	5	RDINA attribute	b_k	0.225	28.4	0.186	0.178	9.1	0.130
			h'_{k1} (observed)	−0.020	36.9	0.772	0.071	22.2	0.568
			h'_{k2} (Latent 1: 2PL)	−0.173	31.0%	0.656	−0.152	19.0	0.401
			h'_{k3} (Latent 2: GRM)	0.303	41.5	0.910	0.261	27.4	0.609
		RDINA item	f_j	−0.017	3.5	0.023	−0.007	2.6	0.010
			d_j	0.026	1.4	0.040	0.013	1.1	0.019
		2PL	b_m	0.225	28.7	0.023	0.209	29.0	0.013
			a_j	0.012	0.9	0.014	0.012	1.0	0.007
		GRM	b_{nm}	0.001	1.7	0.018	−0.009	0.7	0.012
			a_m	0.001	2.2	0.166	0.027	1.3	0.108

Note. IRT = item response theory; MSE = mean squared error; GRM = graded response model; 2PL = two-parameter logistic; RDINA = reparameterized D deterministic input, noisy, "and" gate.

sample size conditions, respectively. For the five-attribute condition, % Bias was 28.4% and 9.1%, respectively.

Moreover, for the three-attribute condition, the % Bias for the observed variable was 6.4%, whereas the % Bias for the latent variable estimated using GRM was 5.7% for the 1,000 sample size condition. This contrasts with the five-attribute condition, where % Bias was 36.9% for the observed variable, while the latent variables estimated with the 2PL and GRM were 31.0% and 41.5%, respectively, for the 1,000 sample size condition (see Table 5).

Condition 5 (covariate RDINA model using two continuous and one dichotomous observed covariates). The parameter recovery for one dichotomous and two continuous observed predictors was fit using the covariate RDINA model (see Online Appendix Table A2). The largest % Bias in the three-attribute model was for the continuous covariate effect (h_{k3}), which was 14.3%. RDINA item-level parameters had % Bias less than 1.6%. In the five-attribute model, the % Bias for the covariate effects ranged between 2.7% and 81.2%, and between 1.3% and 36.3% for sample sizes 1,000 and 2,000, respectively.

Classification and class sizes. Regardless of condition, number of attributes, and sample size, P_c was greater than .92 (see Online Appendix Table A3). When compared with a covariate-based model (Condition 5), P_c estimates were .94 for the three- and five-attribute conditions across samples, respectively. For latent class sizes, the % Bias was less than 1.5% across conditions, number of attributes, and sample sizes.

Summary of simulation results. Overall, parameter recovery was consistent across conditions, number of attributes, and sample sizes when single latent or observed predictor was specified (Conditions 1, 2, and 3). In particular, the recovery of IRT parameters was particularly notable, as % Bias was all less than 2.3%. Moreover, the RDINA item parameters had similar stability in recovery with % Bias less than 2.3%. The largest variability in estimation was at the attribute level, which could be due to predictors specified at this level. When only a single predictor was specified at the attribute level, the bias for attribute difficulty (b_k) was modest. However, when multiple predictors were specified—particularly for multiple latent predictors (Condition 4)—bias increased substantially. This was found to be the case when more attributes were specified in the model, noting the larger bias in the five-attribute condition relative to the three-attribute condition. When comparing these results with a covariate RDINA with only observed predictors, the largest % Bias for the structural attribute parameters (h and h' parameters in the three- and five-attribute conditions based on sample size 2,000) were 3.9% and 27.4% in the explanatory RDINA model, whereas they were 14.2% and 36.3% in the covariate RDINA model, respectively. This shows that parameter recovery was better for the explanatory model than for the covariate RDINA in the structural parameters. Classification and recovery of latent class sizes were consistent, regardless of condition, number of attributes, and sample size.

Discussion

Combining observed or latent predictors to CDMs to explain patterns in skill profiles addresses both methodological issues that advance explanatory models for latent classes and also facilitates answering important substantive questions for applied educational and psychological research. Many large-scale assessment data contain background variables. For example, prior research has shown that access to educational technology (e.g., computer and Internet access) is linked with improved computational ability of students in elementary and middle school mathematics and science performance (e.g., Chang & Kim, 2009; Tienken & Wilson, 2007). When applied researchers can identify specific skills (rather than broad domain scores) that are related

to educational technology resources, such information can be meaningful for answering applied educational research questions; it can lead to resource allocation discussions and also contribute to underscoring factors that affect problem solving. As such, when explanatory variables can be linked with CDMs, they could inform instructional needs and identify whether certain educational resources may be effective.

Results from the real-world data analysis showed that student background variables from large-scale international testing data, such as TIMSS, could be used to link with a CDM. In this application, two sets of items (dichotomous and polytomous) were used as indicators of the latent construct, by fitting IRT models (2PL and GRM), and as such were simultaneously specified to predict attribute mastery. Findings indicated differences in effect sizes by attribute. Depending on the substantive nature of the explanatory variables and their relationship with the attributes, this information could be used to inform instruction and learning needs for students. Moreover, results varied when compared with a two-step approach (covariate model), where latent variables were first estimated, and then used to regress as covariates on the attributes. In this comparison, results showed that estimates from the two-step approach were different for some attributes, possibly due to attenuation from measurement error, signaling the value of simultaneously estimating latent predictors. Similar to findings from the SEM literature (e.g., Bedeian, Day, & Kelloway, 1997), the explanatory approach proposed in this study provides meaningful implications for simultaneously estimating latent factors within this framework.

The simulation study showed that IRT parameters were consistently recovered even with sample sizes of 1,000. While modest differences were found between the three- and five-attribute conditions, attribute-level estimates were stable. Additional attention may be given to the attribute difficulty parameter (intercept, b_k) when multiple predictors are specified. Generally, results from the simulation study indicate that incorporating both latent and observed predictors at the attribute level of the RDINA did not affect estimation greatly. Classification was also consistent, regardless of conditions examined. However, based on findings when multiple latent variables are specified with multiple attributes (i.e., five attributes with two latent and one observed predictor), bias in estimates increased substantially, with mean attribute-level % Bias of 19.4% and 15.0% for even the 2,000 sample size condition. These results indicate a need to investigate the effect of multiple latent predictors when a large number of attributes are used in CDMs.

This study provides implications to expand CDMs to incorporate other measurement models that can be linked using a structural component, resembling SEMs. While the focus of this study was on the RDINA, extensions of this model would yield more flexible and generalized CDMs to allow specification of explanatory predictors. Other extensions to the covariate or explanatory frameworks can also be made. For example, using the von Davier (2010) multilevel diagnostic model, covariates or other explanatory factors can be specified at a higher level, which could be used to answer other substantive questions. Future studies may investigate these extensions to advance explanatory models for CDMs.

Declaration of Conflicting Interests

Funding

Supplemental Material

Supplementary material is available for this article online.

References

Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for measurement error attenuation in structural equation models: Some important reminders. *Educational and Psychological Measurement*, *57*, 785-799.

Chang, M., & Kim, S. (2009). Computer access and computer use for science performance of racial and linguistic minority students. *Journal of Educational Computing Research*, *40*, 469-501.

Dayton, C. M., & Macready, G. B. (1988). A latent class covariate model with applications to criterion-referenced testing. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 129-143). New York, NY: Plenum Press.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York, NY: Springer.

DeCarlo, L. T. (2005). *On the use of covariates in a latent class signal detection model for essay grading*. Unpublished manuscript.

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, *35*, 8-26.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*, 115-130.

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

DiBello, L. V., Roussos, L. A., & Stout, W. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics: Vol. 26. Psychometrics* (pp. 979-1030). Amsterdam, The Netherlands: Elsevier.

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

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

Huang, G. H., & Bandeen-Roche, K. (2004). Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika*, *69*, 5-32.

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.

Lee, Y.-S., Park, Y. S., & Taylan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the U.S. national sample using the TIMSS 2007. *International Journal of Testing*, *11*, 144-177.

Park, Y. S., & Lee, Y.-S. (2014). An extension of the DINA model using covariates: Examining factors affecting response probability and latent classification. *Applied Psychological Measurement*, *38*, 376-390.

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

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society.

Tienken, C. H., & Wilson, M. J. (2007). The impact of computer assisted instruction on seventh-grade students' mathematics achievement. *Planning and Changing*, *38*, 181-190.

Vermunt, J. K., & Magidson, J. (2013). *Technical guide for latent GOLD 5.0: Basic, advanced, and syntax*. Belmont, MA: Statistical Innovations.

von Davier, M. (2005). *A general diagnostic model applied to language testing data* (Research Report RR-05-16). Princeton, NJ: Educational Testing Service.

von Davier, M. (2007). *Mixture general diagnostic models* (Research Report RR-07-32). Princeton, NJ: Educational Testing Service.

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

von Davier, M. (2010). Hierarchical mixtures of diagnostic models. *Psychological Test and Assessment Modeling*, *52*, 8-28.

von Davier, M., Xu, X., & Carstensen, C. H. (2011). Measuring growth in a longitudinal large-scale assessment with a general latent variable model. *Psychometrika*, *76*, 318-336.

Xu, X., & von Davier, M. (2008). *Fitting the structured general diagnostic model to NAEP data* (RR-08-27, ETS Research Report). Princeton, NJ: ETS.