Cognitively Diagnostic Psychometric Models: An Integrative Review

Jianbin Fu

University of Wisconsin, Madison

Yanmei Li

University of Wisconsin, Madison

Correspondence may be sent to:

Jianbin Fu

Educational Testing Service
MS 13-P
Rosedale Road
Princeton, NJ 08541

Email: jfu@ets.org

Cognitively Diagnostic Psychometric Models: An Integrative Review

Abstract

A hierarchical structure and a taxonomy are constructed to classify the existing cognitively diagnostic psychometric models (CDPMs, totaling 62 models) and show their interrelationships. The taxonomy is built based on some important model characteristics with regard to knowledge structure, item structure, and time component. Especially, the characteristics of seven types of attribute structures modeled in CDPMs are discussed in detail. CDPM is loosely defined to include all explicitly and implicitly multidimensional (at test level) psychometric models, most of which are considered as item response models. In addition, this review discusses the building blocks of CDPMs, an upper level CDPM assembled from these blocks and future development of CDPMs.

Key words: cognitively diagnostic psychometric model, item response theory, review, taxonomy, attribute structure

Cognitively Diagnostic Psychometric Models: An Integrative Review

1. The Scope

With the coinciding developments in psychometrics and cognitive science in the past fifty years, more and more researchers are interested in combining these two fields to a new psychometric area, often called Cognitively Diagnostic Assessment (CDA) (Nichols, 1994; Nichols, Chipman, & Brennan, 1995). Some researchers have heralded CDA as the new testing paradigm in the 21st century and have called for increasing research and use of CDA (e.g., Embretson, 1999; Stout, 2002). This new emphasis in psychometrics is motivated by both theoretical and practical needs. On the theoretical side, CDA and its associated psychometric models provide a mechanism for validation of cognitive theories (Embretson, 1999). Societal demand for CDA is even stronger. In accordance with the new educational goal in America of shifting from well educating a small proportion of the population to helping all individuals succeed (National Governors' Association, 1990), amplified by the current *No Child left Behind Act of 2001*, contemporary assessments are expected to provide more informative diagnostic reports to students, parents, teachers and principals that enable successful instructional intervention. This stands in contrast to the traditional role of assessment whose main function was to rank students on an underlying latent variable for purposes of selecting those most capable for advanced educational opportunities (Nichols, 1994; Stout, 2002). Baxter and Glaser (1998), Nichols and Sugrue (1999), and L. B. Resnick and D. P. Resnick (1992) provided extensive discussions on these and other motivations behind CDA.

Research on CDA can be applied to the development of new cognitively diagnostic tests. Nichols (1994), Embretson (1999), Embretson and Groin (2001), and Mislevy, Almond and

Lukas (2003) have proposed frameworks for this purpose. In particular, Mislevy et al.'s (2003) conceptual assessment framework, referred to as evidence-centered design (ECD) and its associated operational framework for the implementation of an assessment, referred to as the four-process model (e.g., Almond, Steinberg, & Mislevy, 2002), have been intensively studied by a group of researchers at the Educational Testing Service. Applying cognitively diagnostic psychometric models (CDPMs) to calibrate tests is an integrated part of each aforementioned framework in developing CDA. Cognitively diagnostic psychometric models (CDPMs)[1] are statistical models developed to determine each examinee's diagnostic status with respect to cognitive components and/or each item's measurement of those cognitive components. While CDPMs are specifically designed for cognitively diagnostic tests, they can also be used on traditional tests, provided a small subset of finely-grained skills can be shown to underlie test performance.

CDPMs have been developed for individual assessment tasks in computer based intelligent tutoring systems (ITS) (e.g., Corbett, Anderson, & O'Brien, 1995; Draney, Pirolli, & Wilson, 1995) as well as stand-alone tests. Two salient differences between these assessments are that (a) an ITS contains much more items than a stand-alone test, and many items are written to measure only one cognitive component (called production in ITS), so that an ITS can also include many more, usually hundreds of, cognitive components that are measured at finer levels than those of stand-alone tests, and (b) ITS is a dynamic learning and assessment system where individuals can be assessed on some cognitive component repeatedly until they master the cognitive component, while standalone tests are static and short (Chipman, Nichols, & Brennan, 1995; Junker, 2001). Because of the inherent differences between these two kinds of assessments, the psychometric models developed for them are quite different too. However, since stand-alone tests reflect the majority of psychometrics to date, almost all CDPMs were developed for stand-alone tests, including the models reviewed here. Then the remaining question is which psychometric models belong to CDPM. Because CDPMs are developed for cognitively diagnostic purposes and are based on cognitive theories, CDPMs should be able to account for the relationships among cognitive components measured by a test and also allow inference of each examinee's mastery status on each cognitive component based on his/her performance on the test. In this respect, the classic test models, unidimensional nonparametric item response theory (IRT) models (e.g., Mokken, 1997, Molenaar, 1997b; van Onna, 2002), linear logistic latent class analysis (e.g., Formann, 1992, 1995; Rost, 1988a, 1988b; Vermunt, 2001), and unidimensional parametric IRT models such as 1, 2 or 3 parameter logistic (or normal ogive) models and their extensions to polytomously-scored items, attitude assessment (unfolding model family, e.g., Andrich, 1997; Hoijtink, 1997; Roberts, Donoghue, & Laughlin, 2000, 2002) and others (for example, the two parameter-like extension of Rasch model, Verhelst & Glas, 1995b; the Rasch model for continuous responses, Muller, 1987), all of which assume only one general latent ability underling a test, are not CDPMs. In addition, the models for response time (e.g., Roskam, 1997; Verhelst, Verstralen, & Jansen, 1997) and multiple attempts on items (e.g., Spray, 1997) are not considered as CDPM here because they are not directly related to cognitive diagnosis. Hereafter we use the term, *ability*, to refer to a broad construct or general cognitive process, while an *attribute* refers to a cognitive component or sub-process. For example, a mixed-number subtraction ability may consist of several finely-grained attributes such as "basic fraction subtraction", "simplify/reduce fraction or mixed number", "separate whole number from fraction", etc. (Mislevy, Almond, Yan, & Steinberg, 1999). In this review, the CDPM is loosely defined to include all explicitly and implicitly multidimensional (at test level) psychometric models, most of which are considered as IRT models. By *implicitly*, we mean that some IRT models only estimate one unidimensional ability in a test, but they either (a) decompose item difficulty parameters into some basic parameters, such as the linear logistic test model (LLTM) family; (b) take the item ordering effect within a test into account to model item response dependence, such as the short term learning models (e.g., the dynamic Rasch model, DRM, Verhelst & Glas, 1995a), or (c) add a time component to assess ability change between test sessions, such as the long term learning models (e.g., the partial credit model for change, PCMC, Fischer, 2001). In such models, the basic parameters can be regarded as attributes required in solving the items, and an item ordering effect and time component can be treated as additional latent dimensions. By *loosely*, we mean that some IRT models were not originally designed for cognitive diagnosis, but are multidimensional IRT models and will be classified here as CDPMs in the broad sense, such as testlet models (Wang, Bradlow, & Wainer, 2002).

Table 1 lists all CDPMs, totaling 62 models, which will be discussed in this review. These models, which include models for dichotomously, polytomously and continuously scored items, multiple time points and multiple strategies, represent the CDPMs most common in the literature as well as many less frequently mentioned ones. However, some models are not included in Table 1 if they are directly related to models already in this table and were not treated intensively by their authors. Note that Table 1, which lists CDPMs in the alphabetical order of their acronyms, serves as a look-up table for the full names and their acronyms. In the text, for most of the CDPMs in Table 1 their full names are not spelled out. This review first presents a hierarchical structure to show the interrelationship among the 62 models, and then adapts and extends a taxonomy proposed by Roussos (1994) to classify and organize the 62 CDPMs based on core components, such as knowledge structure, item structure and time component. This taxonomy as well as the hierarchical structure facilitates the understanding of the psychological meanings of these CDPMs, and makes the differences and similarities among CDPMs very clear in terms of model structure and function. Finally, this review discusses the cognitive limitations of CDPMs, the building blocks of CDPMs and an upper level CDPM assembled from these blocks, and the future research of CDPMs.

<Insert Table 1 About Here>

2. A Hierarchical Structure of CDPMs

Inspired by the ideas of the unified model (DiBello, Stout, & Roussos, 1993), Roussos (1994) proposed a general CDPM form and derived a taxonomy for CDPMs based on it. His general conceptual form is revised and extended to accommodate a much larger range of CDPMs reviewed here. This general form employs a probabilistic model on item responses and takes into account four sources of stochasticity (as discussed first in DiBello et al., 1993): multiple strategy, Q completeness, positivity, and slips. This general conceptual model also provides a framework to describe the interaction of the item structure, knowledge structure, and time component in a cognitively diagnostic assessment. In a CDPM the relationship between item response categories and the skill attributes could be defined through an item-by-category-by-attribute incidence matrix, referred to as a Q matrix (K. K. Tatsuoka, 1990). Attributes contained in a Q matrix are called Q attributes. However, there may be attributes which have influence on responses to an item but are omitted from the Q matrix due to human errors or limited prior knowledge. Those attributes are called non-Q attributes and their roles in an item response should be taken into account. Denote an examinee j's response to item i as x_{ij}^{R}, which could be any real number. Then, the general form of a CDPM is given as the conditional probability of x_{ij}^R on Q attributes, non-Q attributes, strategies and random slip rate at time t:

$$P[x_{ij}^R \mid \boldsymbol{\psi}_j(t), \boldsymbol{\Gamma}_j(t), \mathbf{S}_i, p_i] = (1-p_i)\times \sum_{s_{il}\in\mathbf{S}_i} P[s_{il} \mid \boldsymbol{\psi}_j(t), \boldsymbol{\Gamma}_j(t)] * P[x_{ij}^R \mid \boldsymbol{\Gamma}_j(t), s_{il}, \boldsymbol{\psi}_j(t), NS], \tag{1}$$

Where

$t = 1,...,T$ denotes time t when the measurement occurs

$\boldsymbol{\psi}_j(t)$ is the Q attribute vector (state) for examinee j at time t;

$\boldsymbol{\Gamma}_j(t)$ is the non-Q attribute vector (state) for examinee j at time t;

$\mathbf{S}_i = \{s_{il}\}$ is the strategy vector for solving item i, with element s_{il} denoting the l^{th} strategy, and $l = 1,\cdots,L_i$;

p_i is the probability of a random slip for item i, which is assumed here to be the same across strategies and examinees, and

NS denotes no slip.

Figure 1 organizes all the 62 models in Table 1 to form a partial order structure under this general model in terms of model complexity. The nested relationships among the models are easily seen by looking at their formulas. See Fu (2005, chap. 2, part I) for a detailed description and comparison of the mathematical formulas, and the associated statistical characteristics of all the 62 CDPMs reviewed here as well as their kin models within this hierarchical structure. In addition, Fu (2005, chap. 3, part I) also provided a systematic summary of identification conditions, estimation methods, and model checking techniques for CDPMs.

<Insert Figure 1 About Here>

3. A Taxonomy of CDPMs

Following Roussos (1994), this review separates three main components from the general CDPM form (Equation 1): knowledge structure, item structure, and time component. Time component describes examinees' attribute state change between test sessions. Knowledge structure describes how attribute states are modeled at any given time; two features will be considered here: dimensionality of the general ability and attribute scale. Item structure has four substructures, which describe item type (x_{ij}^R), strategy selection ($P[s_{il} \mid \boldsymbol{\psi}_j(t), \boldsymbol{\Gamma}_j(t)]$), Q incompleteness ($\boldsymbol{\Gamma}_j(t)$) and attribute structure in solving an item given a strategy ($P[x_{ij}^R \mid \boldsymbol{\Gamma}_j(t), s_{il}, \boldsymbol{\psi}_j(t), NS]$), respectively. The three main components provide a framework to classify all the CDPMs in a systematic way. This taxonomy of CDPM is described in detail in the following. Again, the position of a particular model in the taxonomy will be evident from its mathematical formula. See Fu (2005, chap. 2, part I) or the references in Table 1 for these formulas if the classification of any CDPM is not completely clear.

First, the classification of these CDPMs from the knowledge structure perspective is presented. Ability dimensionality refers to how many different attributes are explicitly measured as examinee parameters by a model at the test level (in contrast to at the item or response category level). Unidimensional (U) and multidimensional (M) models are the two main categories. As mentioned before, the unidimensionality in CDPMs is actually implicit multidimensionality. A few models can be classified as either U or M depending on how the measured attributes are interpreted. These models form a new category, named *U or M* (U|M). As for attribute scale, the continuous scale and the dichotomous scale (mastery vs. non-mastery) form the two basic categories, which comprise the majority of CDPMs. BN-WAT is the only model in the polytomous category, and UM and FM-P are the two models containing both dichotomous and continuous attributes. A few models contain both polytomous and continuous attributes. Table 2 shows the cross table of CDPMs by attribute scale and dimensionality.

<Insert Table 2 About Here>

Second, whether a CDPM can accommodate time component is considered. The basic criterion to judge whether a model can incorporate a time component is its ability to handle repeated measurements (between test sessions) designs. Note that in this review time component only refers to multiple test occasions, not item ordering during a test session. The long term learning models (such as SLTM, MRMLC and PCMC) are explicitly designed for this. However, all the other CDPMs, although not explicitly developed for measuring change, can easily handle the repeated measurements test data by equating methods, although the equating methods for the CDPMs including ordinal attributes have not been fully developed yet. For the separate test calibration equating, the linking between tests on different occasions can be performed through the usual anchor item design. Under the concurrent test calibration equating, there are usually three ways to account for time components: the "virtual person" approach, the "virtual item" approach (Fischer, 1997, p. 227), and the combination of the two approaches. Note that most of the long term learning models adopt these techniques too.

Third, the item structure in CDPMs is examined. As mentioned before, the item structure contains four substructures: item type, Q incompleteness, strategy selection and attribute structure. Table 3 presents the cross-classification of CDPMs by the first three aspects of item structure: item type, Q incompleteness, and strategy selection. There are three item score types in CDPMs: dichotomously scored (D), polytomous scored (P) and continuously scored (from negative infinite to positive infinite) (C). As for the other three features, they are discussed in details in the following sections.

<Insert Table 3 About Here>

3.1. Attribute Structure

Attribute structure describes the attribute interaction in a cognitive process involved in an item tackling strategy. The discussion of attribute structures in the literature was not consistent: some of the key terminologies such as compensatory, non-compensatory, conjunctive and disjunctive were defined differently by different authors (see, e.g., Embretson, 1984; Maris 1995, 1999; Mislevy, et al., 2001; Jannarone, 1986; Junker, 1999; Roussos, 1994; van Mechelen, De Boeck, & Rosenberg, 1995; Whitely, 1980). Here we attempt to provide a set of clear definitions of these key terms based on model components and informal psychological theory. The CDPMs reviewed here address five kinds of basic attribute structures (compensatory, partially compensatory, non-compensatory, conjunctive and disjunctive) as well as three kinds of their extended attribute structures, either combining two basic attribute structures or taking into account interaction effect. The classification of CDPMs based on those attribute structures are shown in Figure 2. Figure 2 also shows the conceptual nested relationships (e.g., the extension of the definition of *compensatory* contains that of *disjunctive*) and the structural nested relationships (e.g., a conjunctive model with interaction can be simplified to a conjunctive model without interaction) among those attribute structures. Note that the TBA and MMRM do not appear in Figure 2 because they are the two unidimensional models at the item level and item category level, respectively, which do not have any assumption on attribute structure. In these models interest focuses on individual attribute execution, rather than the composite of them. To demonstrate how these attribute structures are represented in CDPMs in general, the following generic example is used in the subsequent sections. An item i has M_i ordered response categories scored as $(0,...,m,...,M_i-1)$, and there are a total of K attributes, $\boldsymbol{\omega}=\{\omega_1,...,\omega_k,...,\omega_K\}$, measured by a test, where ω_k could be on a continuous, dichotomous or polytomous scale. Note that for simplifying notation a given item administered on multiple test sessions and/or to multiple groups of examinees is treated as two different virtual items. An examinee j's attribute state is denoted as $\boldsymbol{\omega}_j=\{\omega_{j1},...,\omega_{jk},...,\omega_{jK}\}$. Responses to category m ($m\geq 1$) (denoted as x_{ijm}=1), or m and above categories (denoted as $x_{ijm'}$=1) are related to a subset of the K attributes, and denote the attribute index subset as $\mathbf{K}_{im}$ or $\mathbf{K}_i$ under some strategy (note that in a cumulative response function the attribute set should be the same across all its score categories, thus the related attribute index set does not have the subscript m). For dichotomously scored items, x_{ijm} and $\mathbf{K}_{im}$ are shortened to x_{ij}^D and $\mathbf{K}_i$, respectively. For each attribute there is a one-to-one monotonously non-decreasing mapping function for category m ($m\geq 1$) of item i or categories larger than or equal to m, namely, $f_{imk}(\omega_{jk})$ or $f_{im'k}(\omega_{jk})$. RP_{ijm} is the response function mapping $f_{imk}(\omega_{jk})$ and/or the composite of them to the category response probability $P(x_{ijm}=1)$, and $RP_{ijm'}$ is the response function mapping $f_{im'k}(\omega_{jk})$ and/or the composite of them to the cumulative response probability $P(x_{ijm'}=1)$. All RP_{ijm} s and $RP_{ijm'}$ s are (coordinate-wise) monotonously non-decreasing except for the RP_{ijm} s of those polytomous models employing the adjacent category logit function (van der Ark, 2001). In those cases

$RP_{ijm}/(RP_{ijm}+RP_{ij(m-1)})$ is (coordinate-wise) monotonously non-decreasing instead. A subscript will be dropped from RP_{ijm}, $RP_{ijm'}$, $f_{imk}(\omega_{jk})$ and $f_{im'k}(\omega_{jk})$ for short if the functions are the same across this subscript, or do not model on this subscript. For example, for a dichotomous model subscript m can be dropped from these symbols. In addition, in the subsequent discussion only one solving strategy is assumed for an item for simplicity.

<Insert Figure 2 About Here>

3.1.1. Conjunctive (without interaction)

Under a conjunctive attribute structure, a composite task can be accomplished if only if all the subtasks (attributes) have been successfully executed. Translating this principle into psychometric modeling implies that successful executions of all the relevant attributes on category $m(>0)$ (or categories m or above) are required in order to respond to this category (or categories equal to or larger than m). In terms of probability, this means that an item response probability can be written as the function of a joint probability of successfully executing all the required attributes. For an item category response probability, this implies

$$P(x_{ijm}=1\,|\,\boldsymbol{\omega}_j)=RP_{ijm}(P(\{y_{ijmk}=1,k\in\mathbf{K}_{im}\})), \tag{2}$$

while for an item cumulative response probability,

$$P(x_{ijm'}=1\,|\,\boldsymbol{\omega}_j)=RP_{ijm'}(P(\{y_{ijm'k}=1,k\in\mathbf{K}_i\})), \tag{3}$$

where

y_{ijmk} denotes examinee j's status on the execution of attribute k to item i's score category m, with 1 meaning success and 0 failure;

$y_{ijm'k}$ denotes examinee j's status on the execution of attribute k to item i's score categories larger than or equal to m, with 1 meaning success and 0 failure;

$\{y_{ijmk} = 1, k \in \mathbf{K}_{im}\}$ represents the event that all the y_{ijmk} s relevant to the score category m of item i equals 1, and

$\{y_{ijm'k} = 1, k \in \mathbf{K}_{i}\}$ represents the event that all the $y_{ijm'k}$ s relevant to the item i's score categories larger than or equal to m equals 1.

The following is an example of a conjunctive cognitive process. A math item may require (a) manipulations of fractions, radicals, powers or decimals, and (b) algebraic factoring or numerical factorization. The correct response to this item relies on simultaneously successfully applying these two attributes to the item (Hartz, 2002, pp.147-155). All the conjunctive models in Figure 2 are explicit multidimensional models at the test level employing either the continuous or ordinal (including dichotomous and polytomous) scales or both on attributes.

The MLTM-WG and the noisy inputs, deterministic “and” gate (NIDA) family of models (Junker & Sijtsma, 2001) including the UM, FM, MCLCM-CJ, 3PL-MIRT-CJ-D, CRM, BN-WAT and MLTM-WOG, usually make two assumptions to simplify the joint probability: (a) local independence of attribute executions (namely that, an attribute execution does not depend on the execution status of any other attribute) or sequential dependence of attribute executions (that is, the attribute executions form a sequence such that an attribute execution only depends on its previous attribute executions) so that the joint probability can be factorized into the simple probabilities of successfully applying each attribute; (b) self dependence of attribute executions, that is, each of the simple probabilities depends only on the attribute itself, not any other attribute (but could depend on other attribute executions for the sequential dependence cases). Therefore, these models are in the multiplicative form of the probabilities of attribute executions, and can be represented by the generic example as follows:

$$P(x_{ijm}=1\,|\,\boldsymbol{\omega}_j)$$

$$=RP_{ijm}(P(\{y_{ijmk}=1,k\in\mathbf{K}_{im}\}))=RP_{ijm}(\prod_{k\in\mathbf{K}_{im}}P(y_{ijmk}=1))=RP_{ijm}(\prod_{k\in\mathbf{K}_{im}}f_{imk}(\omega_{jk}))\text{, and}$$

$$RP_{ijm}(h)=C_{im}*[(1-p_i)h+\eta_i(1-h)]\text{,} \tag{4}$$

or

$$P(x_{ijm'}=1\,|\,\boldsymbol{\omega}_j)$$

$$=RP_{ijm'}(P(\{y_{ijm'k}=1,k\in\mathbf{K}_i\}))=RP_{ijm'}(\prod_{k\in\mathbf{K}_i}P(y_{ijm'k}=1))=RP_{ijm'}(\prod_{k\in\mathbf{K}_i}f_{im'k}(\omega_{jk}))\text{, and}$$

$$RP_{ijm'}(h)=(1-p_i)h+\eta_i(1-h)\text{,} \tag{5}$$

where

$0\le f_{imk}(\omega_{jk})\le 1$ denotes the conditional probability that examinee j successfully executes attribute k to category m of item i, namely, $f_{imk}(\omega_{jk})=P(y_{ijmk}=1)$;

$0\le f_{im'k}(\omega_{jk})\le 1$ denotes the conditional probability that examinee j successfully executes attribute k to item i's categories equal to or larger than m, namely, $f_{im'k}(\omega_{jk})=P(y_{ijm'k}=1)$;

p_i is the probability of a random slip on item i;

η_i is the probability of solving item i by guessing or other strategies (note that, however, these strategies are not modeled in an explicitly cognitive way, so that Equations 4 and 5 should be still considered as representing single-strategy models), and

C_{im} is a predefined constant for category m of item i.

For continuous attributes in the MLTM-WG and NIDA models (including 3PL-MIRT-CJ-D, CRM, FM, MLTM, UM), $f_{imk}(\omega_{jk})$ or $f_{im'k}(\omega_{jk})$ is the 1-parameter or 3-parameter logistic or probit response function. For example, in the MLTM-WG where the item responses are dichotomous, the response function can be rewritten in the form of Equation 4 as:

$$P(x_{ij}^{D}=1\,|\,\boldsymbol{\omega}_j) = RP\Big(\prod_{k\in\mathbf{K}_i} f_{ik}(\omega_{jk})\Big) = (1-p)\prod_{k\in\mathbf{K}_i} f_{ik}(\omega_{jk}) + \eta\Big[1-\prod_{k\in\mathbf{K}_i} f_{ik}(\omega_{jk})\Big],$$

where $f_{ik}(\omega_{jk})$ is the Rasch response function. For ordinal attributes,

$$f_{imk}(\omega_{jk}) = \sum_{o=0}^{O_k-1} \pi_{imko} * I(\omega_{jk}=o)\,, \tag{6}$$

or

$$f_{im'k}(\omega_{jk}) = \sum_{o=0}^{O_k-1} \pi_{im'ko} * I(\omega_{jk}=o)\,, \tag{7}$$

where

π_{imko} and $\pi_{im'ko}$ denote the conditional probability of correctly executing attribute k on item i's category m, and categories equal to or larger than m, respectively, given level o on attribute k;

O_k is the number of ordered categories in attribute k, and

$I()$ is the indicator function with value 1 if its argument is true, and 0 if false.

For the NIDA models , $f_{imk}(\omega_{jk})$ and $f_{im'k}(\omega_{jk})$ are between 0 and 1 and estimated from the data, and except for UM both p_i and η_i in Equations 4 and 5 are constrained to 0. For example, (a) for the MCLCM-CJ where both items and attributes are dichotomous, its response function in the form of Equations 4 and 6 is given by:

$$P(x_{ij}^{D}=1\,|\,\boldsymbol{\omega}_j)=RP(\prod_{k\in\mathbf{K}_i} f_{ik}(\omega_{jk}))=\prod_{k=\mathbf{K}_i}\pi_{ik}{}^{\omega_{jk}}r_{ik}{}^{(1-\omega_{jk})},$$

where

π_{ik} denotes the conditional probability of correctly executing attribute k on item i given attribute k is mastered;

r_{ik} represents the conditional probability of correctly executing attribute k on item i given attribute k is not mastered, and

ω_{jk} is examinee j's state on dichotomous attribute k (0 = nonmaster, 1 = master).

(b) the response function of the binomial FM-P (Templin, He, Roussos & Stout, 2003) can be written in the form of Equations 4 and 6 as:

$$P(x_{ijm}=1\,|\,\boldsymbol{\omega}_j)=RP_{im}(\prod_{k\in\mathbf{K}_i} f_{ik}(\omega_{jk}))=RP_{im}(\prod_{k\in\mathbf{K}_{i,-k_{rs}}}\pi_{ik}^{\omega_{jk}}r_{ik}^{(1-\omega_{jk})}*f_i(\omega_{jk_{rs}}))$$

$$=RP_{im}(f_i(\omega_{jk_{rs}})*\pi_i^{*}*\prod_{k\in\mathbf{K}_{i,-k_{rs}}}r_{ik}^{*(1-\omega_{jk})})$$

$$=\binom{M_i-1}{m}[f_i(\omega_{jk_{rs}})*\pi_i^{*}*\prod_{k\in\mathbf{K}_i}r_{ik}^{*(1-\omega_{jk})}]^{m}*[1-f_i(\omega_{jk_{rs}})*\pi_i^{*}*\prod_{k\in\mathbf{K}_i}r_{ik}^{*(1-\omega_{jk})}]^{M_i-m-1},$$

where

k_{rs} denotes the index for the residual attribute;

$\mathbf{K}_{i,-k_{rs}}$ denotes the set of attribute indexes excluding the residual attribute;

$\pi_i^{*}=\prod_{k\in\mathbf{K}_{i,-k_{rs}}}\pi_{ik}^{\omega_{jk}}$;

$r_{ik}^{*}=\dfrac{r_{ik}}{\pi_{ik}}$ $(0\le r_{ik}^{*}\le 1)$ and

$f_i(\omega_{jk_{rs}})$ is the Rasch response function.

Note that in this case the specified constant, C_{im}, equals $\binom{M_i - 1}{m}$, the number of response combinations of virtual dichotomously scored items given the number of correct responses. A special note should be given to BN-WAT (Mislevy et al., 2001). In Mislevy et al. (2001) the conjunctive and disjunctive structures were defined totally different from the traditional view, and the compensatory version of BN-WAT is actually a conjunctive model. To see that, the compensatory BN-WAT is actually the discrete version of 2PL-MIRT-C-P, that is, the attributes are on ordinal scale instead of on continuous scale and could be written as

$$P(x_{ijm'} = 1 \mid \boldsymbol{\omega}_j) = \frac{\exp(\sum_{k \in K_i} a_{ik}\omega_{jk} - b_{im'})}{1 + \exp(\sum_{k \in K_i} a_{ik}\omega_{jk} - b_{im'})} \tag{8}$$

where

ω_{jk} s are on the ordinal scale;

a_{ik} is the discrimination parameter of attribute k on item i, and

$b_{im'}$ is the difficulty parameter of item i 's categories larger than or equal to m.

This is not a compensatory model because under a limit number of attribute categories attributes can not be totally compensated by each other (see more discussion in section 3.1.3). It can be proved that Equation 8 is equivalent to

$$P(x_{ijm'} = 1 \mid \boldsymbol{\omega}_j) = \pi^*_{im'} * \prod_{k \in \mathbf{K}_i} \sum_{o=0}^{O_k - 1} r^*_{im'ko} I(\omega_{jk} = o)$$

which can be seen as a polytomous extension (on both item response and attribute scale) of FM-D (see Fu, 2005, pp. 146-148).

For the deterministic inputs, noisy "and" gate (DINA) family of models (Junker & Sijtsma, 2001), the assumption that they make to simplify the joint probability in Equations 2 and 3 is different from those of the MLTM-WG and NIDA models. In particular, instead of simplifying the joint probability into the product of the independent probability of each element in the joint set, DINA models directly categorize the joint probability into a categorical variable by making assumptions. For example, for dichotomous attributes and items it is commonly assumed in DINA models that the masters of all the required attributes will successfully execute these attributes with probability 1, and the nonmasters of any one of these attributes will certainly fail to execute successfully all these attributes. And then based on their discrete joint probabilities examinees are classified into groups, which has the same number of categories as those used in partitioning the joint probability. The response probabilities for examinees within a group are not necessarily the same, and the within-group heterogeneity is accounted for by a response function depending on attributes other than those used to classify examinees. Formally, in terms of the generic example DINA models can be represented as:

$$
\begin{aligned}
&P(x_{ijm} = 1 \mid \boldsymbol{\omega}_j) \\
&= RP_{im}(P(\{y_{ijmk} = 1, k \in \mathbf{K}_{im}^c\}), \{f(\omega_{jh}), h \in \mathbf{K}_{im}^d\}) \\
&= \sum_{g=0}^{G-1} \vartheta_{img}(\{\omega_{jh}, h \in \mathbf{K}_{im}^d\}) * I(SP(P(\{y_{ijmk} = 1, k \in \mathbf{K}_{im}^c\})) = g) \\
&= \sum_{g=0}^{G-1} \vartheta_{img}(\{\omega_{jh}, h \in \mathbf{K}_{im}^d\}) * I(SP(\{\omega_{jk}, k \in \mathbf{K}_{im}^c\}) = g),
\end{aligned}
\tag{9}
$$

or

$$
\begin{aligned}
&P(x_{ijm'} = 1 \mid \boldsymbol{\omega}_j) \\
&= RP_{im'}(P(\{y_{ijm'k} = 1, k \in \mathbf{K}_i^c\}), \{f(\omega_{jh}), h \in \mathbf{K}_i^d\}) \\
&= \sum_{g=0}^{G-1} \vartheta_{im'g}(\{\omega_{jh}, h \in \mathbf{K}_i^d\}) * I(SP(P(\{y_{ijm'k} = 1, k \in \mathbf{K}_i^c\})) = g) \\
&= \sum_{g=0}^{G-1} \vartheta_{im'g}(\{\omega_{jh}, h \in \mathbf{K}_i^d\}) * I(SP(\{\omega_{jk}, k \in \mathbf{K}_i^c\}) = g),
\end{aligned}
\tag{10}
$$

where

$\mathbf{K}_{im}^{c}$ and $\mathbf{K}_{i}^{c}$ denote the set of indexes of attributes used to classify examinees via the ordinal joint probability;

$\mathbf{K}_{im}^{d}$ and $\mathbf{K}_{i}^{d}$ denote the set of indexes of attributes used to account for the within-group heterogeneity of item responses;

$SP()$ is a function of classification attributes for categorizing the joint probability of attribute executions;

G is the number of categories (examinee groups) in the categorized joint probability;

$\vartheta_{img}()$ is the response function for examinees in group g to category m of item i, and

$\vartheta_{im'g}()$ is the response function for examinees in group g to item i's categories equal to or larger than m.

The DINA models in Figure 2 can be separated into three groups. The first group is called the within-group homogenous DINA models, including the BN-WOAT, BS, RS, LPOCM-D, SLM and HICLAS-CJ. Those models are within-group homogenous because an item response probability is the same for all the examinees within a group. Because of this, there is no attribute needed to account for within-group response heterogeneity; thus all the attributes are used to classify examinees. In addition, these models all deal with dichotomous attributes and dichotomously-scored items, and make the common assumption regarding the joint probability as stated above. Their basic formula in the form of Equation 9 can be written as:

$$P(x_{ij}^{D}=1|\boldsymbol{\omega}_j) = RP_i(SP(P(\{y_{ijk}=1, k\in\mathbf{K}_{im}\}))) = RP_i(\prod_{k\in\mathbf{K}_i}\omega_{jk}) = \begin{cases} 1-p_i & \prod_{k\in\mathbf{K}_i}\omega_{jk}=1 \\ \eta_i & \prod_{k\in\mathbf{K}_i}\omega_{jk}=0 \end{cases},$$

where

$SP(P(\{y_{ijk} = 1, k \in \mathbf{K}_{im}\})) = \prod_{k \in \mathbf{K}_i} \omega_{jk}$;

p_i denotes the probability of a slip on item i for the group of examinees who master all the attributes required by item i, and

η_i is the probability of solving item i by guessing for the group of examinees who do not master all of the required attributes.

The second group of DINA models are called within-group heterogeneous DINA model, and contains the MixRM and MLT, where an item response probability could be different for examinees within a group. In the MixRM and MLT, examinees are first grouped based on a grouping attribute, and then the within-group heterogeneity in item responses is modeled via the partial credit model (MixRM), or the graded response model (in the probit form) (MLT) (assuming the grouping attribute is different from the attributes in the IRT models). The continuous attributes used to differentiate examinees' response probabilities within a group are considered as on the same dimension in the MixRM, but on different dimensions in MLT (Uebersax, 1999). As an example, the formulation of the MixRM in the form of Equation 9 is given as:

$$P(x_{ijm} = 1 | \boldsymbol{\omega}_j)$$

$$= RP_{im}(P(y_{ijm} = 1), \{f(\omega_{jh}), h \in \mathbf{K}^d\})$$

$$= RP_{im}(P(y_{ijm} = 1), \boldsymbol{\omega}_j^d)$$

$$= \sum_{g=0}^{G-1} \vartheta_{img}(\omega_{jg}^{d}) * I(SP(P(y_{ijm}=1)) = g)$$
$$= \sum_{g=0}^{G-1} \vartheta_{img}(\omega_{jg}^{d}) * I(SP(\omega_{j}^{c}) = g) \quad (11)$$
$$= \sum_{g=0}^{G-1} \vartheta_{img}(\omega_{jg}^{d}) * I(\omega_{j}^{c} = g),$$

where

$f(\omega_{jh}) = \omega_{jh}$;

$SP(P(y_{ijm}=1)) = \omega_{j}^{c}$, namely, the ordinal attribute used to classify examinees;

$\boldsymbol{\omega}_{j}^{d}$ is a set of continuous attributes to account for within-group heterogeneity of item responses;

ω_{jg}^{d} is an element in $\boldsymbol{\omega}_{j}^{d}$ relevant to group g, and

$\vartheta_{img}(\omega_{jg}^{d}) = \dfrac{\exp(m\omega_{jg}^{d} - b_{img})}{\sum_{v=0}^{M_i-1} \exp(v\omega_{jg}^{d} - b_{ivg})}$ is the response function of the partial credit model.

HRLC and HYBRID compose the third group, the hybrid DINA models, where both homogenous and heterogeneous latent groups are allowed, and the within-group difference is modeled by the partial credit model and 3PL, respectively. HRLC is a generalization of the MixRM that introduces homogenous latent groups. In the HYBRID model, all but one of the groups are homogenous, and these groups are partitioned based on the probability of successfully executing a (grouping) attribute. The only one heterogeneous group can be treated as a “don’t know” group in the sense that the response patterns of the examinees in this group do not provide enough information to calculate the execution probability in order to assign them a latent group membership.

3.1.2. Disjunctive (without interaction).

In the disjunctive case, successfully executing any one of the attributes related to an item is sufficient to get the item right. For example, in a spatial visualization test examinees are asked to judge whether a three-dimensional target is the same as the original one after rotation and change of one or more key features. There are two ways to solve the items: (a) rotating the target mentally the required degree and finding the match, or (b) employing analytical reasoning to detect feature matches without performing rotation (Mislevy & Verhelst, 1990). In terms of the joint probability of attribute executions, the disjunctive attribute structure can be written as:

$$P(x_{ijm}=1 \mid \boldsymbol{\omega}_j) = RP(P(\forall y_{ijmk}=1, k \in \mathbf{K}_{im})),$$

or

$$P(x_{ijm'}=1 \mid \boldsymbol{\omega}_j) = RP(P(\forall y_{ijm'k}=1, k \in \mathbf{K}_i)),$$

where $\forall y_{ijmk}=1$ ($\forall y_{ijm'k}=1$) denotes that any one of y_{ijmk} s ($y_{ijm'k}$ s) equals 1. To simplify the joint probability an attribute execution is assumed to be locally independent and self-dependent. Then the joint probability can be factorized as:

$$P(x_{ijm}=1 \mid \boldsymbol{\omega}_j) = RP(P(\forall y_{ijmk}=1, k \in \mathbf{K}_{im}))$$

$$= RP(1 - P(\{y_{ijmk}=0, k \in \mathbf{K}_{im}\}))$$

$$= RP(1 - \prod_{k \in K_{im}} P(y_{ijmk}=0)) \qquad (12)$$

$$= RP(1 - \prod_{k \in \mathbf{K}_{im}} (1 - P(y_{ijmk}=1 \mid \omega_{jk})))$$

$$= RP(1 - \prod_{k \in \mathbf{K}_{im}} (1 - f_{imk}(\omega_{jk})))$$

$$= 1 - \prod_{k \in \mathbf{K}_{im}} (1 - f_{imk}(\omega_{jk})),$$

where $f_{imk}(\omega_{jk}) = P(y_{ijmk} = 1 \mid \omega_{jk})$, and $RP(h) = h$. A similar decomposition can be made on the joint probability for $P(x_{ijm'} = 1)$. HICLAS-DJ and MCLCM-DJ are the two disjunctive models in Figure 2. They both deal with dichotomous attributes and items, and take the form of Equation 12. They are, however, different in what values $P(y_{ijmk} = 1 \mid \omega_{jk})$ can take; in HICLAS-DJ this probability is constrained to have a value of either 0 (for an attribute nonmaster, ω_{jk} =0) or 1 (for an attribute master, ω_{jk} =1), while in MCLCM-DJ it could be any values between 0 and 1 (inclusive).

3.1.3. Compensatory (without interaction)

Under the compensatory attribute structure, ANY attribute can be compensated by the other attribute(s) to ANY DEGREE (Roussos, 1994) so that the expected performance on an item would not change. According to this criterion, the disjunctive models discussed above should be considered as compensatory models. Another way to achieve “total compensation” in psychometric modeling is to employ the usual logistic or probit function with the (weighted) sum of all the related continuous attributes as the argument, such as the well known 2PNO-MIRT-C-D (McDonald, 1997) and 3PL-MIRT-C-D (Reckase, 1997). As a real case example, the performance on a reading comprehension test may reflect the additive effect of the reading attribute and the degree of familiarity with the reading content. The basic formula for this type of compensatory models can be written as: for the item category response function,

$$P(x_{ijm} = 1 \mid \boldsymbol{\omega}_j) = RP_{ijm}\Big(\sum_{k \in \mathbf{K}_{im}} f_{imk}(\omega_{jk})\Big), \quad (13)$$

and for the item cumulative category response function,

$$P(x_{ijm'} = 1 \mid \boldsymbol{\omega}_j) = RP_{ijm'}(\sum_{k \in \mathbf{K}_i} f_{ik}(\omega_{jk})) \,. \tag{14}$$

In both cases, ω_{jk} s are on the continuous scale, and RP_{ijm} and $RP_{ijm'}$ are either the logistic or probit function.

For the explicit multidimensional IRT at response category level models (2PNO-MIRT-C-D, 3PL-MIRT-C-D, 3PNO-MIRT-C-D, MRM, MRMLC, SLTM, 2PNO-MIRT-C-P, 2PL-MIRT-C-P, MPRM, MRCMLM, MULTRA, TM, and 2PNO-MIRT-C-C) each ω_{jk} is treated as an examinee parameter such that $f_{ik}(\omega_{jk}) = a_{ik}\omega_{jk}$ and $f_{imk}(\omega_{jk}) = a_{imk}\omega_{jk}$, where a_{ik} (a_{imk}) is the (specified or freely estimated) discrimination parameter of attribute k on (category m of) item i. Note that in a cumulative response function the discrimination parameter of an attribute in an item is assumed invariant across the score categories of this item, so that $f_{ik}()$ in Equation 14 does not have the subscript m'. For those polytomous models employing the adjacent category logit function (van der Ark, 2001) including MRCMLM, MPRM and MULTRA, a sufficient condition for them to be compensatory models is that the relevant attributes are the same across an item's response categories and their discrimination parameters a_{imk} are invariant within an item category. A well known example is the multidimensional partial credit model which is a sub-model of MRCMLM, MPRM and MULTRA. However, this is not a necessary condition; there are other forms of parameterization which lead to compensatory models also. Depending on the parameterization, these models could also be partially compensatory or conjunctive models (see the discussion in sections 3.1.5 and 3.1.6).

The implicit multidimensional IRT models at the response category level can be separated into three groups. The first group is the LLTM family of models (LLRA, LLTM, LPCM, LRSM, MLPCM, MLRSM, and MIRID), which treat each attribute as an item parameter, that is, $f_{ik}(\omega_k) = \beta_{ik}\omega_k$, where ω_k is the basic parameter related to attribute k (ω_{ik} for MIRID because the basic parameters in MIRID can be item-specific) and β_{ik} is the specified (freely estimated for MIRID) weight of basic parameter k on item i. An examinee parameter in these models can be considered as a quantity related to the composite of the basic parameters. All the RP_{ijm} s in this group (the subscript j is retained because an item response function depends on an examinee's general ability) employ the logistic function. The MRCMLM is a parent model of LLTM model family, thus can decompose an item difficulty into several basic parameters too. The polytomous models in this group (LPCM, LRSM, MLPCM, MLRSM and MRCMLM) employ the adjacent category logit function too so that the sufficient condition discussed above on model parameterization applies here also in order for them to be compensatory models. In SALTUS, the response function given a cognitive stage is actually an LLTM. In this regard, each item can be thought of as having several virtual items, each for a cognitive stage. However, unlike LLTM where examinees' class memberships (namely, the specification of $\boldsymbol{\beta}$ matrix) are known as a priori, the cognitive stage of each examinee in SALTUS is subject to estimation. The second group is the short term learning models (DRM, DTM, and MSLM), where besides the general ability the item response functions take into account examinees' learning effect within a test session due to, for example, responses to the previous items, and/or the feedbacks prior to answer an item. In these models, the learning effect is considered as additional latent dimensions and all the RP_i s have the logistic form. For example, MSLM can be written in the form of Equation 13,

$$P(x_{ij}^D = 1) = RP_i\left(\sum_{k=1}^{2} f(\omega_{jk})\right) = RP_i\left(\sum_{k=1}^{2} \omega_{jk}\right) = \frac{\exp[\omega_{j1} + (i-1) * \omega_{j2} - b_i]}{1 + \exp[\omega_{j1} + (i-1) * \omega_{j2} - b_i]},$$

where $f(\omega_{jk}) = \omega_{jk}$, ω_{j1} and ω_{j2} are examinee j's general ability and learning rate respectively, and b_i is the difficulty parameter of item i. The final group consists of only one model, PCMC, which is a long term learning model, where examinees' general ability change between test occasions is estimated. If we treat the learning gain as a separate latent dimension then the PCMC is actually a multidimensional partial credit model.

3.1.4. Non-compensatory with Interaction

As its name suggests non-compensatory structure is the opposite of compensatory structure: Under non-compensatory attribute structure ANY attribute can not be COMPLETELY compensated by the other attribute(s) in terms of the expected performance on a composite task. The conjunctive models discussed above are a kind of non-compensatory models. Here discussed is another kind of non-compensatory models which are also capable of taking into account attribute interaction effect. The CMDLV and MixRM-LL are the two CDPMs developed under hierarchical log-linear modeling, where an item response probability is the function of the additive effect of item response, attributes and their interactions. The attributes are ordinal so that the compensation from the other attributes is limited. Hence, CMDLV and MixRM-LL are non-compensatory models. Furthermore, the CMDLV and MixRM-LL can explain higher order attribute interaction effects in item responses, because under hierarchical log-linear modeling adding these higher order interaction terms is straightforward.

3.1.5. Partially Compensatory (without interaction)

The partially compensatory structure is an intermediate place between the compensatory structure and the non-compensatory structure. In this structure, SOME BUT NOT ALL attribute(s) can be COMPLETELY compensated by the other attribute(s). Under certain parameterizations those polytomous models using the adjacent category logit function such as the LLTM family models (LPCM, LRSM, MLPCM and MLRSM) and the generalized Rasch family models (MRCMLM, MPRM and MULTRA) are partially compensatory models. For example, the response function of MULTRA can be written as

$$P(x_{ijm} = 1 \mid \boldsymbol{\omega}_j) = \frac{\exp(\sum_{k \in \mathbf{K}_i} a_{imk}\omega_{jk} - b_{im})}{\sum_{h=0}^{M_i - 1} \exp(\sum_{k \in \mathbf{K}_i} a_{imk}\omega_{jk} - b_{im})}, \quad (15)$$

where ω_{jk} is on the continuous scale and $m > 0$. If for an item i, $\mathbf{K}_i$ = {1, 2}, M_i = 3, and all a_{imk}s equal to 1 except $a_{i12} = 0$, then attribute 1 (ω_{j1}) can be completely compensated by attribute 2 (ω_{j2}) but not the other way around. However, it is yet to be determined the exact parameterization conditions for these models to be compensatory, partially compensatory or non-compensatory (or conjunctive).

3.1.6. Conjunctive with Interaction

Unlike those conjunctive models discussed in section 3.1.1, a conjunctive model with interaction can explicitly estimate attribute interaction and/or attribute execution interaction. Specifically, this kind of models can test the following dependences. First, an attribute execution depends on the execution status of other attributes (or the status of previous executions of this attribute), that is, $P(y_{imk} \mid \{y_{imk'}, k' \neq k\}) \neq P(y_{imk})$, where $k, k' \in \mathbf{K}_{im}$. If this is the case, it also means that this attribute execution depends on other attributes, namely $P(y_{imk} \mid \boldsymbol{\omega}) \neq P(y_{imk} \mid \omega_k)$. Furthermore, it is also allowed in some of these models that besides the execution status of other attributes an attribute execution depends on other attributes (see Hoskens & De Boeck, 2001; also see the discussion of "higher mental process" in Samejima, 1995), which can be formally represented as $P(y_{imk} \mid \{y_{imk'}, k' \neq k\}, \boldsymbol{\omega}) \neq P(y_{imk} \mid \{y_{imk'}, k' \neq k\}, \omega_k)$. To measure these dependences observations on each attribute execution are required. Note that, as mentioned before, the NIDA models can handle sequentially dependent attribute executions. However, this dependence is an assumption built into the tests and the models, and consequently, it cannot be verified by the NIDA models.

In the CRK and LRM, the response functions are set up at the test level instead of item level. In other words, the response functions are item pattern response functions. A correct response to an item is considered as the successful executions of all the attributes related to this item. LRM is an implicit multidimensional model where all items are assumed to measure one same attribute and the interaction effect among attribute executions (or called as learning effect) can be taken into account to any higher order. In the MRCMLM, MPRM, LPCM and MULTRA their item response functions can be considered as item category pattern response functions and have the similar form as those in the CRK and LRM: they all belong to the divide-by-total models (Thissen & Steinberg, 1986). Therefore, under proper parameterizations these models can model the conjunctive structure and/or interaction at item level just as the CRK and LRM do at test level. Actually, MRCMLM is a parent model of CRK and LRM (see Figure 1; also see Fu, 2005, p. 30, for a demonstration). For example, if we specify Equation 15 as follows: $\mathbf{K}_i$ = {1, 2}, M_i = 4, and all a_{imk} s equal to 1 except $a_{i11} = 0$ and $a_{i22} = 0$, then Equation 15 is a conjunctive model with interaction, where the second response category requires successful execution of attribute 1, the third one requires that of attribute 2, and the fourth one requires that of both attributes 1 and 2. However, in this case the joint probability of successful execution of both attributes is not the product of the simple probability of successful execution of each attribute; there is interaction effect between the two attributes. Note that when MRCMLM, MPRM, LPCM and MULTRA are parameterized to be conjunctive models the scoring of item response categories may only have the nominal property instead of the ordinal property under their typical uses. MCUIRT is a Markov chain 3-parameter logistic (3PL) model where the item response function is the response function of 3PL multiplied by a transition parameter determined by the answer status (right or wrong) of the pervious item in a test. Therefore, MCUIRT is actually a short term learning model which takes into account the learning effect coming from the pervious execution of an attribute.

3.1.7. Compensatory and Conjunctive

The models in this category include both compensatory and conjunctive attribute structure. The 3PL-MIRT-C&CJ-D has a parameter to control the relative proportion of the compensatory component and the conjunctive component. However, this model seems to lack the cognitive foundation. Basically, GLTM (-WG and –WOG) is a function of the product of several LLTM models, where each LLTM represents the probability of an attribute execution. Since LLTM is a compensatory model, GLTM is a hybrid model having mixed compensatory and conjunctive attribute structure. It can be shown that the response functions of GLTM can be parameterized using MRCMLM so that MRCMLM is a parent model of GLTM (see Figure 1; also see Fu, 2005, p. 33, for a demonstration). Therefore, under proper parameterization MRCMLM could be a model containing both compensatory and conjunctive structures.

3.2. Q incompleteness

Q incompleteness means that a model takes into account the attributes outside those specified in Q matrix. These non-Q attributes are often called residual attributes, and usually their cumulative effect is represented by one composite attribute in the model. As a result, it makes sense to use the continuous scale for the composite residual attribute. Based on this principle, all CDPMs with all continuous attributes except for some models requiring observations on subtasks or for special purposes (see below) are Q incomplete, because the composite residual attribute is easy to incorporate into these models without changing their mathematical forms. Some models requiring observational data on subtasks (such as the MIRID, MLTM-WOG, GLTM-WOG) are Q complete, because the executions on the residual attribution are not observable. However, MLTM-WG and GLTM-WG are Q incomplete because in these models the aggregate effect of all the non-Q attributes (including guessing) is included in the guessing term. Most CDPMs including discrete attributes are Q complete, because inserting a continuous residual attribute into these models will change their formulations and estimation processes. The exceptions are those models explicitly taking into account the continuous residual attribute, such as UM and FM. Some implicit multidimensional models with all continuous attributes are developed for special purposes: (a) modeling short-term learning (DRM, DTM, MSLM, LRM and MCUIRT); (b) modeling learning stages (SALTUS), and (c) modeling long-term learning (PCMC). Therefore, the concept of Q incompleteness is not applicable for these models.

3.3. Cognitive strategy

There are only three models explicitly designed for multiple cognitive strategies: the MLTM-WG, UM, and SLM. However, cognitive strategy can be thought of as a discrete attribute in a CDPM. Consequently, those CDPMs having discrete attribute(s), which include all models in the "D", "P", "C&P", and "C&D" rows in Table 2 except TBA and RS[2], can be used to model multiple cognitive strategies. Note the explanatory difference in modeling multiple strategies between the models with disjunctive structure (HICLAS-DJ and MCLCM-DJ) and the remaining models. In the disjunctive structure models, what is estimated is an examinee's posterior probability of mastery status of all the involved strategies on an item, while in the remaining models it is an examinee's posterior probability of applying each of the strategies involved on this item. Based on the posterior probability the strategy mastery status or the actual strategy applied on this item can be determined, for example, choosing the one having highest posterior probability (Note that the prior probabilities are just the proportions of examinees in each group, often called group weights). To see the difference, an example is given. Suppose an item can be solved by two strategies. In a disjunctive structure model the two strategies are represented by two dichotomous attributes (1 denotes mastery of a strategy, and 0 non-mastery), respectively. The disjunctive structure model estimates the posterior probabilities of the mastery statuses on these two attributes for each examinee. In MCLCM-DJ also estimated are the probabilities of correctly applying a strategy conditional on the mastery status of this strategy, while in HICLAS-DJ this conditional probability for a non-master is assumed to be 0, and 1 for a master, and then the only scenario that an examinee will fail this item is that he/she does not master both strategies[3]. By contrast, in the remaining models these two strategies are represented by one dichotomous attribute (0 denotes applying strategy 1, and 1 applying strategy 2). Those models estimate an examinee' posterior probability of applying each of the two strategies to this item, and some of them also estimate the associated probability of solving this item by a strategy. In this regard, some constraints due to the ordinal property of attributes may be relaxed in these models, for example, the restriction that the probability of successfully executing an attribute conditional on a high level of this attribute should be no less than that conditional on a low level.

As for the CDPMs with all continuous attributes, they can not handle multiple cognitive strategies naturally, unless they are explicitly developed for multiple strategies. Currently, only MLTM-WG can serve this purpose in this category of CDPMs.

4. Discussions and Future Research of CDPMs

4.1. CDPMs and Cognitive Theories

The concept of an attribute in CDPMs is often vague for most cognitive theories. Most cognitive theories are so complex that a high level cognitive process usually contains hundreds of basic cognitive elements. However, it is computationally infeasible and practically unattractive for a CDPM to include too many cognitive elements. The implications of this are that (a) a CDPM is usually not an exact statistical modeling of a cognitive theory, but, at best, a simplification, and (b) an *attribute* in a CDPM is a concept related to a relatively high level cognitive process which may consist of many basic cognitive elements. In contrast, the psychometric models developed for computer based intelligent tutoring systems are much more complicated than CDPMs in terms of the number of basic cognitive elements and cognitive structures considered, because of the dynamic testing and computational nature of intelligent tutoring systems (Chipman, Nichols, & Brennan, 1995; Junker, 2001). On the other hand, the general attributes in a CDPM have the advantage of communicative practicality: The diagnostic information is easily understood by teachers, students and parents, and compatible with the instructional or learning decisions they are likely to make (Chipman, Nichols, & Brennan, 1995).

As discussed in section 3.1.1, the conjunctive models make the two assumptions to simplify a joint probability of attribute executions: local independence or sequential dependence of attribute executions and self dependence of attribute executions. Under most circumstances these two assumptions are reasonable and make conjunctive models tractable. However, if an underlying theory holds against these assumptions and supports dependence of attribute executions (on other attribute executions and/or other attributes), an experiment should be designed so that observations on each attribute execution can be obtained and then the conjunctive models with interaction can be applied to test the dependence. At the first thought, it seems awkward to allow an attribute execution conditional on other attributes besides itself. However, considering that an attribute in CDPMs often refers to a relatively high level cognitive process then this kind of dependence should not be surprising.

4.2. Future Development of CDPM

4.2.1. Developing new CDPMs and elaborating existing CDPMs

The general model proposed in section 2 is conceptual because it does not provide the specific forms to model strategy selection and attribute structures. However, after examining the structures of the existing CDPMs in the above sections we now can construct concrete CDPMs at the very high level. First, we may separated some basic elements out of those existing CDPMs : (a) attribute scales (dichotomous, polytomous, and continuous), (b) item types (dichotomous, polytomous, and continuous), (c) item response functions (logistic and probit), (d) number of item parameters (1, 2 and 3), (e) attribute structure (compensatory, partially compensatory, non-compensatory, conjunctive, disjunctive, and interaction), (f) dimensionality of ability (multidimensional, unidimensional), (g) decomposition of item difficulty (decomposed, non-decomposed), and (h) mixed latent class model and IRT (mixed, non-mixed). Denote that Element A is a higher level element than Element B if Element A contains Element B. Thus, "polytomous item" is a higher level element than "dichotomous item", and the same applies for multidimensional vs. unidimensional, 3 item parameters vs. 2 item parameters, 2 item parameters vs. 1 item parameter, decomposed vs. non-decomposed, and mixed vs. non-mixed. Then a more general model contains more of the above elements, or more higher level elements. This is the principle under the hierarchical structure of the CDPMs presented in section 2 (also see Rost, 2001, for an example of building up a hierarchical structure in the Rasch family of models in the similar way). A very general model can be constructed as mixed 3-parameter compensatory and conjunctive, multidimensional and decomposed logistic (or probit) model with interaction and continuous attributes for polytomously scored items. The basic idea behind this model is similar to that of GLTM-WOG. Specifically, in this model an item response involves a conjunctive cognitive process with high level attributes, and furthermore, the executions of the high level attributes are the results of compensatory processes with low level attributes. Assuming that (a) an item i has M_i response categories scored as $0,1,\cdots,M_i-1$, and all categories require K high level attributes; (b) the executions of high level attributes are sequentially dependent, and (c) the interactions only exist in the compensatory parts of the model, then the logistic form of this model can be formally written as:

$$
\begin{aligned}
P(x_{ijm'}=1\,|\,\boldsymbol{\Theta}_j,g) &= P(y_{ijm'1}=1,\cdots,y_{ijm'k}=1,\cdots,y_{ijm'K}=1\,|\,\boldsymbol{\Theta}_j,g)\\
&= c_{igm'}+(1-c_{igm'})\prod_{k=1}^{K}P(y_{ijm'k}=1\,|\,y_{ijm'1}=1,\cdots,y_{ijm'(k-1)}=1,\boldsymbol{\Theta}_j,g),
\end{aligned}
\tag{16}
$$

and

$$P(y_{ijm'k}=1 \mid y_{ijm'1}=1,\cdots,y_{ijm'(k-1)}=1,\mathbf{\Theta}_j,g)=\frac{\exp(k_{ijgm'k}-\sum_{p=1}^{P}\beta_{igm'pk}\xi_p)}{1+\exp(k_{ijgm'k}-\sum_{p=1}^{P}\beta_{igm'pk}\xi_p)},$$

$$\sum_{g=1}^{G}\pi_g=1,$$

$$P(x_{ij0'}=1 \mid \mathbf{\Theta}_j,g)=1,$$

$$P(y_{ij0'k}=1 \mid y_{ijm'1}=1,\cdots,y_{ijm'(k-1)}=1,\mathbf{\Theta}_j,g)=1,$$

where

$x_{ijm'}=1$ denotes that examinee j's response to item i is in a category equal to or larger than m $(m=0,\cdots,M_i-1)$;

$y_{ijm'k}=1$ denotes the examinee j's successful execution of high level attribute k to categories equal to or larger than m of item i;

g represents a latent group;

π_g is the proportion of examinees in the latent group g;

$\mathbf{\Theta}_j=\{\theta_{jv}\}$, is the low level attribute vector of examinee j with order V;

$$k_{ijgm'k}=\sum_{v=1}^{V}a_{igm'kv}\theta_{jv}q_{im'kv}+\sum_{v=1}^{V}\sum_{v^*=v+1}^{V}a_{igm'kvv^*}\theta_{jv}\theta_{jv^*}q_{im'kv}q_{im'kv^*}+\cdots+a_{igm'k12\cdots V}\theta_{j1}\theta_{j2}\cdots\theta_{jV}q_{im'k1}q_{im'k2}\cdots q_{im'kV};$$

$q_{im'kv}$ equals to 1 if low level attribute v is required by the execution of high level attribute k to the categories equal to or larger than m of item i, 0 otherwise;

a is the discrimination parameter;

$c_{igm'}$ is the guessing parameter, and

$\sum_{p=1}^{P} \beta_{igm'pk} \xi_p$ is the difficulty parameter, which is decomposed into p basic parameters ξ_p with weights $\beta_{igm'pk}$, which could be specified, or estimated if $y_{ijm'k}$ s are observable.

If we consider the ordinal attribute scale as a restricted case of the continuous attribute scale (note that a model using the ordinal attribute scale is not statistically nested in a model employing the continuous attribute scale), then most of the models in Figure 1 can be seen as the sub-models of Equation 16. Specifically, in Figure 1 all the models in the first row under the general CDPM, except for MCLCM-DJ, SLM, TBA, RS, 3PL-MIRT-CJ&C-D, CMDLV and BN-WAT can be shown as the restricted versions of Equation 16, by dropping model element(s), or moving from high level element(s) to low level element(s). Although the usefulness of this complex upper-level model on test practice and the feasibility of its estimation need to be further explored, it definitely provides a framework to develop new CDPMs, which may have both practical and theoretical value, by filling in gap between models. The following are some examples: (a) a 3 parameter compensatory multidimensional logistic (or normal ogive) model for polytomous-scored items, maybe even with interaction terms; (b) a mixed 3-parameter unidimensional (or multidimensional) IRT and latent class model with multiple latent grouping variables, (c) a mixed MRCMLM with latent class model, and (d) a mixed Fusion Model (FM) with latent class model.

Some existing models could be elaborated further. For example, (a) the estimation method for MCUIRT has not been worked out; (b) the same situation applies for 3PL-MIRT-CJ&C-D; (c) the marginal maximum likelihood or Bayesian modal estimation of the MCLCM-CJ under the parameterization of FM (Equation 18) remains to be developed[4].

4.2.2. Other Topics

1. A comparison of the item response functions of the MixRM (or MixRM-LL) and the FM-D shows that they are quite similar: For the MixRM, the item response function for a group of examinees (having the same attribute state) can be considered as a Rasch model with a universal item difficulty parameter (across groups) and a Saltus item difficulty parameter specific for this group, while for the FM-D the same function can be interpreted as a Rasch model with a universal item difficulty parameter and a weight on this Rasch model specific for this group. For a small number of latent groups (attributes) and a large number of items, the MixRM is much more parsimonious than FM-D in terms of the number of estimated parameters, and thus may be a good replacement for FM-D. Actually many parameters in FM-D may not be estimated very well because of the complexity of the model (Hartz, 2002). Bolt (1999) first explored this possibility, but much more can be done along this direction in the future.

2. CMDLV and Rabe-Hesketh, Skrondal, and Pickles' (2004) generalized linear latent and mixed models framework (GLLAMM) are two very flexible modeling approaches with broad applications in cognitive diagnostic testing, such as modeling causal relationships, incorporating interaction terms, and detecting DIF items. However, the CMDLV was developed and popular in the latent class analysis community, and the GLLAMM in biostatistics. These frameworks are largely unnoticed by the educational measurement community. It will be a fruitful effort to explore their rich applications in educational measurement.

3. Constructing and verifying a Q matrix are always challenging in cognitively diagnostic assessment (CDA). Besides experts' domain knowledge, test data often can help in these regards. Hartz (2002, chap. 5) demonstrated a naïve way to incorporate items' p values (proportion of correct answers) in Q matrix construction. And Q matrix exploration can be

easily carried out via model comparisons. In addition, HICLAS (-CJ and –DJ) has a built-in mechanism to explore Q matrixes, although its application in CDA is very limited. Maris, De Boeck, and van Mechelen (1996) extended HICLAS to the probability matrix decomposition models (PMD). However, PMD has many unsolved issues in applications to CDA, which remains to be explored and elaborated, including the identification conditions.

4. It will be very interesting to carry out some comparison studies among the FM-D family of models, including the FM-D, the MCLCM-CJ in the FM parameterization, Junker and Sijtsma's (2001) simplified version of MCLCM-CJ, Bolt and Mroch's (2004) extension of Junker and Sijtsma's model, and the one proposed here. The response function of the FM-D is written as:

$$P(x_{ij}^{D} = 1 \mid \boldsymbol{\alpha}_j, \theta_j) = \pi_i^* \prod_{k=1}^{K} r_{ik}^{*(1-\alpha_{jk})\times q_{ik}} P_{c_i}(\theta_j), \tag{17}$$

where

x_{ij}^{D} is examinee j's response to item i which is a dichotomous value;

$\boldsymbol{\alpha}_j = \{\alpha_1, \cdots, \alpha_k\}$ is examinee j's dichotomous attribute state;

K is the number of attributes in the Q matrix;

π_i^* is the probability of successfully applying all attributes required by item i as specified in Q matrix, given that an examinee masters all those attributes;

r_{ik}^* $(0 \le r_{ik}^* \le 1)$ is the ratio of (a) the probability of correctly executing attribute k to item i given that an examinee masters attribute k, and (b) the probability of correctly executing attribute k to item i given an examinee does not master attribute k;

θ_j is the examinee j's residual ability;

$P_{c_i}(\theta_j) = P_{c_i}(x_{ij} = 1 | \theta_j) = \dfrac{1}{1 + \exp(-\theta_j - c_i)}$, is the unidimensional Rasch model with item easiness parameter c_i ($0 \le c_i \le 3$) to account for the effect of examinee j's residual ability in answering item i correctly.

The MCLCM-CJ in the FM parameterization is just the FM-D excluding the residual part:

$$P(x_{ij}^D = 1 | \boldsymbol{\alpha}_j) = \pi_i^* \prod_{k=1}^{K} r_{ik}^{*(1-\alpha_{jk}) \times q_{ik}} \ . \tag{18}$$

Junker and Sijtsma (2001) proposed a simplified version of the original MCLCM-CJ (Maris, 1999). Their model can be written as:

$$P(x_{ij}^D = 1 | \boldsymbol{\alpha}_j) = \prod_{k=1}^{K} \pi_k^{\alpha_{jk} * q_{ik}} r_k^{(1-\alpha_{jk}) * q_{ik}} \ , \tag{19}$$

where

π_k is the probability of successfully applying attribute k given an examinee masters this attribute, and

r_k is the probability of successfully applying attribute k given an examinee is a nonmaster of this attribute.

Note that π_k and r_k are the same across items. Bolt and Mroch (2004) extended Junker and Sijtsma's model by introducing item-specific attribute mastery, that is, an attribute master on an item may not necessarily be an master of this attribute on another item, to account for, to some degree, the variation in the execution of an attribute across items. Thus, the α_{jk} s in Equation 19 should add a subscript i:

$$P(x_{ij}^D = 1 | \boldsymbol{\alpha}_j) = \prod_{k=1}^{K} \pi_k^{\alpha_{ijk} * q_{ik}} r_k^{(1-\alpha_{ijk}) * q_{ik}} \ . \tag{20}$$

If a researcher is interested in estimating an examinee's probability of mastering each measured attribute, instead of an examinee's mastery status, and assumes that a master of an attribute has probability 1 to execute it successfully to all the items in a test, then a MCLCM-CJ like model can be written as:

$$P(x_{ij}^{D} = 1 \mid \pi_{j1}, \cdots, \pi_{jK}) = \prod_{k=1}^{K} \pi_{jk}^{q_{ik}} , \tag{21}$$

where π_{jk} is examinee j's probability of mastering attribute k. Note that this model does not have any item parameter and actually is a simplified version of Maris, et al. (1996) conjunctive version of PMD with Q matrix fixed. A sufficient condition for Equations 19 through 21 to be identified is a full rank Q matrix. The relationship among these models can be viewed from the following three respects. First, in terms of model complexity, they form an order as (from complex to simple): Equation 17, Equation 18, Equation 20, Equation 19 and Equation 21. Secondly, in terms of the heterogeneity of the probabilities of successfully executing an attribute across items given an examinee, the ordering (from high to low) is {Equation 17, Equation 18}, Equation 20, and {Equation 19, Equation 21}, where models within {} are at the same level. Thirdly, in terms of the heterogeneity of the probabilities of successfully executing an attribute across examinees given an item, the ordering (from high to low) is Equation 21, and {other models}. In real applications, it is particularly worth comparing these models to find a best fitting model. Actually, based on their differences on the above three respects, researchers often can speculate which model will be favored most by a given test. In addition, all these models could be extended to handle polytomously scored items, for example, by using the cumulative response function.

5. Some machine learning algorithms, for example, classification and regression tree, artificial neutral network and support vector machine (Mitchell, 1997), may have potential

applications in cognitively diagnostic testing. For example, the ideal response patterns (as in RS) may provide training samples for these algorithms, and the resulting models can be used to classify examinees. Levine, Williams, and Jun (1993) is one example of research in this direction.

5. Concluding Remarks

Multidimensional IRT models have been developed since the 1960s (Reckase, 1997), and are becoming popular in recent years because of the societal needs for cognitively diagnostic testing. Especially, CDPMs with discrete attributes have gained much attention because of their nice interpretability, that is, the straightforward meanings of attributes to common audiences. The FM, RS and BN are significant models developed in the past 10 years, and are still under intensive study. The conjunctive attribute structure is often assumed in the current cognitively diagnostic testing practice, for example, the PSAT test (Stout, 2002).

Psychometrics is a relatively narrow field in applied statistics with a simple and definitive research objective: inferring examinees' true states on some psychological or cognitive features from their responses on a test. Therefore, psychometric models are much more similar than the models in other fields of applied statistics, such as biostatistics or econometrics, as can be seen from the above review. The good thing about this is that estimating new CDPMs, which is usually considered as a hard part of developing new statistical models, should not be a difficult task in general; estimation can always be derived in a manner similar to those used in other comparable models. Furthermore, with the advance of numeric methods and the wide availability of optimization routings, maximum likelihood estimation becomes so straightforward that even derivatives are not necessary. The advance of Bayesian inference and Markov Chain

Monte Carlo (MCMC) techniques simplify CDPM estimation further. MCMC provides a unified sampling-based approach for the estimation of all kinds of statistical models from the Bayesian perspective, including model checking and comparison. Using general purpose MCMC software, such as WinBugs1.4 (Spiegelhalter, Thomas, Best, & Lunn, 2003), a very complicated model can be set up for estimation by straightforward programming. Writing a computer program for a specific model implementing MCMC algorithms is not hard either. However, easy estimation of complicated models should not be seen as a sign to encourage the use of complicated models. Instead, people should use the simpler models wherever possible, because (a) the complicated models are often difficult to be understood and interpreted; (b) the estimations of complicated models are less stable than simple models, and (c) complicated models often involve subtle identification conditions, which are difficult to be evaluated. For instance, the identification issues in Maris et al.'s (1996) PMD models, Mislevy, et al.'s (2001) BN-WAT models, and Junker and Sijtsma's (2001) simplified version of MCLCM-CJ need more consideration. Molenaar (1997a) showed, in the unidimensional IRT context, the complicated models usually would not, in practice, change the interested conclusions based on the simpler models.

References

Ackerman, T., & Spray, J. A. (1987). *A general model for item dependency* (ACT Research Report Series No. 87-9). Iowa City, IA: American College Testing Program.

Adams, R. J., Wilson, M., & Wang, W.-c. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1-23.

Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). A four-process architecture for assessment delivery, with connections to assessment design. Retrieved October 8, 2004, from http://www.education.umd.edu/EDMS/mislevy/papers/ProcessDesign.pdf

Andersen, E. B. (1973). Conditional inference for multiple-choice questionnaires. *British Journal of Mathematical & Statistical Psychology, Vol. 26*(1), 31-44.

Andrich, D. (1997). An hyperbolic cosine IRT model for unfolding direct responses of persons to items. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 399-414). New York, NY: Springer-Verlag.

Baxter, G. P., & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement: Issues & Practice, 17*(3), 37-46.

Béguin, A. A., & Glas, C. A. W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika, 66*(4), 541-561.

Bolt, D. M. (1999). *Applications of an IRT mixture model for cognitive diagnosis.* Paper presented at the AERA annual meeting, Montreal, Quebec, CANADA.

Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speediness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement, 39*(4), 331-348.

Bolt, D. M., & Mroch, A. (2004). *Application of a mixture IRT model for cognitive diagnosis.* Paper presented at the Psychometric Society 2004 annual meeting, Monterey, CA.

Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64*(2), 153-168.

Butter, R., De Boeck, P., & Verhelst, N. D. (1998). An item response model with internal restrictions on item difficulty. *Psychometrika, 63*(1), 47-63.

Chipman, S. F., Nichols, P., & Brennan, R. L. (1995). Introduction. In P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 1-18). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Corbett, A. T., Anderson, J. R., & O'Brien, A. T. (1995). Student modeling in the act programming tutor. In P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 19-41). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

De Boeck, P., & Rosenberg, S. (1988). Hierarchical classes: Model and data analysis. *Psychometrika, 53*(3), 361-381.

DiBello, L. V., Stout, W., & Roussos, L. (1993). *Unified cognitive/psychometric diagnosis: Foundations and application.* Paper presented at the AERA annual meeting, Atlanta, Ga.

DiBello, L. V., Stout, W., & Roussos, L. (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: Lawrence Erlbaum Associates, Inc.

Draney, K. L., Pirolli, P., & Wilson, M. (1995). A measurement model for a complex cognitive skill. In P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 103-125). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Embretson, S. E. (1984). A general latent trait model for response processes. *Psychometrika, 49*(2), 175-186.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika, 56*(3), 495-515.

Embretson, S. E. (1997). Structured ability models in tests designed from cognitive theory. In M. Wilson, G. J. Engelhard & K. L. Draney (Eds.), *Objective measurement* (Vol. 4, pp. 223-236). Greenwich, CT: Ablex Publishing Corporation.

Embretson, S. E. (1999). Cognitive psychology applied to testing. In F. T. Durso (Ed.), *Handbook of applied cognition* (pp. 629-660). New York, NY: John Wiley & Sons Ltd.

Embretson, S. E., & Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement, 38*(4), 343-368.

Falmagne, J.-C. (1989). A latent trait theory via a stochastic learning theory for a knowledge space. *Psychometrika, 54*(2), 283-303.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, Vol. 37*(6), 359-374.

Fischer, G. H. (1977). Linear logistic latent trait models: Theory and applications. In H. Spada & W. Kempf (Eds.), *Structural models of thinking and learning* (pp. 203-225). Berne: Huber.

Fischer, G. H. (1997). Unidimensional linear logistic Rasch models. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 225-244). New York, NY: Springer-Verlag.

Fischer, G. H. (2001). Gain scores revisited under an IRT perspective. In A. Boomsma, M. A. J. van Duijn & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 43-68). New York, NY: Springer-Verlag.

Fischer, G. H., & Parzer, P. (1991). An extension of the rating scale model with an application to the measurement of change. *Psychometrika*, *56*(4), 637-651.

Fischer, G. H., & Ponocny, I. (1994). An extension of the partial credit model with an application to the measurement of change. *Psychometrika*, *59*(2), 177-192.

Fischer, G. H., & Ponocny, I. (1995). Extended rating scale and partial credit models for assessing change. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 181-202). New York, NY: Springer-Verlag.

Fischer, G. H., & Seliger, E. (1997). Multidimensional linear logistic models for change. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 232-346). New York, NY: Springer-Verlag.

Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, *87*, 476-486.

Formann, A. K. (1995). Linear logistic latent class analysis and the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 239-256). New York, NY: Springer-Verlag.

Fu, J. (2005). *A polytomous extension of the fusion model and its Bayesian parameter estimation.* Unpublished doctoral dissertation, University of Wisconsin, Madison.

Fu, J., & Bolt, D. M. (2004). *A polytomous extension of the fusion model and its Bayesian parameter estimation.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Glas, C. A. W., & Verhelst, N. D. (1989). Extensions of the partial credit model. *Psychometrika, 54*(4), 635-659.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement, 26*(4), 301-321.

Hagenaars, J. A. (1990). *Categorical longitudinal data: Loglinear panel, trend and cohort analysis.* Newbury Park, CA: Sage.

Hagenaars, J. A. (1993). *Loglinear models with latent variables*. Thousand Oaks, CA: Sage Publications, Inc.

Hagenaars, J. A. (1998). Categorical causal modeling. *Sociological Methods & Research, 26*(4), 436-487.

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

Hoijtink, H. (1997). PARELLA: An IRT model for parallelogram analysis. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 415-430). New York, NY: Springer-Verlag.

Hoskens, M., & De Boeck, P. (2001). Multidimensional componential item response theory models for polytomous items. *Applied Psychological Measurement, 25*(1), 19-37.

Jannarone, R. J. (1986). Conjunctive item response theory kernels. *Psychometrika, 51*(3), 357-373.

Jannarone, R. J. (1997). Models for locally dependent response: Conjunctive item response theory. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 271-286). New York, NY: Springer-Verlag.

Junker, B. W. (1999). Some statistical models and computational methods that may be useful for cognitively-relevant assessment. Retrieved October 2, 2005, from http://www.stat.cmu.edu/~brian/nrc/cfa/

Junker, B. W. (2001). On the interplay between nonparametric and parametric IRT, with some thoughts about the future. In A. Boomsma, M. A. J. van Duijn & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 247-276). New York, NY: Springer-Verlag.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258-273.

Kelderman, H. (1984). Loglinear Rasch model tests. *Psychometrika, 49*(2), 223-245.

Kelderman, H. (1997). Loglinear multidimensional item response model for polytomously scored items. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 287-304). New York, NY: Springer-Verlag.

Kelderman, H., & Macready, G. B. (1990). The use of loglinear models for assessing differential item functioning across manifest and latent examinee groups. *Journal of Educational Measurement, 27*(4), 307-327.

Kempf, W. (1977). Dynamic models for the measurement of 'traits' in social behavior. In W. Kempf & B. H. Repp (Eds.), *Mathematical models for social psychology* (pp. 14-58). Berne: Huber.

Klauer, K. C., & Sydow, H. (2001). Modeling learning in short-term learning tests. In A. Boomsma, M. A. J. van Duijn & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 69-88). New York, NY: Springer-Verlag.

Levine, M. V., Williams, B., & Jun, Y. (1993). *Cognitive diagnosis and the prediction of educational outcomes.* Paper presented at the conference on alternative diagnostic assessment, Iowa City.

Maris, E. (1995). Psychometric latent response models. *Psychometrika, 60*(4), 523-547.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika, 64*(2), 187-212.

Maris, E., De Boeck, P., & van Mechelen, I. (1996). Probability matrix decomposition models. *Psychometrika*, *61*(1), 7-29.

McDonald, R. P. (1997). Normal-Ogive multidimensional model. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 258-270). New York, NY: Springer-Verlag.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. Retrieved October 8, 2004, from http://www.education.umd.edu/EDMS/mislevy/papers/BriefIntroECD.pdf

Mislevy, R. J., Almond, R. G., Yan, D., & Steinberg, L. S. (1999). Bayes nets in educational assessment: Where do the numbers come from? In K. B. Laskey & H. Prade (Eds.), *Proceedings of the fifteenth conference on uncertainty in artificial intelligence* (pp. 437-446). San Francisco, CA: Morgan Kaufmann.

Mislevy, R. J., Senturk, D., Almond, R. G., Dibello, L. V., Jenkins, F., Steinberg, L. S., et al. (2001). *Modeling conditional probabilities in complex educational assessments.* Paper presented at the International Meeting of the Psychometric Society, Osaka, Japan.

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55*(2), 195-215.

Mislevy, R. J., & Wilson, M. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika, 61*(1), 41-71.

Mitchell, T. M. (1997). *Machine learning*. New York, NY: McGraw-Hill.

Mokken, R. J. (1997). Nonparametric models for dichotomous responses. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 351-368). New York, NY: Springer-Verlag.

Molenaar, I. W. (1997a). Lenient or strict application of IRT with an eye on practical consequences. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 38-49). New York: Waxmann.

Molenaar, I. W. (1997b). Nonparametric models for polytomous responses. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 369-380). New York, NY: Springer-Verlag.

Moustaki, I. (2000). A latent variable model for ordinal variables. *Applied Psychological Measurement, 24*(3), 211-223.

Muller, H. (1987). A Rasch model for continuous ratings. *Psychometrika, 52*(2), 165-181.

Muraki, E., & Carlson, J. E. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement, 19*(1), 73-90.

National Governors' Association. (1990). *Educating America: State strategies for achieving the national education goals.* Washington, DC: Author.

Nichols, P., & Sugrue, B. (1999). The lack of fidelity between cognitively complex constructs and conventional test development. *Educational Measurement: Issues & Practice, 18*(2), 18-30.

Nichols, P. D. (1994). A framework for developing cognitively diagnostic assessments. *Review of Educational Research, 64*(4), 575-604.

Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Eds.). (1995). *Cognitively diagnostic assessment*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69*(2), 167-190.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the iv Berkeley symposium on mathematical statistics and probability* (Vol. IV, pp. 321-333). Berkeley: University of California Press.

Reckase, M. D. (1997). A linear logistic multidimensional model for dichotomous item response data. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 271-286). New York, NY: Springer-Verlag.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Norwell, MA: Kluwer.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*(1), 3-33.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2002). Characteristics of MML/EAP parameter estimates in the generalized graded unfolding model. *Applied Psychological Measurement, 26*(2), 192-208.

Roskam, E. E. (1997). Models for speed and time-limit tests. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 187-208). New York, NY: Springer-Verlag.

Rost, J. (1988a). Rating scale analysis with latent class models. *Psychometrika, 53*(3), 327-348.

Rost, J. (1988b). Test theory with qualitative and quantitative latent variables. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 147-171). New York, NY: Plenum Press.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*(3), 271-282.

Rost, J. (1991). A logistic mixture distribution model for polychotomous item responses. *British Journal of Mathematical & Statistical Psychology, 44*(1), 75-92.

Rost, J. (2001). The growing family of Rasch models. In A. Boomsma, M. A. J. van Duijn & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 25-42). New York, NY: Springer-Verlag.

Rost, J., & Carstensen, C. H. (2002). Multidimensional Rasch measurement via item component models and faceted designs. *Applied Psychological Measurement*, *26*(1), 42-56.

Roussos, L. (1994). *Summary and review of cognitive diagnosis models*. Unpublished manuscript, University of Illinois at Urbana-Champaign.

Samejima, F. (1974). Normal Ogive model on the continuous response level in the multidimensional latent space. *Psychometrika, Vol. 39*(1), 111-121.

Samejima, F. (1995). A cognitive diagnosis method using latent trait models: Competency space approach and its relationship with Dibello and stout's unified cognitive-psychometric diagnosis model. In P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 391-410). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sheehan, K. M. (1997). A tree-based approach to proficiency scaling and diagnostic assessment. *Journal of Educational Measurement, 34*(4), 333-352.

Spiegelhalter, D. J., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS user manual. Retrieved October 9, 2004, from http://www.mrc-bsu.cam.ac.uk/bugs

Spray, J. A. (1997). Multiple-attempt, single-item response models. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 209-220). New York, NY: Springer-Verlag.

Spray, J. A., Davey, T. C., Reckase, M. D., Ackerman, T., & Carlson, J. E. (1990). *Comparison of two logistic multidimensional item response theory models.* (Research Report No. ACT-RR-90-8, ONR90-8). Iowa City, IA: American College Testing Program; Office of Naval Research, Cognitive and Neural Sciences Div.

Stegelmann, W. (1983). Expanding the Rasch model to a general model having more than one dimension. *Psychometrika, 48*(2), 259-267.

Stout, W. (2002). Psychometrics: From practice to theory and back: 15 years of nonparametric multidimensional IRT, dif/test equity, and skills diagnostic assessment. *Psychometrika, 67*(4), 485-518.

Sympson, J. B. (1978). A model for testing with multidimensional items. In D. J. Weiss (Ed.), *Proceedings of the 1977 computerized adaptive testing conference* (pp. 82-98). Minneapolis, MN: University of Minnesota, Department of Psychology, Psychometric Methods Program.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Applied Statistics, 51*(3), 337-350.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453-488). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327-359). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Templin, J., He, X., Roussos, L., & Stout, W. (2003). *The pseudo-item method: A simple technique for analysis of polytomous data with the fusion model* (Draft Technical Report). Urbana-Champaign, IL: University of Illinois at Urbana-Champaign.

Templin, J., Roussos, L., & Stout, W. (2003). *An extension of the current fusion model to treat polytomous attributes* (Draft Technical Report). Urbana-Champaign, IL: University of Illinois at Urbana-Champaign.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 50*(4), 567-577.

Uebersax, J. S. (1999). Probit latent class analysis with dichotomous or ordered category measures: Conditional independence/dependence models. *Applied Psychological Measurement, 23*(4), 283-297.

van der Ark, L. A. (2001). Relations and properties of polytomous item response theory models. *Applied Psychological Measurement, 25*(3), 273-283.

van Mechelen, I., De Boeck, P., & Rosenberg, S. (1995). The conjunctive model of hierarchical classes. *Psychometrika, 60*(4), 505-521.

van Onna, M. J. H. (2002). Bayesian estimation and model selection in ordered latent class models for polytomous items. *Psychometrika, 67*(4), 519-538.

Verguts, T., & De Boeck, P. (2000). A Rasch model for detecting learning while solving an intelligence test. *Applied Psychological Measurement, 24*(2), 151-162.

Verhelst, N. D., & Glas, C. A. W. (1995a). Dynamic generalizations of the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 181-202). New York, NY: Springer-Verlag.

Verhelst, N. D., & Glas, C. A. W. (1995b). The one parameter logistic model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 215-238). New York, NY: Springer-Verlag.

Verhelst, N. D., Verstralen, H. H. F. M., & Jansen, M. G. H. (1997). A logistic model for time-limit tests. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 169-186). New York, NY: Springer-Verlag.

Vermunt, J. K. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Applied Psychological Measurement, 25*(3), 283-295.

von Davier, M., & Rost, J. (1997). Self monitoring: A class variable? In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 296-305). New York: Waxmann.

Wainer, H., Bradlow, E. T., & Du, Z. (2001). Testlet response theory: An analog for the 3pl model useful in adaptive testing. In W. J. Van der Lindern & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 245-270). Boston, MA: Kluwer-Nijhoff.

Wang, X., Bradlow, E. T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. *Applied Psychological Measurement, 26*(1), 109-128.

Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika, 45*(4), 479-494.

Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin, 105*(2), 276-289.

Yamamoto, K. (1989). *A hybrid model of IRT and latent class models* (ETS Research Report No. RR-89-41). Princeton, NJ: Educational Testing Service.

Zwinderman, A. A. (1997). Response models with manifest predictors. In W. J. van der Lindern & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 245-257). New York, NY: Springer-Verlag.

Acknowledgment

The authors thank Brain W. Junker and Daniel M. Bolt for their helpful comments which led to an improvement on the early version of this paper.

Address

Correspondence may be sent to:

Jianbin Fu

Educational Testing Service
MS 13-P
Rosedale Road
Princeton, NJ 08541

Email: jfu@air.org

Footnotes

[1] Cognitively diagnostic psychometric models have also been called cognitive psychometric models in Embretson (1999), psychometric skills diagnostic models in Stout (2002), and cognitive diagnosis models in Roussos (1994).

[2] In TBA each item relates to only one low level attribute. Thus, it is not so meaningful to treat those attributes as strategies. In RS, ideal response patterns are not easy to be identified for strategies.

[3] It may not be appropriate if people interpret the probability as successfully applying a strategy to the item, because it does not make much sense that an examinee solves an item by applying both strategies simultaneously.

[4] Maris (1999) developed these estimation methods for the original MCLCM-CJ under the assumption that all items require the same set of attributes, and imposed strong constraints to identify the model, which make this model not very cognitively attractive.

Table 1

Cognitively Diagnostic Psychometric Models: Full Names and Acronyms (in alphabetical order of model acronyms)

Acronym	Full name
2PNO-MIRT-C 2PNO-MIRT-C-D 2PNO-MIRT-C-P	Compensatory 2-Parameter Normal-Ogive Multidimensional IRT Model (2PNO-MIRT-C): for Dichotomous Items (2PNO-MIRT-C-D) (McDonald, 1997); for Polytomous Items (2PNO-MIRT-C-P) (Muraki & Carlson, 1995)
2PNO-MIRT-C-C	Compensatory 2-Parameter Normal-Ogive Multidimensional IRT Model for the Continuous Response (2PNO-MIRT-C-C) (Samejima, 1974)
2PL-MIRT-C-P	Compensatory 2-Parameter Logistic Multidimensional IRT Model for Polytomous Items (2PL-MIRT-C-P) (Moustaki, 2000)
3PL-MIRT-C-D	Compensatory 3-Parameter Logistic Multidimensional IRT Model for Dichotomous Items (3PL-MIRT-C-D) (Reckase, 1997)
3PL-MIRT-C&CJ-D	Generalized Compensatory and Conjunctive Three Parameter Logistic Model for Dichotomous Items (3PL-MIRT-C&CJ-D) (Spray, Davey, Reckase, Ackerman, & Carlson, 1990)
3PL-MIRT-CJ-D	Conjunctive 3-Parameter Logistic Multidimensional IRT Model for Dichotomous Items (3PL-MIRT-CJ-D) (Sympson, 1978)
3PNO-MIRT-C-D	Compensatory 3-Parameter Normal-Ogive Multidimensional IRT Model for Dichotomous Items (3PNO-MIRT-C-D) (Beguin & Glas, 2001);
BN BN-WOAT BN-WAT	Bayesian Networks (BN): without attribute transformation (only for Dichotomous Items) (BN-WOAT) (Mislevy, Almond, Yan, & Steinberg, 1999); with attribute transformation (BN-WAT) (Mislevy, et al. 2001)
BS	Binary Skills Model (BS) (Haertel, 1989)
CMDLV	Causal Models with Discrete Latent Variables (CMDLV) (Hagenaars, 1990, 1993, 1998)
CRK	Conjunctive Rasch Kernel (CRK) (Jannarone, 1986, 1997)
CRM	Conjunctive Rasch Model (CRM) (Maris, 1995)
DRM	Dynamic Rasch Model (DRM) (Verhelst & Glas, 1995a; Verguts & De Boeck, 2000)
DTM	Dynamic Test Model (DTM) (Kempf, 1977)
FM FM-D FM-P	Fusion Model (FM):for Dichotomous Items (FM-D) (Hartz, 2002;Templin, Roussos & Stout, 2003); for Polytomous Items (FM-P) (Fu & Bolt, 2004; Templin, He, Roussos & Stout, 2003)
GLTM GLTM-WOG GLTM-WG	General Component Latent Trait Model (GLTM): Without Guessing (GLTM-WOG); with Guessing (GLTM-WG) (Embretson, 1984)
HICLAS-DJ	Disjunctive Hierarchical Class Model (HICLAS-DJ) (De Boeck & Rosenberg, 1988)

Table 1 (cont.)

Cognitively Diagnostic Psychometric Models: Full Names and Acronyms (in alphabetical order of model acronyms)

Acronym	Full name
HICLAS-CJ	Conjunctive Hierarchical Class Model (HICLAS-CJ) (van Mechelen, De Boeck, & Rosenberg, 1995)
HRLC	Hybrid Rasch-Latent Class Model (HRLC) (Davier & Rost, 1997)
HYBRID	HYBRID Model (HYBRID) (Yamamoto, 1989)
LLRA	LLTM with Relaxed Assumptions (LLRA) (Fischer, 1977; Fischer & Seliger, 1997)
LLTM	Linear Logistic Test Model (LLTM) (Fischer, 1973, 1997)
LPCM	Linear Partial Credit Model (LPCM) (Fischer & Ponocny, 1994,1995; Glas & Verhelst, 1989)
LPOCM-D	Latent Partially Ordered Classification Models for Dichotomous Items (LPOCM-D) (C. Tatsuoka, 2002)
LRM	Loglinear Rasch Model (LRM) (Kelderman, 1984)
LRSM	Linear Rating Scale Model (LRSM) (Fischer & Parzer, 1991)
MCLCM-DJ	Disjunctive Multiple Classification Latent Class Model (MCLCM-DJ) (Maris, 1999)
MCLCM-CJ	Conjunctive Multiple Classification Latent Class Model (MCLCM-CJ) (Maris, 1999)
MCUIRT	Markov Chain Unidimensional IRT Model (MCUIRT) (Ackerman & Spray, 1987)
MIRID	Model with Internal Restrictions on Item Difficulty (MIRID) (Butter, De Boeck, & Verhelst, 1998)
MixRM	Mixture Rasch Model (MixRM) (Bolt, 1999; Bolt, Cohen, & Wollack, 2002; Mislevy & Verhelst, 1990; Rost, 1990, 1991)
MixRM-LL	Mixture Rasch Model—Log-Linear Modeling (MixRM-LL) (Kelderman & Macready, 1990)
MLPCM	Multidimensional Linear Partial Credit Model (MLPCM) (Fischer & Ponocny, 1994,1995)
MLRSM	Multidimensional Linear Rating Scale Model (MLRSM) (Fischer & Parzer, 1991)
MLT	Mixed Latent Trait Models (MLT) (Uebersax, 1999)
MLTM MLTM-WOG MLTM-WG	Multicomponent Latent Trait Model (MLTM): without Guessing (MLTM-WOG) (Whitely, 1980); with Guessing (MLTM-WG) (Embretson, 1984)
MMRM	Multicategorical and Multidimensional Rasch Model (MMRM) (Rasch, 1961; Andersen, 1973)
MPRM	Multidimensional Polytomous Rasch Model (MPRM) (Kelderman, 1997)
MRCMLM	Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM) (Adams, Wilson, & Wang, 1997)
MRM	Multicomponent Rasch Model (MRM) (Stegelmann, 1983)
MRMLC	Multidimensional Rasch Model for Learning and Change (MRMLC) (Embretson, 1991)
MSLM	Measurement-Structural Learning Model (MSLM) (Klauer & Sydow, 2001)

Table 1 (cont.)

Cognitively Diagnostic Psychometric Models: Full Names and Acronyms (in alphabetical order of model acronyms)

Acronym	Full name
MULTRA	Multidimensional Rasch Model (MULTRA) (Rost & Carstensen, 2002)
PCMC	Partial Credit Model for Change (PCMC) (Fischer, 2001)
RS	Rule Space (RS) (K.K. Tatsuoka, 1995; DiBello, 2002)
SALTUS	Saltus Model (SALTUS) (Wilson, 1989; Mislevy & Wilson, 1996)
SLM	Stochastic Learning Model (SLM) (Falmagne, 1989)
SLTM	Structured Latent Trait Model (SLTM) (Embretson, 1997)
TBA	Tree-Based Approach (TBA) (Sheehan, 1997)
TM TM-2PNO-D TM-2PNO-P TM-3PL-D	Testlet Model (TM): Two Parameter Normal Ogive Model for Dichotomous Items (TM-2PNO-D) (Bradlow, Wainer, & Wang, 1999), and for Polytomous Items (TM-2PNO-P) (Wang, Bradlow, & Wainer, 2002); Three Parameter Logistic Model for Dichotomous Items (TM-3PL-D) (Wainer, Bradlow, & Du, 2001)
UM	Unified Model (UM) (DiBello, Stout, & Roussos, 1993, 1995)

Table 2

Cross-Classification of CDPMs by Attribute Scale and Dimensionality

Attribute scale[a]	Dimensionality[b]		
	U	M	U|M
C	DRM DTM LLTM LPCM LRM LRSM MCUIRT MIRID MSLM PCMC SALTUS	2PNO-MIRT-C 2PNO-MIRT-C-C 2PL-MIRT-C-P 3PL-MIRT-C-D 3PL-MIRT-C&CJ-D 3PL-MIRT-CJ-D 3PNO-MIRT-C-D CRM GLTM LLRA MLPCM MLRSM MLTM MMRM MPRM MRCMLM MRM MULTRA TM	MRMLC SLTM CRK
D		BN-WOAT BS HICLAS-DJ HICLAS-CJ LPOCM-D MCLCM-DJ MCLCM-CJ RS SLM TBA	
P		BN-WAT	
C&P		CMDLV FM-D HYBRID	HRLC MixRM MixRM-LL MLT
C&D		FM-P UM	

Note. For the full name of each acronym see Table 1.
[a]C = Continuous, D = Dichotomous, P = Polytomous, & = And.
[b]M = Multidimensional, U = Unidimensional, | = Or.

Table 3

Cross-Classification of CDPMs by Item Type, Q Incompleteness and Cognitive Strategy

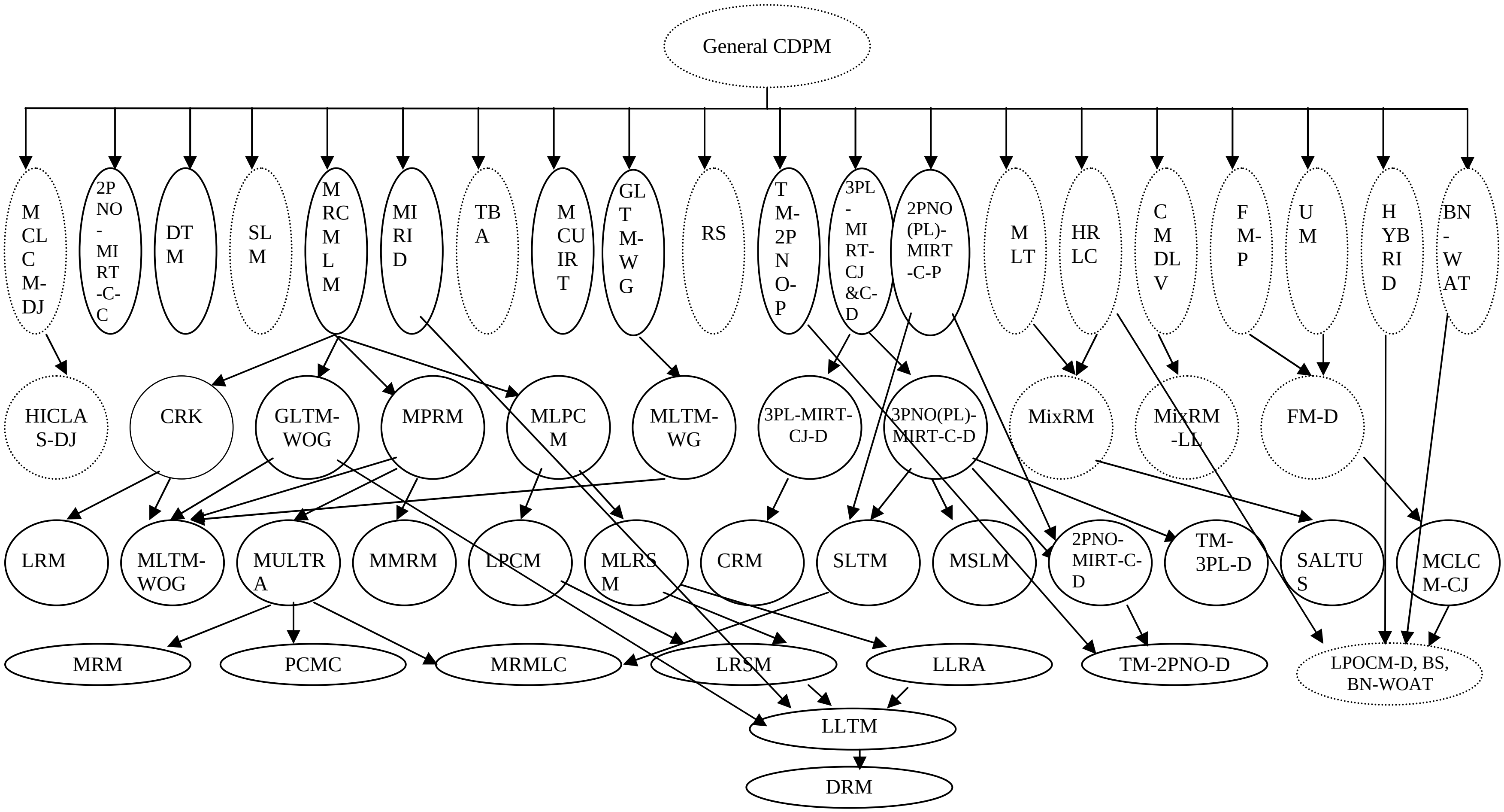
Item type	Q incompleteness	Yes		No		NA	
	Cognitive strategy	Single	Multiple	Single	Multiple	Single	Multiple
Dichotomous		2PNO-MIRT-C-D 3PL-MIRT-C-D 3PNO-MIRT-C-D 3PL-MIRT-C&CJ-D 3PL-MIRT-CJ-D CRM CRK GLTM-WG LLRA LLTM MRM MRMLC SLTM	FM-D MLTM- WG UM	GLTM- WOG MIRID RS TBA	BN-WOAT BS HICLAS-DJ HICLAS-CJ HYBRID LPOCM-D MCLCM-DJ MCLCM-CJ MLTM-WOG SLM	DRM DTM MCUIRT MSLM LRM	SALTUS
Polytomous		2PNO-MIRT-C-P 2PL-MIRT-C-P LPCM LRSM MLPCM MLRSM MPRM MRCMLM MULTRA TM	BN-WAT FM-P	MMRM	CMDLV HRLC MixRM MixRM-LL MLT	PCMC	
Continuous		2PNO-MIRT-C-C					

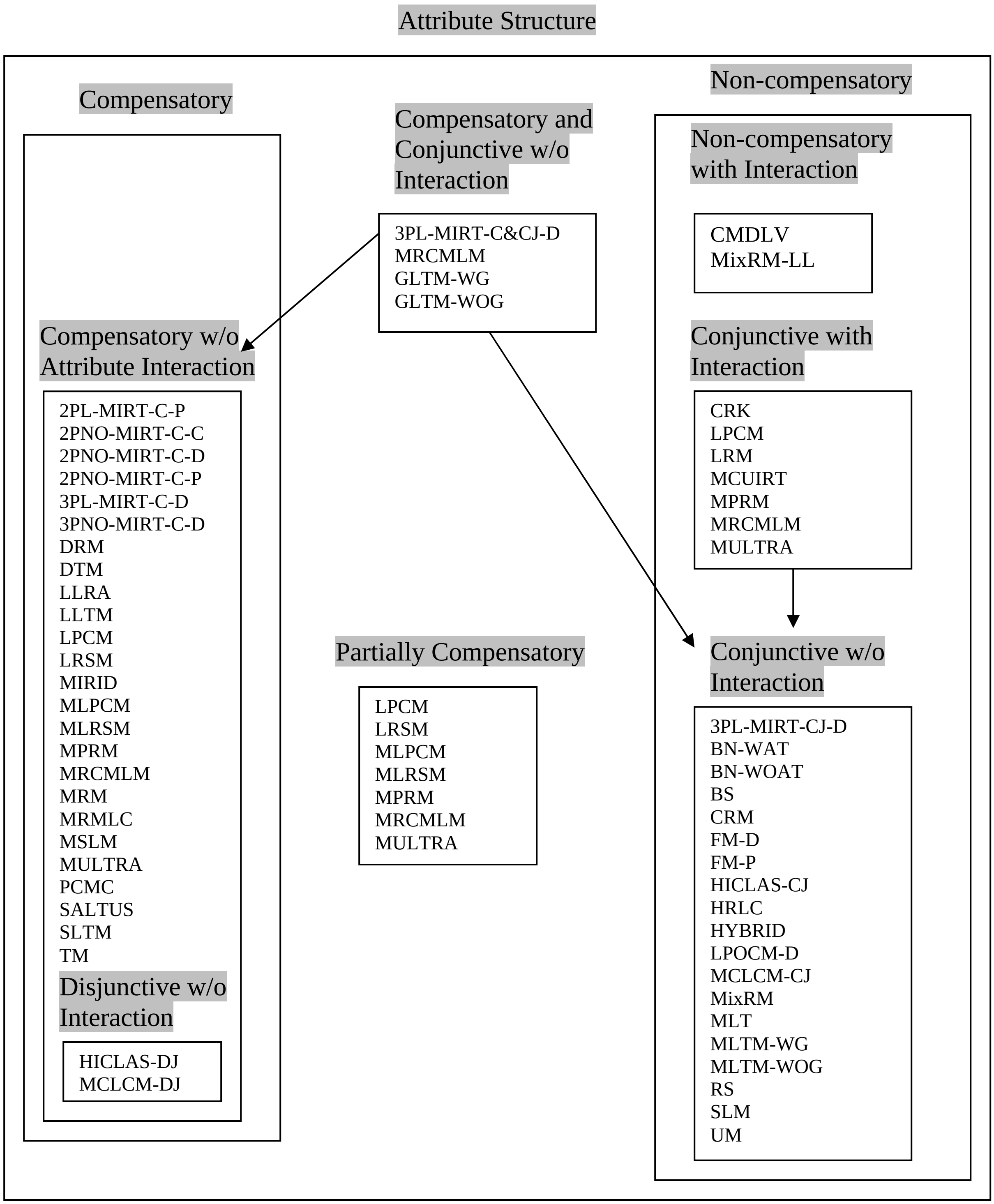
Note. For the full name of each acronym see Table 1.

Figure Caption

Figure 1. The hierarchical structure of CDPMs. Solid ovals represent CDPMs with all continuous attributes, and dashed ovals represent CDPMs having ordinal attribute(s). Arrows represent nesting relationship; for example, all the models are the sub-models of the general CDPM. See Table 1 for the full name of each acronym.

Figure 2. The classification of CDPM attribute structures. Embedded rectangles represent conceptual nested relationship (e.g., the extension of the definition of *compensatory* contains that of *disjunctive*) and arrows represent structural nested relationships (e.g., a conjunctive model with interaction can be simplified to a conjunctive model without interaction). See Table 1 for the full name of each acronym.


General CDPM
MCLCM-DJ
2PNO-MIRT-C-C
DTM
SLM
MRCMLM
MIRID
TBA
MCUIRT
GLTM-WG
RS
TM-2PNO-P
3PL-MIRT-CJ&C-D
2PNO(PL)-MIRT-C-P
MLT
HRLC
CMDLV
FM-P
UM
HYBRID
BN-WAT
HICLAS-DJ
CRK
GLTM-WOG
MPRM
MLPCM
MLTM-WG
3PL-MIRT-CJ-D
3PNO(PL)-MIRT-C-D
MixRM
MixRM-LL
FM-D
LRM
MLTM-WOG
MULTRA
MMRM
LPCM
MLRSM
CRM
SLTM
MSLM
2PNO-MIRT-C-D
TM-3PL-D
SALTUS
MCLCM-CJ
MRM
PCMC
MRMLC
LRSM
LLRA
TM-2PNO-D
LPOCM-D, BS, BN-WOAT
LLTM
DRM


Attribute Structure
Compensatory
Compensatory w/o Attribute Interaction
2PL-MIRT-C-P
2PNO-MIRT-C-C
2PNO-MIRT-C-D
2PNO-MIRT-C-P
3PL-MIRT-C-D
3PNO-MIRT-C-D
DRM
DTM
LLRA
LLTM
LPCM
LRSM
MIRID
MLPCM
MLRSM
MPRM
MRCMLM
MRM
MRMLC
MSLM
MULTRA
PCMC
SALTUS
SLTM
TM
Disjunctive w/o Interaction
HICLAS-DJ
MCLCM-DJ
Compensatory and Conjunctive w/o Interaction
3PL-MIRT-C&CJ-D
MRCMLM
GLTM-WG
GLTM-WOG
Partially Compensatory
LPCM
LRSM
MLPCM
MLRSM
MPRM
MRCMLM
MULTRA
Non-compensatory
Non-compensatory with Interaction
CMDLV
MixRM-LL
Conjunctive with Interaction
CRK
LPCM
LRM
MCUIRT
MPRM
MRCMLM
MULTRA
Conjunctive w/o Interaction
3PL-MIRT-CJ-D
BN-WAT
BN-WOAT
BS
CRM
FM-D
FM-P
HICLAS-CJ
HRLC
HYBRID
LPOCM-D
MCLCM-CJ
MixRM
MLT
MLTM-WG
MLTM-WOG
RS
SLM
UM