British Journal of Mathematical and Statistical Psychology (2019)

The British Psychological Society

www.wileyonlinelibrary.com

Cognitive diagnosis models for multiple strategies

Wenchao Ma* and Wenjing Guo

Department of Educational Studies in Psychology, Research Methodology and Counseling, The University of Alabama, Tuscaloosa, Alabama, USA

Cognitive diagnosis models (CDMs) have been used as psychometric tools in educational assessments to estimate students' proficiency profiles. However, most CDMs assume that all students adopt the same strategy when approaching problems in an assessment, which may not be the case in practice. This study develops a generalized multiple-strategy CDM for dichotomous response data. The proposed model provides a unified framework to accommodate various condensation rules (e.g., conjunctive, disjunctive, and additive) and different strategy selection approaches (i.e., probability-matching, over-matching, and maximizing). Model parameters are estimated using the marginal maximum likelihood estimation via expectation-maximization algorithm. Simulation studies showed that the parameters of the proposed model can be adequately recovered and that the proposed model was relatively robust to some types of model misspecifications. A set of real data was analysed as well to illustrate the use of the proposed model in practice.

3 condensation rules:
-conjunctive
-disjunctive
-additive

strategy selection approach
- probability matching
-over matching
-maximizing

Estimation procedure
- MMLE via EM-ALGGO

1. Introduction

Cognitive diagnostic assessments are intended to build students' proficiency profiles of a number of skills of interest. In the psychometric literature, skills, such as addition, subtraction, and multiplication in an elementary mathematics test, are usually replaced by a generic term 'attribute', which can also represent cognitive process and psychological disorders, among others. To infer students' attribute profiles, cognitive diagnosis models (CDMs) have been shown to be promising. As restricted latent class models, CDMs group students into different latent classes, each with a unique attribute profile. The attribute profile depicts a student's strengths and weaknesses, and thus has the potential to facilitate teaching and learning.

A large number of CDMs have been developed in the literature (see Rupp, Templin, & Henson, 2010), but most of them assume that all students use the same strategy to solve problems. This might be too restrictive in practice. As noted by Peters, Smedt, Torbeyns, Ghesquiere, and Verschaffel (2013), current mathematics education emphasizes the importance of teaching students a variety of strategies to solve problems efficiently and flexibly. As a result, students often have multiple ways of approaching a problem. For instance, Star and Rittle-Johnson (2008) discussed three strategies for solving algebra equation problems such as $3(x+1) = 15$, and Peters *et al.* (2013) distinguished two strategies for solving two-digit subtraction problems such as 75–52. In addition to elementary mathematics, Larkin, McDermott, Simon, and Simon (1980) presented

*Correspondence should be addressed to Wenchao Ma, Department of Educational Studies in Psychology, Research Methodology and Counseling, University of Alabama, Box 870231, Room 307B, 520 Colonial Drive, Tuscaloosa 35487, AL, USA (email: wenchao.ma@ua.edu).

DOI:10.1111/bmsp.12155

different strategies that can be used for solving problems in physics, and Olshavsky (1976) identified 10 strategies involved in reading comprehension.

Ignoring multiple strategies and fitting data using single-strategy CDMs could result in model misspecifications and an inadequate model-data fit, which, in turn, causes concerns as to the validity of inferences. Until now, only a few CDMs have been capable of accommodating multiple strategies. One of these is the Unified model (DiBello, Stout, & Roussos, 1995), which, unfortunately, cannot be estimated statistically. The diagnostic tree model (DTM; Ma, 2019) is another CDM that allows multiple strategies used in polytomous response data from constructed-response items. However, it requires that the strategy used by each student for an item can be identified, which is challenging for dichotomous response data from multiple-choice items. Additionally, the mixture distribution models, such as the mixture item response models (e.g., Mislevy & Verhelst, 1990; Rost, 1990) and mixture general diagnostic model (von Davier, 2007, 2008b), have been used for accommodating multiple strategies in educational assessment (for an overview, see von Davier, 2010). One major limitation of the mixture models is that individuals are usually assumed to use a single strategy for all questions in a test, although Mislevy and Verhelst (1990) discussed briefly a possible solution to relax this assumption and Molenaar and de Boeck (2018) proposed a response mixture model for responses and response time that allows individuals to switch strategies. Another approach to multiple strategies is the multiple-strategy deterministic inputs, noisy 'and' gate (MS-DINA; de la Torre & Douglas, 2008) model. It assumes that all items involve the same number of strategies and that, to answer an item correctly, students are expected to master all attributes required by at least one strategy. In addition, this model assumes that item parameters across different strategies are the same and that the use of each strategy is equally difficult. These assumptions make the MS-DINA model highly restrictive.

MS-DINA:
- assumes that all items involve the same number of strategies
- to answer an item correctly, students are expected to master all attributes required by at least one strategy.
- assumes that item parameters across different strategies are the same and that the use of each strategy is equally difficult

In this study, we develop a generalized multiple-strategy CDM for dichotomous response data, which substantially extends the MS-DINA model to provide a more flexible framework for modelling students' responses under multiple strategies. The proposed model uses several popular single-strategy dichotomous CDMs as building blocks, and takes various strategy selection approaches into consideration. The rest of the paper is organized as follows. In the next section, we review several single-strategy CDMs and the MS-DINA model. In Section 3, we introduce the generalized multiple-strategy CDM and we discuss its relation with other psychometric models. In Section 4, we describe in detail two simulation studies for evaluating the viability of the proposed model under varied conditions. We analyse a set of fraction subtraction data in Section 5 to illustrate the use of the proposed model in practice. We conclude in Section 6 with a summary of current work, and a discussion of directions for future research.

-Q-matrix: does item j measure item k

- k attributes yield c = 2^k latent classes

-condensation rule: tells what attributes I need to possess to get the item correct

2. Background

Most CDMs, if not all, consist of two major components: the Q-matrix (Tatsuoka, 1983), which relates items to attributes; and the condensation rule (Maris, 1999), which defines how attributes interact with each other. Suppose a test has J items measuring K binary attributes. The Q-matrix is of dimensions $J \times K$, with element $q_{jk} = 1$ if item j measures attribute k and $q_{jk} = 0$ if item j does not measure attribute k. Also, K binary attributes yield 2^K latent classes, each with a unique attribute profile. The attribute profile for latent class c

is denoted as $\boldsymbol{\alpha}_c = [\alpha_{c1}, \ldots, \alpha_{cK}]^\top$, where $\alpha_{ck} = 1$ if attribute k is mastered and $\alpha_{ck} = 0$ if attribute k is not mastered. Let Y_{ij} be a response variable of individual i to item j, and y_{ij} a realization. Also, denote $\boldsymbol{Y}_i$ as item response vector of individual i and $\boldsymbol{Y} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N]^\top$ the item responses of N individuals.

IRF: prob of ind i with profile c, answering item j correctly

The CDMs based on different condensation rules have distinct forms of item response function (IRF), which is the probability of an individual with attribute profile α_c answering item j correctly, and is denoted by $P_j(\boldsymbol{\alpha}_c)$. The deterministic inputs, noisy 'and' gate (DINA; Haertel, 1989) model is based on the conjunctive rule, and its IRF can be written by

a_{c1}q_{11} a_{c2}q_{12} ... a_{ck}q_{1k}

1 - P = s

slipping parameter = aka probability of failure

$$P_j(\boldsymbol{\alpha}_c) = \delta_{j0} + \delta_{j1} \prod_{k=1}^{K} \alpha_{ck}^{q_{jk}}, \tag{1}$$

probability of success for indiv with profile alpha sub c

We only concern ourself with required attributes of subcripts k in which q vector is 1 (for DINA, we have to possess those to get the product notation 1.)

where δ_{j0} is the guessing parameter, representing the success probability for individuals who do not master all required attributes, and δ_{j1} is the increase in success probability when all required attributes are mastered. In contrast, the deterministic input, noisy 'or' gate (DINO; Templin & Henson, 2006) model is based on the disjunctive condensation rule with the following IRF:

$$P_j(\boldsymbol{\alpha}_c) = \delta_{j0} + \delta_{j1}\left[1 - \prod_{k=1}^{K}(1-\alpha_{ck})^{q_{jk}}\right], \tag{2}$$

where δ_{j0} is the guessing parameter, and δ_{j1} represents the increase in the success probability when any of the required attributes are mastered. In addition, de la Torre (2011) defines additive CDMs (ACDMs) under identity, log and logit link functions. The IRFs of these models can be written by

$$g[P_j(\boldsymbol{\alpha}_c)] = \delta_{j0} + \sum_{k=1}^{K} \delta_{jk}\alpha_{ck}q_{jk}, \tag{3}$$

parang logistic regression lang kapag ang link ay logit. However, in this case, estimated ang alpha, and that is needed to get the value for the categorical variable alpha*q. Unlike sa logistic regression na given sya as data

where δ_{j0} is the guessing parameter, and δ_{jk} represents the contribution due to the mastery of attribute k. As noted by de la Torre (2011), the logit and log link ACDMs are equivalent to the linear logistic model (LLM; Maris, 1999), and the reduced reparametrized unified model (RRUM; Hartz, 2002), respectively. Therefore, unless otherwise stated, the ACDM refers to the identity link model in this article.

Unlike the aforementioned CDMs, the MS-DINA model (de la Torre & Douglas, 2008) assumes that there are M q-vectors for item j corresponding to M strategies. Let $\boldsymbol{q}_j^m = \{q_{jk}^m\}$ denote the q-vector of item j for strategy m. The IRF of the MS-DINA model can be formulated as

$$P_j(\boldsymbol{\alpha}_c) = \delta_{j0} + \delta_{j1}\left[1 - \prod_{m=1}^{M}\left(1 - \prod_{k=1}^{K}\alpha_{ck}^{q_{jk}^m}\right)\right], \tag{4}$$

which assumes that individuals who master all attributes required by at least one strategy have a success probability of $\delta_{j0} + \delta_{j1}$, whereas those who do not master all attributes required by at least one strategy have a success probability of δ_{j0}. As a result, there are two parameters for each item only.

3. A generalized multiple-strategy model

In this section, a generalized multiple-strategy (GMS) CDM is presented. Before introducing the model, we summarize some literature on three strategy selection approaches. The first approach is called the maximizing approach. Based on this approach, individuals choose the strategy that has the highest success probability to maximize the expected item score. For example, if the probabilities of an individual answering an item correctly using strategies A and B are 0.8 and 0.5, respectively, the individual will always choose strategy A. The maximizing approach is in accordance with the expected utility theory, which is one of the most important theories on decisions under uncertainty in economics. However, there is little dispute that humans often violate this rational decision-making principle because of cognitive limitations and memory failure (e.g., Fazio, Dewolf, & Siegler, 2016; Wozny, Beierholm, & Shams, 2010). The second approach, a probability-matching approach, is found to be used by humans for various cognitive tasks (e.g., Gaissmaier & Schooler, 2008; Lovett & Anderson, 1995; Wozny *et al.*, 2010). The probability-matching approach is in line with the adaptive strategy choice model of Siegler, Adolph, and Lemaire (1996, p. 92), which assumes that the 'probability of choosing a particular strategy is proportional to that strategy's projected strength [a function of the strategy's past accuracy when all else being equal] relative to that of all strategies combined'. For the aforementioned example, the student will have a probabilities of 0.8/(0.8 + 0.5) and 0.5/(0.8 + 0.5) to use strategies A and B, respectively. Unlike the maximizing approach, the probability-matching approach is suboptimal in that it assumes that individuals might use inferior strategies. In addition to the maximizing and probability-matching approaches, Lovett and Anderson (1995, p. 269) found that some individuals tended to use the over-matching approach, which means individuals 'use the more successful strategy more often than its proportion of solutions [according to the probability-matching approach]'. For the previous example, according to the over-matching approach, the probability of a student adopting strategy A would be greater than 0.8/(0.8+0.5).

3.1. Model specification

Suppose that item j has M_j strategies, each involving a unique subset of the K attributes. This implies that different items might have different numbers of strategies and that different strategies correspond to distinct q-vectors. In addition, let $\boldsymbol{q}_j^m = \{q_{jk}^m\}$ denote the q-vector of item j for strategy m. The GMS-CDM defines the likelihood of observing response vector $\boldsymbol{y}_i$ for individual i with attribute profile $\boldsymbol{\alpha}_c$ as

joint probability of the data given theta

$$L(\boldsymbol{y}_i) = \sum_c \pi_c \prod_j^J \underbrace{P_j(\boldsymbol{\alpha}_c)^{y_{ij}}}_{\text{prob of success}} \underbrace{[1 - P_j(\boldsymbol{\alpha}_c)]^{1-y_{ij}}}_{\text{prob of failure}}, \tag{5}$$

where π_c (which is discussed later) defines the joint attribute distribution, and the IRF, or $P_j(\boldsymbol{\alpha}_c)$ term, is formulated as

prob of success for each item

$$P_j(\boldsymbol{\alpha}_c) = \sum_{m=1}^{M_j} P(Y_{ij} = 1|\boldsymbol{\alpha}_c, m)P_j(m|\boldsymbol{\alpha}_c). \tag{6}$$

Here, $P(Y_{ij} = 1|\boldsymbol{\alpha}_c, m)$ is the probability of individual i given attribute profile $\boldsymbol{\alpha}_c$ answering item j correctly using strategy m, which is referred to as a strategy-specific IRF. For simplicity, subscript i will be dropped when there is no ambiguity. In addition, $P_j(m|\boldsymbol{\alpha}_c)$ is the probability of an individual with attribute profile $\boldsymbol{\alpha}_c$ choosing strategy m for item j. We further define

$$P_j(m|\boldsymbol{\alpha}_c) = \frac{P(Y_j = 1|\boldsymbol{\alpha}_c, m)^s}{\sum_{m=1}^{M_j} P(Y_j = 1|\boldsymbol{\alpha}_c, m)^s}, \tag{7}$$

where s is referred to as a strategy selection parameter. When $s = 0$, all strategies have the same probability of being selected. When $s = 1$, the strategy is chosen using the probability-matching approach. When $s > 1$, over-matching is modelled, which means the strategy with a higher success probability is more likely to be selected. The larger s is, the more likely the strategy with the highest success probability is chosen. As a result, when $s \to +\infty$, the strategy is chosen approximately using the maximizing approach.

For illustration, Figure 1 displays $P_j(\boldsymbol{\alpha}_c)$ for different strategy selection parameters under two strategies. For Strategy A, $P(Y_j = 1|\boldsymbol{\alpha}_c, m = A)$ is given on the x-axis and ranges

s =100, is the maximizing probability. this means that across different probs of success, choosing any available strategy means choosing it so that the strategy associated to highest prob of success is chosen

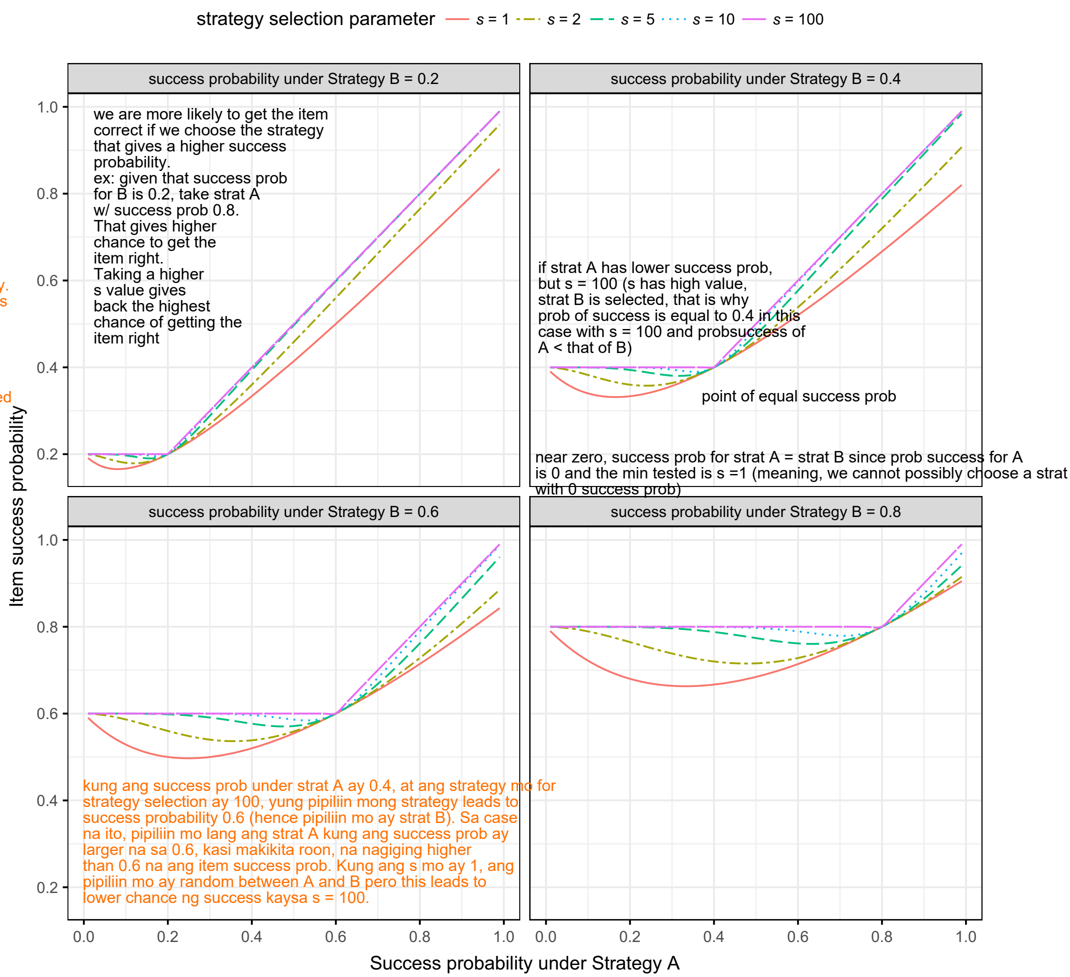

Figure 1. Success probabilities with different strategy selection parameters.

from 0 to 1. Four panels represent $P(Y_j = 1|\boldsymbol{\alpha}_c, m = B) = 0.2$, 0.4, 0.6 and 0.8, respectively, for Strategy B. Based on equations (6) and (7), $P_j(\boldsymbol{\alpha}_c)$ is calculated and shown on the y-axis. Five lines represent s = 1, 2, 5, 10 and 100, respectively. Several interesting findings can be observed. First, when s = 10 or 100, we have

$$P_j(\boldsymbol{\alpha}_c) \approx \max[P(Y_j = 1|\boldsymbol{\alpha}_c, m = A), P(Y_j = 1|\boldsymbol{\alpha}_c, m = B)]. \tag{8}$$

max between success prob under strat A and success prob under strat B

This implies that the GMS-CDM with s = 10 could provide a good approximation to the maximizing strategy selection approach. Secondly, the maximizing selection approach always yielded the highest item success probabilities, and the probability-matching approach the lowest. Item success probabilities based on the over-matching selection approach were in-between. Thirdly, when the success probabilities were identical for two strategies, or very close to zero for any one of the strategies, the item success probabilities based on different strategy selection methods were virtually identical.

Equations (6) and (7) define the framework of the GMS-CDM within which the strategy-specific IRF can be modelled using various dichotomous CDMs. In this study, we assume that for an item, all strategies involve the same condensation rule, and thus all strategy-specific IRFs for the same item have the same form. Specifically, the strategy-specific IRF of the GMS-DINA model is parametrized as

all strategies involve the same condensation rule - ALL strategy specific IRFs for the same item have the same form (ex: if DINA sa item 1, all other possible strategies ay DINA rin)

$$P(Y_j = 1|\boldsymbol{\alpha}_c, m) = \delta_{j0} + \delta_{jm1} \prod_{k=1}^{K} \alpha_{ck}^{q_{jk}^m}, \tag{9}$$

where δ_{j0} is the success probability of individuals who do not master all attributes required by any of the M_j strategies, and δ_{jm1} is the increase in success probability as a result of mastering all the required attributes of strategy m. Compared with the MS-DINA model (de la Torre & Douglas, 2008), the GMS-DINA model has $M_j + 1$ parameters for item j and allows for individuals following different strategies to have different success probabilities.

The strategy-specific IRF of the GMS-DINO model is formulated as

M_j + 1 = j different delta_jm1 + guessing parameter

$$P(Y_j = 1|\boldsymbol{\alpha}_c, m) = \delta_{j0} + \delta_{jm1}\left[1 - \prod_{k=1}^{K}(1 - \alpha_{ck})^{q_{jk}^m}\right], \tag{10}$$

where δ_{j0} is the success probability of individuals who do not master any required attributes for any of the M_j strategies, and δ_{jm1} is the increase in success probability as a result of mastering any required attributes of strategy m. Like the GMS-DINA model, the GMS-DINO model involves $M_j + 1$ parameters for item j and allows different strategies to have different difficulties.

For GMS-ACDM, LLM and RRUM, the strategy-specific IRF can be written by

$$g[P(Y_{ij} = 1|\boldsymbol{\alpha}_c, m)] = \delta_{j0} + \sum_{k=1}^{K} \delta_{jk}\alpha_{ck}q_{jk}^m, \tag{11}$$

where $g[\cdot]$ is the identity, logit or log link function. Intercept δ_{j0} is the baseline parameter for individuals who do not master any required attributes for all M_j strategies and δ_{jk} is the increase as a result of mastering attribute k. Note that $\delta_{jk} \equiv 0$ if $\sum_{m=1}^{M_j} q_{jk}^m = 0$. Also, under different strategies that involve attribute k, the contributions of attribute k are assumed to be identical. As a result, there are $1 + K_j$ parameters for item j, where K_j is the number of unique attributes required by all M_j strategies; it can be calculated by $K_j = \sum_{k=1}^{K}[1 - \prod_{m=1}^{M_j}(1 - q_{jk}^{m})]$.

The joint attribute distribution π_c in equation (5) can be parametrized in various ways (Maris, 1999). The saturated model treats π_c as parameters for $c = 1, \ldots, 2^K$, with the constraint that $\sum_{c=1}^{2^K} \pi_c = 1$, where π_c represents the proportion of individuals who have an attribute profile $\boldsymbol{\alpha}_c$. To reduce the number of parameters involved in the joint attribute distribution, a loglinear model (Xu & von Davier, 2008) or a higher-order IRT model (de la Torre & Douglas, 2004) can be used. In this study, the saturated model is used because it is the most general parametrization.

3.2. Parameter estimation

Let $\boldsymbol{\gamma} = [\boldsymbol{\delta}^\top, \boldsymbol{\pi}^\top]^\top$ denote a vector of model parameters, including item parameters $\boldsymbol{\delta}$ involved in the strategy-specific IRFs and mixing proportion parameters $\boldsymbol{\pi} = \{\pi_c\}$. Note that the strategy selection parameter s is assumed to be known and does not need to be estimated. As in Bock and Aitkin (1981), γ can be estimated using the marginal maximum likelihood estimation via expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). Specifically, instead of finding γ that directly maximizes

$$L(\boldsymbol{y}) = \prod_{i=1}^{N} L(\boldsymbol{y}_i), \tag{12}$$

the EM algorithm maximizes the complete-data log-likelihood iteratively. More specifically, the EM algorithm consists of two steps: the expectation (E) step and the maximization (M) step. In the E-step, we calculate the so-called Q-function (Dempster *et al.*, 1977), which is the expected log-likelihood of the complete-data conditional on the observed data and current parameter estimates. With independent individuals, on the $(t+1)$ th iteration, the Q-function takes the following form:

$$Q(\gamma|\gamma^{(t)}) = \sum_{c=1}^{2^K} n_c \log[\pi_c] + \sum_{j=1}^{J}\sum_{c=1}^{2^K}\Big[r_{jc}\log\big[P_j(\boldsymbol{\alpha}_c)\big] + (n_c - r_{jc})\log\big[1 - P_j(\boldsymbol{\alpha}_c)\big]\Big]. \tag{13}$$

Here, $\gamma^{(t)}$ is the vector of parameters from the tth iteration, n_c is the expected number of individuals in latent class c and r_{jc} is the expected number of individuals in latent class c who answer item j correctly. These can be calculated by

$$\begin{aligned} n_c &= \sum_{i=1}^{N} P(\boldsymbol{\alpha}_c|\boldsymbol{y}_i, \gamma^{(t)}), \\ r_{jc} &= \sum_{i=1}^{N} y_{ij} P(\boldsymbol{\alpha}_c|\boldsymbol{y}_i, \gamma^{(t)}), \end{aligned} \tag{14}$$

where $P(\boldsymbol{\alpha}_c|\boldsymbol{y}_i, \gamma^{(t)})$ is the posterior probability of individual i being assigned to latent class c. It can be calculated using the Bayes rule:

$$P(\boldsymbol{\alpha}_c|\boldsymbol{y}_i, \boldsymbol{\gamma}^{(t)}) = \frac{P(\boldsymbol{y}_i|\boldsymbol{\alpha}_c, \boldsymbol{\gamma}^{(t)})\pi_c^{(t)}}{\sum_c P(\boldsymbol{y}_i|\boldsymbol{\alpha}_c, \boldsymbol{\gamma}^{(t)})\pi_c^{(t)}}, \tag{15}$$

and

$$P(\boldsymbol{y}_i|\boldsymbol{\alpha}_c, \boldsymbol{\gamma}^{(t)}) = \prod_{j=1}^{J} \left[P(Y_{ij} = 1|\boldsymbol{\alpha}_c, \boldsymbol{\gamma}^{(t)})\right]^{y_{ij}} \left[1 - P(Y_{ij} = 1|\boldsymbol{\alpha}_c, \boldsymbol{\gamma}^{(t)})\right]^{1-y_{ij}}. \tag{16}$$

In the M-step, the Q-function is maximized. The E- and M-steps repeat until certain convergence criteria have been met.

3.3. Relation to other psychometric models

The proposed GMS-CDMs are related to several existing psychometric models. First, the MS-DINA model of de la Torre and Douglas (2008) can be viewed as a constrained GMS-DINA model. Specifically, when $\delta_{jm1} = \delta_{j1}$ for all M_j strategies, and $s \to +\infty$, the GMS-DINA is equivalent to the MS-DINA model. With these constraints, there are only two parameters for each item regardless of the number of strategies. In practice, the GMS-DINA with a large s (e.g., 10) can provide a good approximation to the MS-DINA model.

Second, the GMS-CDMs are also related to the mixture distribution models (e.g., Mislevy & Verhelst, 1990; Rost, 1990; von Davier, 2008b, 2010). The mixture approach typically assumes that a student uses a certain strategy for all items in an assessment, although different students might adopt distinct strategies. In contrast, the GMS-CDMs allow a student to switch strategies for different items. This is aligned with the overlapping waves theory (Siegler *et al.*, 1996), which claims that students tend to adaptively select strategies according to the questions and situations. In addition, the mixture model usually involves the estimation of the probability of adopting each strategy for each individual. Nevertheless, the GMS-CDMs assume that students with the same attribute profile have the same probability of adopting each strategy for an item. The probability can be calculated based on strategy-specific IRFs and the strategy selection parameter. In summary, compared with mixture distribution models, the GMS-CDMs not only allow us to calculate the probability of a student choosing each strategy for each item, but also allow students to switch strategies for different items.

Last, the GMS-CDMs can be viewed as a variant of the DTM of Ma (2019), which incorporates CDMs into tree models to accommodate multiple strategies for polytomous response data from constructed-response items. Similar to the DTM approach, the GMS-CDMs proposed in this paper can be presented using the tree diagram, as shown in Figure 2, where the first node (i.e., X_0) represents the task for strategy selection with M possible outcomes for M strategies. After strategy m is selected, students need to undertake corresponding tasks, which are represented by node X_m with two possible outcomes (i.e., success or failure). Note that the outcomes related to X_0 are unobserved but those for X_m are observed. Although both the proposed GMS-CDM and the DTM can be represented using tree diagrams, they differ substantially in the parametrization of each node.

4. Simulation studies

Two simulation studies were conducted to evaluate the parameter recovery of the GMS-CDMs based on the MMLE/EM algorithm, and to examine the impact of model

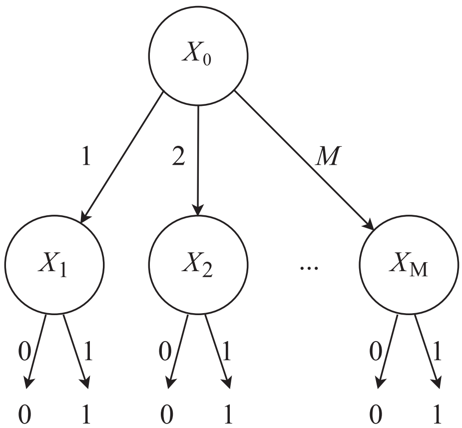

Figure 2. Tree diagram of GMS-CDMs.

misspecifications under various conditions. Although several GMS-CDMs were proposed in the previous section, only GMS-DINA and GMS-ACDM were examined because (1) the DINA and DINO models are similar due to the duality relation (Köhn & Chiu, 2016; Liu, Xu, & Ying, 2011), and (2) the ACDM, LLM and RRUM are related in that they are all main-effect models with the only difference in the link functions (e.g., Ma, Iaconangelo, & de la Torre, 2016).

4.1. Simulation study I: Parameter recovery

4.1.1. Design

For the first simulation study, the sample size $N = 1{,}000$, 2,000 and 4,000. The test length was fixed at $J = 30$ and the number of attributes was $K = 5$. There were two strategies for each item. The Q-matrix for each strategy is given in Table 1, which was created to ensure that each attribute was measured the same number of times under the two strategies and that the Q-matrix was complete (Chiu, Douglas, & Li, 2009) for each strategy. The GMS-DINA and GMS-ACDM with strategy selection parameter $s = 1$, 2 and 10 (representing the probability-matching, over-matching and maximizing strategy selection approaches) were used to generate item responses. Specifically, $\delta_{j0} \sim U[0.05, 0.35]$ and $P(Y_j|\boldsymbol{\alpha}_c = 1, m) \sim U[0.65, 0.95]$. For the GMS-DINA model, $\delta_{jm1} = P(Y_j|\boldsymbol{\alpha}_c = 1, m) - \delta_{j0}$. For the GMS-ACDM, δ_{jk} was randomly chosen from uniform distributions with the constraint that $\sum_{k=1}^{K} \delta_{jk}\alpha_{ck}q_{jk}^{m} = P(Y_j = 1|\boldsymbol{\alpha}_c, m) - \delta_{j0}$. The attribute profiles of individuals were drawn from either a uniform distribution, where all possible attribute profiles have the same chance of being selected, or a higher-order distribution (de la Torre & Douglas, 2004), where the conditional probability of attribute k being mastered is

$$P(\alpha_k = 1|\theta, b_k) = \frac{\exp[1.7(\theta - b_k)]}{1 + \exp[1.7(\theta - b_k)]}. \tag{17}$$

Here, $\theta \sim N(0, 1)$ is the higher-order ability and $b_k = -1 + 0.5(k - 1)$ is the difficulty of attribute k for $k = 1, \ldots, K$. Under each condition, 500 data sets were generated and fitted using the generating model. Parameter estimation for the GMS-CDMs were implemented in R (R Core Team, 2018) and is now included in the GDINA R package (Ma & de la Torre, 2019).

Table 1. Q-matrix for simulation studies

	Strategy A					Strategy B				
Item	α_1	α_2	α_3	α_4	α_5	α_1	α_2	α_3	α_4	α_5
1	1	0	0	0	0	0	1	0	0	0
2	0	0	1	0	0	0	0	0	1	0
3	0	0	0	0	1	1	0	0	0	0
4	0	1	0	0	0	0	0	1	0	0
5	0	0	0	1	0	0	0	0	0	1
6	1	1	0	0	0	0	0	1	1	0
7	0	0	1	0	1	1	0	0	1	0
8	0	1	0	1	0	1	0	0	0	1
9	0	0	0	1	1	0	1	1	0	0
10	1	0	1	0	0	0	1	0	0	1
11	1	1	1	0	0	0	0	1	1	1
12	0	1	0	1	1	1	1	0	0	1
13	0	1	1	1	0	1	1	0	1	0
14	1	0	0	1	1	0	1	1	0	1
15	1	0	1	0	1	1	0	1	1	0
16	1	0	0	0	0	0	1	0	0	0
17	0	0	1	0	0	0	0	0	1	0
18	0	0	0	0	1	1	0	0	0	0
19	0	1	0	0	0	0	0	1	0	0
20	0	0	0	1	0	0	0	0	0	1
21	1	1	0	0	0	0	0	1	1	0
22	0	0	1	0	1	1	0	0	1	0
23	0	1	0	1	0	1	0	0	0	1
24	0	0	0	1	1	0	1	1	0	0
25	1	0	1	0	0	0	1	0	0	1
26	1	1	1	0	0	0	0	1	1	1
27	0	1	0	1	1	1	1	0	0	1
28	0	1	1	1	0	1	1	0	1	0
29	1	0	0	1	1	0	1	1	0	1
30	1	0	1	0	1	1	0	1	1	0

4.1.2. Criteria

To evaluate item parameter recovery of the GMS-DINA and GMS-ACDM, the average bias and root mean square error (RMSE) of item parameter estimate $\hat{\delta}_{j\cdot}$ were calculated by

$$\text{bias} = \frac{\sum_{r=1}^{R}\sum_{j=1}^{J}(\hat{\delta}_{j\cdot} - \delta_{j\cdot})}{R \times J} \tag{18}$$

and

$$\text{RMSE} = \sqrt{\frac{\sum_{r=1}^{R}\sum_{j=1}^{J}(\hat{\delta}_{j\cdot} - \delta_{j\cdot})^2}{R \times J}}, \tag{19}$$

respectively. Person parameter recovery was evaluated using the proportion of correctly classified attribute (PCA) and the proportion of correctly classified attribute vectors (PCV) defined as

$$\text{PCA} = \frac{\sum_{r=1}^{R}\sum_{i=1}^{N}\sum_{k=1}^{K} 1^{(r)}_{[\alpha_{ik}=\hat{\alpha}_{ik}]}}{R\times N\times K}, \tag{20}$$

and

$$\text{PCV} = \frac{\sum_{r=1}^{R}\sum_{i=1}^{N} 1^{(r)}_{[\boldsymbol{\alpha}_i=\hat{\boldsymbol{\alpha}}_i]}}{R\times N}, \tag{21}$$

respectively. Here, $1^{(r)}_{[\boldsymbol{\alpha}_i=\hat{\boldsymbol{\alpha}}_i]}$ is an indicator variable with an outcome of 1 if the estimated attribute vector matches the true for the rth replication, and 0 otherwise. Note that N, R and J are the number of individuals, replications and items, respectively.

4.1.3. Results

Tables 2 and 3 give the biases of item parameter estimates and Tables 4 and 5 present the RMSEs under varied conditions. There are three parameters per item for the GMS-DINA model (i.e., δ_{j0} and δ_{j11} for strategy 1 and δ_{j21} for strategy 2) and at most six parameters per item for the GMS-ACDM (i.e., δ_{j0} and δ_{jk} for $k = 1,\ldots,5$). Several findings can be observed. First, the biases of item parameter estimates were <0.01 under almost all conditions. The only exception is δ_{j11} of the GMS-DINA model, which had a bias of -0.011 when $N = 1{,}000$, $s = 10$ and higher-order attribute distribution. Second, compared with other parameters, δ_{j0} always had smaller RMSEs, but sometimes slightly larger biases. For the GMS-DINA model, δ_{j11} produced similar RMSEs as δ_{j21} when attributes were uniformly distributed, but larger RMSEs when attributes were drawn from higher-order distributions. For the GMS-ACDM, δ_{jk} yielded similar RMSEs for all attributes when they were drawn from uniform distribution. When attributes were drawn from the higher-order distribution, δ_{jk} had smaller RMSEs if attribute k was easier, and larger RMSEs if attribute k was harder. Third, it is not unexpected that as N increased, the RMSEs decreased for both GMS-DINA and GMS-ACDM. Last, in general, the RMSEs were larger when $s = 1$ and smaller when $s = 10$, and the RMSEs for $s = 2$ were in between these. Exceptions occurred for the GMS-DINA model when attributes were drawn from the higher-order distribution.

Tables 6 and 7 give the PCA and PCV for different models under varied conditions. As N or s increased, both PCA and PCV improved regardless of the model used. In addition,

Table 2. Biases of item parameter estimates of the GMS-DINA model across 30 items.

		$s = 1$			$s = 2$			$s = 10$		
Attribute	N	δ_{j0}	δ_{j11}	δ_{j21}	δ_{j0}	δ_{j11}	δ_{j21}	δ_{j0}	δ_{j11}	δ_{j21}
Uniform	1,000	−0.000	−0.002	−0.001	−0.000	−0.001	−0.000	−0.000	−0.001	−0.001
	2,000	−0.000	−0.001	0.000	−0.000	−0.000	0.000	−0.000	−0.000	−0.000
	4,000	−0.000	−0.000	−0.000	−0.000	0.000	−0.000	−0.000	−0.000	0.000
Higher-order	1,000	−0.001	−0.007	−0.003	−0.001	−0.007	−0.004	−0.001	−0.011	−0.005
	2,000	−0.000	−0.003	−0.001	−0.001	−0.004	−0.001	−0.000	−0.008	−0.003
	4,000	−0.000	−0.001	−0.000	−0.000	−0.002	−0.001	−0.000	−0.005	−0.001

Table 3. Biases of item parameter estimates of GMS-ACDM across 30 items

Attribute	N	s	δ_{j0}	δ_{j1}	δ_{j2}	δ_{j3}	δ_{j4}	δ_{j5}
Uniform	1,000	1	−0.006	0.002	0.004	0.003	0.003	0.005
		2	−0.006	0.002	0.003	0.003	0.005	0.002
		10	−0.004	0.001	0.003	0.001	0.003	0.002
	2,000	1	−0.003	0.001	0.001	0.001	0.002	0.002
		2	−0.003	0.002	0.001	0.002	0.002	0.001
		10	−0.003	0.000	0.001	0.001	0.001	0.001
	4,000	1	−0.002	−0.001	0.000	0.000	0.001	0.001
		2	−0.002	−0.000	0.001	0.000	0.001	0.001
		10	−0.002	0.000	−0.000	0.001	0.000	0.001
Higher-order	1,000	1	−0.002	−0.001	0.002	0.004	0.004	0.002
		2	−0.003	−0.001	0.002	0.003	0.004	−0.001
		10	−0.002	0.001	0.000	0.001	0.001	−0.001
	2,000	1	−0.001	−0.000	0.001	0.003	0.002	0.002
		2	−0.001	0.000	0.001	0.002	0.003	0.001
		10	−0.001	0.000	−0.000	0.002	0.001	−0.001
	4,000	1	−0.001	0.000	0.002	0.002	0.002	0.001
		2	−0.001	0.000	0.001	0.002	0.003	0.002
		10	−0.000	0.000	0.001	0.001	0.001	0.000

Note. N, sample size; s, strategy selection parameter.

Table 4. RMSEs of item parameter estimates of the GMS-DINA model across 30 items

		s = 1			s =2			s = 10		
Attribute	N	δ_{j0}	δ_{j11}	δ_{j21}	δ_{j0}	δ_{j11}	δ_{j21}	δ_{j0}	δ_{j11}	δ_{j21}
Uniform	1,000	0.026	0.063	0.063	0.024	0.050	0.051	0.023	0.048	0.048
	2,000	0.018	0.044	0.043	0.017	0.035	0.035	0.016	0.033	0.033
	4,000	0.013	0.031	0.031	0.012	0.024	0.025	0.011	0.023	0.023
Higher-order	1,000	0.026	0.089	0.074	0.024	0.084	0.066	0.023	0.091	0.068
	2,000	0.018	0.061	0.051	0.016	0.059	0.045	0.016	0.070	0.046
	4,000	0.013	0.041	0.035	0.012	0.041	0.032	0.011	0.052	0.033

Note. N, sample size; s, strategy selection parameter.

individuals tend to be better classified under higher-order attribute distribution than the uniform distribution. This is inconsistent with the observation that the item parameters were better recovered under the uniform attribute distribution. A possible reason is the expected numbers n_c and r_{jc} in equation (14) were estimated more stably under the uniform distribution because none of the latent classes was too small. This, in turn, led to more accurate item parameter estimates. Last, the GMS-DINA model tended to yield better person classifications than the GMS-ACDM under all conditions.

4.2. Simulation study II: Person classification accuracy under model misspecifications

The goal of the second simulation study is twofold: (1) to examine whether individuals can be accurately classified when the GMS-CDMs are employed for data generated using

Table 5. RMSEs of item parameter estimates of GMS-ACDM across 30 items

Attribute	s	N	δ_{j0}	δ_{j1}	δ_{j2}	δ_{j3}	δ_{j4}	δ_{j5}
Uniform	1	1,000	0.046	0.065	0.063	0.064	0.068	0.066
		2,000	0.031	0.045	0.044	0.043	0.047	0.045
		4,000	0.022	0.032	0.030	0.030	0.033	0.032
	2	1,000	0.043	0.054	0.052	0.053	0.056	0.054
		2,000	0.030	0.038	0.037	0.037	0.039	0.038
		4,000	0.020	0.027	0.026	0.026	0.027	0.027
	10	1,000	0.040	0.049	0.049	0.048	0.050	0.048
		2,000	0.027	0.034	0.033	0.034	0.035	0.034
		4,000	0.020	0.024	0.024	0.023	0.024	0.024
Higher-order	1	1,000	0.034	0.059	0.067	0.078	0.097	0.119
		2,000	0.023	0.042	0.046	0.052	0.068	0.083
		4,000	0.016	0.029	0.032	0.037	0.048	0.056
	2	1,000	0.032	0.049	0.056	0.066	0.085	0.113
		2,000	0.023	0.034	0.039	0.045	0.059	0.077
		4,000	0.016	0.024	0.028	0.032	0.042	0.053
	10	1,000	0.032	0.046	0.053	0.059	0.076	0.106
		2,000	0.022	0.032	0.035	0.042	0.052	0.072
		4,000	0.015	0.022	0.025	0.029	0.037	0.051

Note. N, sample size; s, strategy selection parameter.

Table 6. PCA of GMS-CDMs when models were correctly specified

		GMS-DINA			GMS-ACDM		
Attribute	N	$s = 1$	$s = 2$	$s = 10$	$s = 1$	$s = 2$	$s = 10$
Uniform	1,000	0.867	0.902	0.918	0.823	0.842	0.859
	2,000	0.872	0.904	0.921	0.833	0.849	0.865
	4,000	0.872	0.904	0.921	0.836	0.851	0.867
Higher-order	1,000	0.919	0.938	0.948	0.875	0.884	0.893
	2,000	0.921	0.940	0.950	0.885	0.892	0.902
	4,000	0.923	0.941	0.951	0.889	0.896	0.907

Note. N, sample size; PCA, proportion of correctly classified attribute; s, strategy selection parameter.

single-strategy models, or when the single-strategy models are fitted to data generated using multiple-strategy models; (2) to investigate the impact of misspecification in strategy selection parameters on person classification accuracy.

4.2.1. Design

The Q-matrices and item parameter generation were the same as in the first simulation study. Sample size was fixed at $N = 2{,}000$ and individuals' attribute profiles were drawn from the higher-order distribution with the same specifications as in the previous simulation study. Each of the DINA and GMS-DINA models with strategy selection parameter $s = 1$, 2 and 10 was used to simulate 500 data sets, which were then fitted by the DINA and GMS-DINA models. Also, each of the ACDM and GMS-ACDM models was

Table 7. PCV of GMS-CDMs when models were correctly specified

		GMS-DINA			GMS-ACDM		
Attribute	N	$s = 1$	$s = 2$	$s = 10$	$s = 1$	$s = 2$	$s = 10$
Uniform	1,000	0.607	0.708	0.754	0.428	0.492	0.538
	2,000	0.619	0.713	0.762	0.447	0.503	0.550
	4,000	0.618	0.713	0.761	0.451	0.507	0.554
Higher-order	1,000	0.709	0.773	0.810	0.547	0.584	0.613
	2,000	0.716	0.780	0.814	0.571	0.604	0.634
	4,000	0.719	0.781	0.816	0.582	0.611	0.646

Note. N, sample size; PCA, proportion of correctly classified attribute; s, strategy selection parameter.

Table 8. PCA of DINA and GMS-DINA models under model misspecifications

	Fitted model			
Generating model	DINA	$s = 1$	$s = 2$	$s = 10$
DINA	0.942	0.939	0.941	0.941
$s = 1$	0.816	0.921	0.919	0.919
$s = 2$	0.821	0.938	0.940	0.940
$s = 10$	0.825	0.946	0.950	0.950

Note. DINA, deterministic input noisy and gate model; PCA, proportion of correctly classified attribute; s = 1, GMS-DINA model with strategy selection parameter 1; s = 2, GMS-DINA model with strategy selection parameter 2; s = 10, GMS-DINA model with strategy selection parameter 10.

Table 9. PCV of DINA and GMS-DINA models under model misspecifications

	Fitted model			
Generating model	DINA	$s = 1$	$s = 2$	$s = 10$
DINA	0.761	0.754	0.760	0.758
$s = 1$	0.361	0.716	0.708	0.706
$s = 2$	0.366	0.771	0.780	0.778
$s = 10$	0.373	0.799	0.812	0.814

Note. DINA, deterministic input noisy and gate model; PCA, proportion of correctly classified attribute; s = 1, GMS-DINA model with strategy selection parameter 1; s = 2, GMS-DINA model with strategy selection parameter 2; s = 10, GMS-DINA model with strategy selection parameter 10.

used to generate 500 data sets, which were then fitted by the ACDM and GMS-ACDMs. Note that to fit the single-strategy DINA model and ACDM, or to generate data using them, the Q-matrix for the first strategy was used.

4.2.2. Results

Tables 8 and 9 present PCA and PCV related to the DINA and GMS-DINA models. Several findings can be observed. First, fitting the data using the generating model (i.e., the true model) always produced the most accurate attribute classifications, and the PCA and PCV

Table 10. PCA of ACDM and GMS-ACDM under model misspecifications

	Fitted model			
Generating model	ACDM	s = 1	s = 2	s = 10
ACDM	0.943	0.938	0.940	0.940
s = 1	0.770	0.885	0.882	0.879
s = 2	0.768	0.890	0.892	0.890
s = 10	0.772	0.894	0.901	0.902

Note. ACDM, additive cognitive diagnosis model; PCA, proportion of correctly classified attribute; s = 1, GMS-DINA model with strategy selection parameter 1; s = 2, GMS-DINA model with strategy selection parameter 2; s = 10, GMS-DINA model with strategy selection parameter 10.

Table 11. PCV of ACDM and GMS-ACDM under model misspecifications

	Fitted model			
Generating model	ACDM	s = 1	s = 2	s = 10
ACDM	0.746	0.729	0.737	0.738
s = 1	0.257	0.571	0.564	0.557
s = 2	0.248	0.591	0.604	0.598
s = 10	0.247	0.601	0.627	0.634

Note. ACDM, additive cognitive diagnosis model; PCA, proportion of correctly classified attribute; s = 1, GMS-DINA model with strategy selection parameter 1; s = 2, GMS-DINA model with strategy selection parameter 2; s = 10, GMS-DINA model with strategy selection parameter 10.

ranged from 0.921 to 0.950, and from 0.716 to 0.814, respectively. The classification rates when fitting the true model can serve as a benchmark for gauging the impact of model misspecifications. Second, when fitting the DINA model to data generated using the GMS-DINA model, the attribute classification rates worsened substantially (i.e., >10% decrease in PCA and >35% decrease in PCV compared with the benchmark). In contrast, when fitting the GMS-DINA model to the data generated using the DINA model, the attribute classification rates were very close to the benchmark (i.e., <1% decrease in both PCA and PCV). Additionally, it seems that when the Q-matrix was correctly specified, the misspecification in strategy selection parameter s had negligible impact on attribute classifications (i.e., <1% decrease in PCA and <2% decrease in PCV compared with the benchmark). The same trends can be observed for ACDM and GMS-ACDM models as shown in Tables 10 and 11.

5. Real data illustration

The data used in this section consist of the responses of 536 students to 15 fraction subtraction items, which have been previously analysed by de la Torre and Douglas (2008) and Mislevy (1996). The test measures seven attributes: α_1, basic fraction subtraction; α_2, simplify/reduce; α_3, separate whole number from fraction; α_4, borrow one from whole number to fraction; α_5, convert whole number to fraction; α_6, convert mixed number to fraction; α_7, column borrow in subtraction. The Q-matrix given in Table 12 is obtained from de la Torre and Douglas (2008), which is based on the work by Mislevy (1996). As discussed by Mislevy (1996), students can use two strategies to solve these items. Students who adopted Strategy A, which involves attributes α_1, α_2, α_3, α_4 and α_5, separated mixed

Table 12. Q-matrix for fraction subtraction items

	Strategy A					Strategy B				
Item	α_1	α_2	α_3	α_4	α_5	α_1	α_2	α_5	α_6	α_7
1	1	0	0	0	0	1	0	0	0	0
2	1	1	1	1	0	1	0	0	1	0
3	1	0	0	0	0	1	0	0	0	0
4	1	1	1	1	1	1	0	1	1	0
5	0	0	1	0	0	0	1	1	1	1
6	1	1	1	1	0	1	1	0	1	0
7	1	1	1	1	0	1	1	0	1	0
8	1	1	0	0	0	1	1	0	0	0
9	1	0	1	0	0	1	0	0	1	0
10	1	0	1	1	1	1	1	1	0	0
11	1	0	1	0	0	1	1	0	1	0
12	1	0	1	1	0	1	1	0	1	0
13	1	1	1	1	0	1	1	0	1	1
14	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	0	1	1	0	1	1

Table 13. AIC and BIC statistics of GMS-CDMs

	AIC			BIC		
Model	$s = 1$	$s = 2$	$s = 10$	$s = 1$	$s = 2$	$s = 10$
GMS-DINA	7,121	7,014	7,023	7,845	7,738	7,747
GMS-DINO	7,239	7,242	7,248	7,963	7,966	7,972
GMS-ACDM	6,967	6,956	6,886	7,858	7,847	7,777
GMS-LLM	6,829	6,846	6,841	7,720	7,737	7,732
GMS-RRUM	6,862	6,841	6,833	7,753	7,732	7,724

numbers into the whole number and fractional part and then performed subtraction separately. If necessary, a one is borrowed from the whole number. However, students who used Strategy B, which involves attributes α_1, α_2, α_5, α_6 and α_7, converted mixed numbers to improper fractions before performing subtraction.

The GMS-DINA, GMS-DINO, GMS-ACDM, GMS-LLM and GMS-RRUM, with strategy selection parameter s = 1, 2 and 10, were estimated using the aforementioned EM algorithm. All δ parameters for GMS-DINA and GMS-DINO models were constrained to be non-negative, and all main effects of GMS-ACDM, GMS-LLM and GMS-RRUM were constrained to be non-negative. These constraints ensured the monotonicity of item success probabilities. Furthermore, the lower and upper bounds of item success probabilities were set to 0.0001 and 0.9999, respectively.

It is well known that parameter estimates of the latent class model using the method of maximum likelihood estimation (e.g., the EM algorithm) might be local maxima. The same issue arose in the context of CDMs, which can be viewed as restricted latent class models. In this study, 300 sets of distinct starting values were used when each model was fitted to the data. With the convergence criterion of 0.001 as the maximal change in negative two times log-likelihood and the maximum number of 3,000 iterations, the estimates with the largest marginalized log-likelihood are reported here, although they are still not necessarily the global solutions.

Table 14. Parameter estimates and standard errors (in parenthesis) of GMS-LLM

Item	δ_{j0}	δ_{j1}	δ_{j2}	δ_{j3}	δ_{j4}	δ_{j5}	δ_{j6}	δ_{j7}
1	−2.437 (0.412)	4.982 (2.112)						
2	−4.656 (1.719)	2.255 (1.709)	2.424 (1.844)	0.000 (0.907)	4.471 (1.755)		4.493 (2.302)	
3	0.131 (0.165)	3.768 (2.295)						
4	−9.210 (0.400)	5.291 (0.903)	0.000 (1.128)	4.064 (1.481)	2.911 (1.295)	5.343 (1.494)	1.217 (1.079)	
5	−8.056 (0.113)		2.620 (0.758)	0.000 (0.115)		0 (1.272)	2.131 (1.083)	14.818 (2.744)
6	−8.005 (0.645)	3.508 (1.197)	2.675 (1.203)	0.552 (1.292)	4.738 (0.926)		2.591 (1.580)	
7	−9.210 (0.107)	3.321 (0.700)	2.986 (0.857)	3.924 (0.831)	7.054 (0.368)		4.655 (1.681)	
8	−1.965 (0.891)	2.991 (0.981)	3.603 (3.802)					
9	−6.572 (1.543)	9.072 (1.469)		9.014 (1.201)			1.464 (1.699)	
10	−9.210 (0.509)	3.987 (1.072)	5.231 (1.230)	4.967 (1.604)	0.000 (1.088)	7.408 (1.660)		
11	−3.467 (1.04)	4.838 (1.08)	4.980 (1.607)	1.014 (1.762)			5.161 (1.866)	
12	−9.210 (0.159)	3.125 (1.288)	0.000 (1.375)	4.948 (1.131)	4.712 (1.129)		17.599 (4.466)	
13	−9.210 (0.711)	6.605(1.287)	1.229 (0.893)	6.147 (1.045)	3.129 (1.269)		7.725 (1.776)	0.129 (0.631)
14	−9.210 (0.022)	4.184 (0.689)	0.730 (1.255)	2.930 (1.102)	2.385 (0.948)	3.145 (0.888)	3.095 (1.417)	0.000 (0.535)
15	−9.21 (0.107)	2.069 (0.747)	2.040 (0.925)	2.925 (1.008)	5.544 (0.628)		6.601 (1.415)	0.000 (0.468)

Note. Standard errors were calculated using the non-parametric bootstrap method based on 1,000 bootstrap samples with the same convergence criteria as the original GMS-LLM estimation.

To compare different models, we calculated the Akaike information criterion (AIC) and Bayesian information criterion (BIC) statistics. As shown in Table 13, the GMS-LLM with $s = 1$ had the smallest AIC and BIC values (i.e., AIC = 6,829, BIC = 7,720). According to Huo and de la Torre (2014), the AIC and BIC statistics of the MS-DINA model are 7,172 and 7,844, respectively. This suggests that the GMS-LLM has a better relative model-data fit than the MS-DINA model.

The item parameter estimates of the GMS-LLM are given in Table 14. Take item 2 as an example. Based on the Q-matrix in Table 12, students who use Strategy A might need to master attributes α_1, α_2, α_3 and α_4, whereas students who use Strategy B might only need to know attributes α_1 and α_6. If students do not master any of these five required attributes, they have a probability of 0.009 (i.e., $\exp(-4.656)/[1+\exp(-4.656)]$) to answer item 2 correctly. If students master attributes α_1 and α_6, they have a probability of 0.083 to answer item 2 correctly if they adopted Strategy A, and a probability of 0.890 if they adopted Strategy B. Because $s = 1$, the chances of these students using Strategies A and B are 0.085 and 0.915, respectively. As a result, the probability of these students answering item 2 correctly is 0.821. In a similar manner, the probability of adopting each strategy for each item given the attribute profile can be calculated. In addition, it is possible to calculate the prevalence of strategy m for item j as in $P_j(m) = \sum_c P_j(m|\boldsymbol{\alpha}_c)\pi_c$. For instance, the prevalence of strategy A is 0.775 for item 2 but only 0.404 for item 11.

After calculating item success probabilities, it can be observed that students who did not master any attributes tended to have low success probabilities for all items (<0.13) except item 3 (i.e., 0.533), and that students who mastered all attributes tended to have high success probabilities for all items (>0.92) except item 6 (i.e., 0.866). Similar to de la Torre (2008) and Ma and de la Torre (2016), $P_j(\mathbf{1}) - P_j(\mathbf{0})$ can be defined as the discrimination index of item j for assessing the quality of the item. It can be found that all items had discrimination indices >0.840 with the only exception being item 3 with a discrimination index of 0.448. In addition, the item parameter estimates for some values of δ were equal to zero, showing that these attributes might not be required for solving the problem.

A close scrutiny of Table 14 reveals that some item parameter estimates are boundary or near-boundary solutions (e.g., $\hat{\delta}_0 = -9.210$ for item 4). Although not presented here, several mixing proportion parameters were also estimated to be zero or very close to zero. The boundary or near-boundary solutions are not uncommon in latent class models when the maximum likelihood method is adopted for parameter estimation. Their occurrence, which could be an indicator of local maximum solution or identification issues (Uebersax, 2000), tends to produce numerical difficulties in the estimation algorithm (Vermunt & Magidson, 2004) and the calculation of parameters' covariance matrix (Garre & Vermunt, 2006). In this study, we calculated the standard error of item parameter estimates using the bootstrap procedure (Efron & Tibshirani, 1994), as suggested by De Menezes (1999) for handling the problem of boundary or near-boundary estimates. However, we employed the non-parametric bootstrap method instead of the parametric procedure used by De Menezes (1999) because the former seems to be more robust when the model or Q-matrix is misspecified (Guo, Ma, & de la Torre, 2017). It can be observed that the boundary solutions tend to have smaller bootstrap standard errors, indicating smaller variation among bootstrap samples probably because the parameter estimates were truncated.

Another important issue in CDMs is whether all latent classes are distinguishable, which is usually referred to as the completeness of the Q-matrix. However, it is not only the property of the Q-matrix, but it is also related to the underlying models (Köhn & Chiu, 2017, 2018). de la Torre and Douglas (2008) and Huo and de la Torre (2014) discussed the possibility of identifying equivalent classes in the DINA and MS-DINA models using ideal response vectors, and Köhn and Chiu (2017) proposed to examine the completeness by assessing whether different latent classes correspond to distinct expected item response vectors in other CDMs. Based on the procedure of Köhn and Chiu (2017), all latent classes are distinguishable when the Q-matrix and GMS-LLM are correctly specified.

6. Summary and discussion

Despite the fact that a large number of CDMs have been developed in the psychometric literature, most are not capable of accommodating multiple strategies. Unfortunately, it is not uncommon for students to adopt multiple strategies when approaching a problem. When multiple strategies are used by students but ignored by the models, the validity of inferences can be of great concern. In this study, we attempt to address this issue by developing a generalized multiple-strategy CDM framework within which a variety of models can be defined. The proposed model has several distinguishing features. First, based on theories and empirical findings about how strategies are chosen in problem-solving processes, the proposed model accommodates probability-matching, over-matching and maximizing strategy selection approaches in a unified framework. Secondly, the model can accommodate the situation that students adopt distinct strategies for different items and it can allow the estimation of the probability of using each strategy for each item given the attribute profiles. Finally, the model can handle different condensation rules and it can allow the use of several widely used CDMs. Simulation studies have shown that parameters of the proposed GMS-CDMs can be recovered adequately using the EM algorithm under various conditions. In addition, the proposed GMS-CDMs are shown to be robust when they are fitted to the data generated using single strategy CDMs and when strategy selection parameters are misspecified.

Results from the real data analysis show that the GMS-LLM has a better fit than other GMS models and the MS-DINA model. We also found that the Q-matrix in the real data was complete based on the procedure of Köhn and Chiu (2017), indicating that all latent classes are distinguishable. Nevertheless, the evaluation of local identifiability of the GMS-LLM, by checking whether the information matrix is positive definite as in von Davier (2014) and Ma (2019), failed as a result of numerical problems caused by the boundary estimates. A possible approach to boundary solutions is to employ the posterior mode estimation, which has been discussed by DeCarlo (2011) in the CDM contexts. Although the identifiability conditions have been studied for some CDMs recently (Gu & Xu, 2018; Xu, 2017; Xu & Shang, 2018), determining the identifiability conditions for more complex models such as the proposed GMS-CDMs could be challenging. Therefore, this is deferred for future research.

Notwithstanding the flexibility and many important features, the proposed GMS-CDMs are not without limitations. As noted by Siegler *et al.* (1996), the speed of the strategy is another important factor that individuals consider when choosing strategies. However, the proposed GMS-CDMs ignore this factor. As a result, the proposed models might be particularly appropriate when all strategies are similarly efficient. Future studies can explore how to incorporate strategy efficiency into the model. In addition, the formulations of the proposed GMS-CDMs can be extended in several ways. For example, the guessing parameters under different strategies are assumed to be equal in the proposed models, and all strategies are assumed to follow the same condensation rule for an item, although distinct condensation rules can be employed for different items.

Moreover, for GMS-ACDM, GMS-LLM and GMS-RRUM, an attribute is assumed to have the same contribution under all strategies that it is involved in. Furthermore, the strategy selection parameter is assumed to be known rather than estimated, although the real data analysis illustrated that relative fit statistics can be used to compare models with different strategy selection parameters. Future research can consider extending the GMS-CDM by relaxing some of these assumptions. Last, several widely used CDMs were used to define the strategy-specific IRF in this study, but it would be possible to employ more general CDMs, such as the generalized DINA model (de la Torre, 2011), loglinear CDM (Henson, Templin, & Willse, 2009) and the general diagnostic model (von Davier, 2008a).

The focus of this study is the development of GMS-CDMs in response to practical needs, and when an appropriate psychometric tool is available, we can expect more applications along this line. However, a well-designed diagnostic assessment is equally, if not more, important as the psychometric models. Research questions, such as how to identify students' strategies for each item and how to correctly define the corresponding q-vectors, also need to be carefully answered to exploit the potential of the proposed models.

References

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459. https://doi.org/10.1007/BF02293801

Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, *74*, 633–665. https://doi.org/10.1007/s11336-009-9125-0

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, *35*, 8–26. https://doi.org/10.1177/0146621610377081

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, *45*, 343–362. https://doi.org/10.1111/j.1745-3984.2008.00069.x

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*, 179–199. https://doi.org/10.1007/s11336-011-9207-7

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333–353. https://doi.org/10.1007/bf02295640

de la Torre, J., & Douglas, J. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, *73*, 595–624. https://doi.org/10.1007/s11336-008-9063-2

De Menezes, L. (1999). On fitting latent class models for binary data: The estimation of standard errors. *British Journal of Mathematical and Statistical Psychology*, *52*, 149–168. https://doi.org/10.1348/000711099159026

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (methodological)*, *39*, 1–38. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

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

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall.

Fazio, L. K., Dewolf, M., & Siegler, R. S. (2016). Strategy use and strategy choice in fraction magnitude comparison overlapping waves theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1–16. https://doi.org/10.1037/xlm0000153

Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, *109*, 416–422. https://doi.org/10.1016/j.cognition.2008.09.007

Garre, F. G., & Vermunt, J. K. (2006). Avoiding boundary estimates in latent class analysis by Bayesian posterior mode estimation. *Behaviormetrika*, *33*, 43–59. https://doi.org/10.2333/bhmk.33.43

Gu, Y., & Xu, G. (2018). The sufficient and necessary condition for the identifiability and estimability of the DINA model. *Psychometrika*, Advanced online publication. https://doi.org/10.1007/s11336-018-9619-8

Guo, W., Ma, W., & de la Torre, J. (2017). *Standard error estimation using bootstrap approaches for cognitive diagnosis models*. Paper presented at the Annual Meeting of the National Council of Measurement in Education, San Antonio, TX.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*, 301–321. https://doi.org/10.1111/j.1745-3984.1989.tb00336.x

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.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*, 191–210. https://doi.org/10.1007/s11336-008-9089-5

Huo, Y., & de la Torre, J. (2014). Estimating a cognitive diagnostic model for multiple strategies via the EM algorithm. *Applied Psychological Measurement*, *38*, 464–485. https://doi.org/10.1177/0146621614533986

Köhn, H.-F., & Chiu, C.-Y. (2016). A proof of the duality of the DINA model and the DINO model. *Journal of Classification*, *33*, 171–184. https://doi.org/10.1007/s00357-016-9202-x

Köhn, H.-F., & Chiu, C.-Y. (2017). A procedure for assessing the completeness of the Q-matrices of cognitively diagnostic tests. *Psychometrika*, *82*, 112–132. https://doi.org/10.1007/s11336-016-9536-7

Köhn, H.-F., & Chiu, C.-Y. (2018). How to build a complete Q-matrix for a cognitively diagnostic test. *Journal of Classification*, *35*, 273–299. https://doi.org/10.1007/s00357-018-9255-0

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335–1342. https://doi.org/10.1126/science.208.4450.1335

Liu, J., Xu, G., & Ying, Z. (2011). *Learning item-attribute relationship in Q-matrix based diagnostic classification models*. arXiv preprint arXiv:1106.0721. Retrieved from http://arxiv.org/pdf/1106.0721v1.pdf

Lovett, M. C., & Anderson, J. R. (1995). Making heads or tails out of selecting problem-solving strategies. In J. Moore & J. Lehman (Eds.), *The seventeenth annual conference of the cognitive science society* (pp. 265–270). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ma, W. (2019). A diagnostic tree model for polytomous responses with multiple strategies. *British Journal of Mathematical and Statistical Psychology*, *72*, 61–82. https://doi.org/10.1111/bmsp.12137

Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology*, *69*, 253–275. https://doi.org/10.1111/bmsp.12070

Ma, W., Iaconangelo, C., & de la Torre, J. (2016). Model similarity, model selection, and attribute classification. *Applied Psychological Measurement*, *40*, 200–217. https://doi.org/10.1177/0146621615621717

Ma, W., & de la Torre, J. (2019). *GDINA: The generalized DINA model framework*. [Computer software version 2.3.2]. Retrieved from https://CRAN.R-project.org/package=GDINA

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*, 187–212. https://doi.org/10.1007/BF02294535

Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, *33*, 379–416. https://doi.org/10.1111/j.1745-3984.1996.tb00498.x

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*, 195–215. https://doi.org/10.1007/bf02295283

Molenaar, D., & de Boeck, P. (2018). Response mixture modeling: Accounting for heterogeneity in item characteristics across response times. *Psychometrika*, *83*, 279–297. https://doi.org/10.1007/s11336-017-9602-9

Olshavsky, J. E. (1976). Reading as problem solving: An investigation of strategies. *Reading Research Quarterly*, *12*, 654. https://doi.org/10.2307/747446

Peters, G., Smedt, B., Torbeyns, J., Ghesquiere, P., & Verschaffel, L. (2013). Children's use of addition to solve two-digit subtraction problems. *British Journal of Psychology*, *104*, 495–511. https://doi.org/10.1111/bjop.12003

R Core Team (2018). *R: A language and environment for statistical computing* [Computer software]. Vienna, Austria: R Core Team. Retrieved from https://www.R-project.org/

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*, 271–282. https://doi.org/10.1177/014662169001400305

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

Siegler, R. S., Adolph, K. E., & Lemaire, P. (1996). Strategy choices across the life span. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 79–121). Mahwah, NJ: Erlbaum.

Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, *18*, 565–579. https://doi.org/10.1016/j.learninstruc.2007.09.018

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*, 345–354. https://doi.org/10.1111/j.1745-3984.1983.tb00212.x

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*, 287–305. https://doi.org/10.1037/1082-989x.11.3.287

Uebersax, J. (2000). *A brief study of local maximum solutions in latent class analysis*. Retrieved from http://www.john-uebersax.com/stat/local.htm

Vermunt, J. K., & Magidson, J. (2004). Latent class analysis. In M. S. Lewis-Beck, A. Bryman & T. F. Liao (Eds.), *The SAGE encyclopedia of social science research methods* (pp. 550–553). Thousand Oaks, CA: Sage

von Davier, M. (2007). Mixture distribution diagnostic models. *ETS Research Report Series*, *2007*, i–21. https://doi.org/10.1002/j.2333-8504.2007.tb02074.x

von Davier, M. (2008a). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*, 287–307. https://doi.org/10.1348/000711007x193957

von Davier, M. (2008b). The mixture general diagnostic model. In G. R. Hancock & K. M. Samuelson (Eds.), *Advances in latent variable mixture models* (pp. 164–175). Charlotte, NC: Information Age Publishing.

von Davier, M. (2010). Mixture distribution item response theory, latent class analysis, and diagnostic mixture models. In S. E. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (pp. 11–34). Washington, DC: American Psychological Association. https://doi.org/10.1037/12074-000

von Davier, M. (2014). The DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology*, *67*, 49–71. https://doi.org/10.1111/bmsp.12003

Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Computational Biology*, *6*, 1–7. https://doi.org/10.1371/journal.pcbi.1000871

Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *The Annals of Statistics*, *45*, 675–707. https://doi.org/10.1214/16-AOS1464

Xu, G., & Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association*, *113*, 1284–1295. https://doi.org/10.1080/01621459.2017.1340889

Xu, X., & von Davier, M. (2008). Fitting the structured general diagnostic model to NAEP data. *ETS Research Report Series*, *2008*, 1–18. https://doi.org/10.1002/j.2333-8504.2008.tb02113.x

Received 21 April 2018; revised version received 4 November 2018