Journal of Educational and Behavioral Statistics

A Multi-strategy Cognitive Diagnosis Model Incorporating Item Response Times Based on Strategy Selection Theories

Journal:	*JEBStats*
Manuscript ID	JEBS-2021-MAN-0178
Manuscript Type:	Manuscript
Abstract:	Cognitive Diagnostic Models (CDMs) have been used as a psychometric tool in educational and psychological evaluation to assess the proficiency of students. However, most CDMs assume that all students use the same one strategy when dealing with problems in the assessment, which ignores the diversity and differences in processing strategies in practice. This study develops a multi-strategy cognitive diagnosis model that combines response times based on strategy selection theories, which integrates individual response accuracy and response time into the same framework to define strategy selection, making it closer to the individual's strategy selection process. Simulation studies showed that the proposed model had reasonable item recovery and classification accuracy for attributes and outperformed the existing multi-strategy CDMs. A set of real data was analyzed as well to illustrate the application and the advantages of the proposed model in practice.

A Multi-strategy Cognitive Diagnosis Model Incorporating Item Response Times Based on Strategy Selection Theories

Abstract: Cognitive Diagnostic Models (CDMs) have been used as a psychometric tool in educational and psychological evaluation to assess the proficiency of students. However, most CDMs assume that all students use the same one strategy when dealing with problems in the assessment, which ignores the diversity and differences in processing strategies in practice. This study develops a multi-strategy cognitive diagnosis model that combines response times based on strategy selection theories, which integrates individual response accuracy and response time into the same framework to define strategy selection, making it closer to the individual's strategy selection process. Simulation studies showed that the proposed model had reasonable item recovery and classification accuracy for attributes and outperformed the existing multi-strategy CDMs. A set of real data was analyzed as well to illustrate the application and the advantages of the proposed model in practice.

Key Words: Cognitive diagnosis model; Multi-strategy; Response time; Strategy selection

Introduction

Cognitive diagnosis models (CDMs) have been used as psychometric tools in educational assessments to estimate students' proficiency profiles. Currently large number of CDMs have been developed by psychometric scholars (Fu & Li, 2008), such as the widely used the Rule Space Model (RSM; Tatsuoka, 1995), the Attribute Hierarchy Model (AHM; Leighton, Gierl, & Hunka, 2004), the deterministic inputs, noise "and" gate model (DINA; Junker & Sijtsma, 2001), the Generalized DINA Model

(G-DINA; de la Torre, 2011), etc. However these models basically assume that all students use the same strategy to solve problems, which ignores the diversity and differences in processing strategies.

Psychological studies have shown that in solving multi-digit multiplication (Siegler & Lemaire, 1997), spelling (Rittle-Johnson & Siegler, 1999), graphical reasoning (DeShon, Chan, & Weissbein, 1995), and other types of cognitive tasks, individual may use multiple problem-solving strategies. Ignoring multiple strategies and using a single-strategy model to fit the multi-strategy test scenario will cause the model to be incorrectly set or insufficiently fitted, which will affect the accuracy and effectiveness of inferences. So far, only few CDMs have been proposed to accommodate multiple strategies. Mislevy and Verhest (1990) proposed a mixed linear logistic test model (LLTM), which allowed different individuals to use different problem-solving strategies, but limited individuals to only use one same strategy to solve all items. In addition, de la Torre and Douglas (2008) developed the multiple-strategy deterministic inputs, noisy "and" gate (MS-DINA) model and assumed 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. Moreover the MS-DINA model also 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 in practice. Therefore, Ma and Guo (2019) proposed a generalized multiple-strategy model (GMS-CDM) for dichotomous response data, which extends the MS-DINA model to provide a more flexible framework for modeling the response and takes various strategy selection approaches into consideration. However, when modeling strategy selection, the above-mentioned multi-strategy cognitive diagnosis models all have certain limitations and deficiencies in which only the influence of the accuracy of the strategy is considered to strategy selection.

Plenty of literatures have addressed how individuals select strategies in the process of problem solving. Rieskamp and Otto (2006) put forward the strategy selection learning theory (SSL theory), which holds that students will adjust their strategies according to the accuracy of strategies to solve practical problems. However, Marewski and Schooler (2011) believed that the SSL theory has limitations and proposed the cognitive niche theory, which believe that students will weigh the speed, accuracy and effort of strategy execution to estimate the usability of the strategy. In addition, adaptive control of thought-rational model (ACT-R Model; Lovett & Anderson, 1996) and adaptive strategy choice model (AS-CM; Siegler et al., 1997) believe that individuals will weigh the speed and accuracy of the strategy to choose the best strategy. It can be clearly seen from literature that most of the existing strategy choice theories believe that individuals will weigh both the accuracy and the speed of strategies to choose the optimal strategy in the process of problem-solving.

We believe the response time (RT) on problem-solving is a very important information of individual process speed and will be very helpful in inferring which problem-solving strategy is used by individual due to that different strategy may present different distributions in response time. For example, in the thought rotation test, some individuals may use the thought rotation strategy, while others may use the analytical rule strategy. Although both two strategies may give the correct response, they have different distributions in response time (Mislevy, Wingersky, Irvine, & Dann, 1991). Siegler (1989) also suggested that response time analyses may be very useful in examining the strategies that examinees use to solve a problem. Hence, the inclusion of response time can be used to analyze the response strategies adopted by individuals and make relevant inferences about the thinking process.

Therefore, our purpose of the current study is to propose a multi-strategy cognitive diagnosis model which incorporates response time in the strategy selection modeling to improve the existing models which only consider the influence of the accuracy in strategy selection. Specifically, the proposed model integrates both information of the accuracy and speed (response time) into the same framework to define strategy selection, rather than being limited to only the impact of accuracy on strategy selection in previous studies, which is more in line with the existing strategic choice theory. Our proposed model is expected to improve the existing CDMs in terms of both the classification accuracy and the diagnosis of strategy selection.

Background

Q-matrix (Tatsuoka, 1983) is an important component of CDMs, which associates items with attributes. The MS-DINA model (de la Torre et al., 2008) assumes that there are M Q-matrices corresponding to M strategies. Let $q_j^m = \{q_{jk}^m\}$ denote the q-vector of item j for strategy m. The IRF of the MS-DINA model can be formulated as,

$$P(x_{ij}=1|\alpha_c) = \delta_{j0} + \delta_{j1}\left[1 - \prod_{m=1}^{M}\left(1 - \prod_{k=1}^{K}\alpha_{ck}^{q_{jk}^m}\right)\right], \tag{1}$$

where q_{jk}^m indicates whether the k-th cognitive attribute needs to be mastered when answering item j using strategy m; α_{ck} indicates whether individual with attribute profile α_c has mastered attribute k. δ_{j0} is the guessing parameter, and δ_{j1} indicates the increase in the success probability when any of the required attributes are mastered, which assumes that individuals who master all attributes required by at least one strategy have a success probability of $\delta_{j0} + \delta_{j1}$. On the contrary, if one does not master all the attributes of at least one strategy, the probability of correct answer is δ_{j0}. So there are only two parameters for each item.

As for GMS-CDM, the IRF is formulated as,

$$P(x_{ij}=1|\alpha_c) = \sum_{m=1}^{M_j} P\,(x_{ij}=1|\alpha_c,m)\,P_j\,(m|\alpha_c). \tag{2}$$

Here $P(x_{ij}=1|\alpha_c,m)$ is the probability of individual i given attribute profile α_c answering item j correctly using strategy m, which is referred to as strategy-specific IRF. Specifically, the strategy-specific IRF of the GMS-DINA model is parameterized as,

$$P(x_{ij}=1|\alpha_c,m) = \delta_{j0} + \delta_{jm1}\prod_{k=1}^{K}\alpha_{ck}^{q_{jk}^m}. \tag{3}$$

Suppose there are M strategies in solving the problem, each involving a unique subset of the K attributes. Let q_{jk}^m denote the q-vector of item j for strategy m. δ_{j0} is the guessing parameter, representing the success probability for individuals who do not master all required attributes, and δ_{jm1} is the increase in success probability when all required attributes of strategy *m* are mastered. Different strategies are allowed to have different accuracy.

$$P_j(m|\alpha_c) = \frac{P(x_{ij}=1|\alpha_c,m)^s}{\sum_{m=1}^{M_j} P(x_{ij}=1|\alpha_c,m)^s}, \tag{4}$$

where $P_j(m|\alpha_c)$ is the probability of an individual with attribute profile α_c choosing strategy *m* for item *j*. *s* is called the strategy selection parameter, which can be set in advance. Therefore, in this modeling, the strategy selection probability is only affected by success probability using a specific strategy.

The Proposed Multi-strategy CDM Incorporating Item Response Times (MS-CDM-RT)

Model Specification

In the proposed model, the correct probability of an individual to solve item *j* is the product sum of the correct probability of using various strategies and their weights, which is defined as follows:

$$P(x_{ij}=1|\alpha_i) = \sum_{m=1}^{M_j} P(x_{ij}=1|\alpha_i,m) w_{ijm}, \tag{5}$$

where ω_{ijm} is the probability of individual *i* choosing strategy *m* for item *j*, which can be used to define strategy selection probability, as a weight coefficient. $P(x_{ij}=1|\alpha_i,m)$ is the probability of individual *i* given attribute profile α_i answering item *j* correctly using strategy *m*, whose expression can be seen in Formula (3) for DINA model (other CDMs can of course be chosen, such as the G-DINA model).

According to theories of strategy selection, the accuracy of response and response time of each strategy will be both taken into consideration when the individuals make strategy selection. We further define it as:

$$\omega_{ijm} = \frac{\pi_m P(x_{ij}=1|\alpha_i, m) f(t_{ij}|m)}{\sum_{m=1}^{M} \pi_m P(x_{ij}=1|\alpha_i, m) f(t_{ij}|m)}, \tag{6}$$

where t_{ij} is the observed response time of person *i* on item *j*, and π_m is the proportion of individuals using strategy *m*, which can be called a mixed ratio. Let $0 < \pi_m < 1$, and $\sum_{m=1}^{M} \pi_m = 1$. For each strategy, the accuracy of response is defined by the part of $P(x_{ij}=1|\alpha_c, m)$, and the response time or speed is defined by the part of $f(t_{ij}|m)$ which can been chosen as the Lognormal Model for Response Times (LMFRT; van der Linden, 2006) as,

$$f(t_{ij}|m) = \frac{\gamma_{jm}}{t_{ij}\sqrt{2\pi}} \exp\left\{-\frac{1}{2}[\gamma_{jm}(\ln t_{ij} - (\beta_{jm} - \tau_{im}))]^2\right\}, \tag{7}$$

where $f(t_{ij}|m)$ is the normal distribution probability density function of the logarithm of the response time of individual *i* on item *j* using strategy *m*. τ_{im} is the speed parameter of individual *i* using strategy *m*. The larger the value, the shorter the time it takes an individual to complete the problem. γ_{jm} indicates the degree of discrimination of the item *j* on the speed level of the subjects using strategy *m*. The higher the value, represents the stronger the ability of the item to distinguish between individuals at different speed levels. β_{jm} reflects the time intensity of the item *j* for subjects using strategy *m*, that is, the time required to complete item *j*.

In the expression of the strategy selection probability ω_{ijm}, its numerator refers to the posterior probability of the individual choosing strategy *m* when weighing the accuracy of the strategy and the response time or speed. Assuming there are *M* strategies in total, the denominator represents the cumulative sum of the *M* posterior probabilities. When $P(x_{ij}=1|\alpha_i, m)$ is larger, that is, the probability that individual *i* uses strategy *m* to answer item *j* correctly, ω_{ijm} will be greater, indicating that individual i has a greater probability of choosing strategy m when answering item j. In the formula $f(t_{ij} \mid m)$, when the time-related parameter value ($\beta_{jm} - \tau_{im}$) under strategy m is closer to the true value, ω_{ijm} will be larger, and the probability of strategy m being selected is also greater.

It should be noted that we use the DINA model as an example to illustrate the modeling framework of the proposed model (we called the proposed model as MS-CDM-RT model), which does not lose the generality and can be easily extended to other models (such as the G-DINA model).

Parameter Estimation

In this study, the Markov Chain Monte Carlo (MCMC) is used to estimate the parameters of the proposed MS-CDM-RT model. The characteristic of the MCMC method is that even if the complexity of the model increases, its calculation process is still easy to implement, and the model parameters can be estimated at the same time. The MCMC method constructs a combination of the observation data and the likelihood function of the prior distribution of the model parameters, and then obtains the posterior distribution of the parameters and samples them to infer the model parameters.

If the attribute mastery mode α_i is given and individuals' responses are independent, then the model likelihood function can be given as:

$$L(\mathbf{X}|\alpha_i, \tau_{im}, \gamma_{jm}, \beta_{jm}, \delta_{j0}, \delta_{jm1}, \pi_m) = \prod_{i=1}^{N} P(x_i|\alpha_i). \tag{8}$$

The prior distribution of each parameter in the model can be chosen as:

$$\alpha_i \sim Bernoulli(Uniform(0,1)), \tag{9}$$

$$\tau_{im} \sim N(0,1), \tag{10}$$

$$\gamma_{jm} \sim Uniform(0.8, 2.5), \tag{11}$$

$$\beta_{jm} \sim N(0,1), \tag{12}$$

$$\delta_{j0} \sim Uniform(0.05, 0.2) \quad , \tag{13}$$

$$\delta_{jm1} = P(x_{ij} = 1 | \alpha_i, m) - \delta_{j0}, \quad (14)$$

and

$$\pi_m \sim Beta(1, 1). \quad (15)$$

Then, the joint posterior distribution of the proposed model can be expressed as,

$$p(\alpha_i, \tau_{im}, \gamma_{jm}, \beta_{jm}, \delta_{j0}, \delta_{jm1}, \pi_m | x_{ij}, t_{ij}) \propto \\ L(\alpha_i, \tau_{im}, \gamma_{jm}, \beta_{jm}, \delta_{j0}, \delta_{jm1}, \pi_m) p(\alpha_i) p(\tau_{im}) p(\gamma_{jm}) p(\beta_{jm}) p(\delta_{j0}) p(\delta_{jm1}) p(\pi_m). \quad (16)$$

The parameter estimation adopts the random moving M-H algorithm under Gibbs sampling, and the estimated value of the posterior distribution of each model parameter is used as the estimated value of the parameter.

Simulation Studies

Two simulation studies were conducted to evaluate the performance of the proposed MS-CDM-RT. The first simulation study (Study **I**) mainly explored the parameter recovery of the proposed model under different experimental conditions, and the second simulation study (Study **II**) mainly compared the proposed model with the existing multi-strategy CDMs (i.e., the MS-DINA and GMS-DINA).

Simulation Study I

Design

For the first simulation study, the sample size is set as N = 500, 1000 and 2000 respectively. The test length was fixed at J = 30 and the number of attributes was set K = 5. There were two strategies for each item. The Q-matrix for each strategy quoted the Q matrix of Ma and Guo (2019), given in Table 1. It was created to ensure that each attribute was measured the same number of times under two strategies and that the Q-matrix was complete (Chiu, Douglas, & Li, 2009) for each strategy. Individuals' attribute profiles were taken from either a uniform distribution, where all possible attribute profiles had 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 given by,

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

Here $\theta \sim N(0,1)$ was the higher-order ability, and $b_k = -1 + 0.5(k-1)$ is the difficulty of attribute k for $k = 1,\ldots,K$.

Specifically, the parameters related to response accuracy were taken from the following distribution, referring to the setting of Ma et al (2019): $\delta_{j0} \sim U[0.05, 0.2]$, $\delta_{jm1} = P(x_{ij} = 1 | \alpha_i = 1, m) - \delta_{j0}$, and the parameters related to the response time were taken from the following distribution: $\tau_{im} \sim N(0,1), \gamma_{jm} \sim Uniform(0.8, 2.5)$, $\beta_{jm} \sim N(0,1)$.

Table 1 Q-matrix for simulation studies (Ma et al., 2019)

Item	Strategy A					Strategy B				
	α1	α2	α3	α4	α5	α1	α2	α3	α4	α5
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

Criteria

To evaluate item parameter recovery of the MS-CDM-RT, the average bias and absolute error (ABSE) of item parameter estimate were used and respectively calculated by

$$bias = \sum_{r=1}^{R} \frac{\hat{\delta}_r - \delta}{R}, \tag{18}$$

What is R? Number of replicates?

delta here is a vector?

and

$$ABSE = \sum_{r=1}^{R} \frac{\left|\hat{\delta}_r - \delta\right|}{R}. \tag{19}$$

Person parameter recovery was evaluated using the pattern correct classification rate (PCCR) and the attribute correct classification rate (ACCR), which were respectively defined as,

$$PCCR = \frac{\sum_{r=1}^{R}\sum_{i=1}^{N}(I(\hat{\alpha}_i, \alpha_i))}{R \times N}, \tag{20}$$

and

I(alpha_hat = alpha)

$$ACCR = \frac{\sum_{r=1}^{R}\sum_{i=1}^{N}(I(\hat{\alpha}_{ik}, \alpha_{ik}))}{R \times N}. \tag{21}$$

Here, R is the number of repeated experiments, and N is the number of individuals.

scalar or vector?

$\hat{\delta}_r$ and δ are the estimated value the true value of the parameter. $\hat{\alpha}_i$ and α_i are the estimated value and the true value of attribute vector matches. I (.) is an indicator function. When the estimated attribute vector matches correctly, that is, I=1, otherwise I=0.

Results of Study I

Table 2 shows the biases of the item parameter estimation and Table 3 shows the ABSEs under various conditions. The MS-CDM-RT model with two strategies has six parameters for each item, including δ_{j0}, δ_{j11}, δ_{j21}, τ_i, β_j and γ_j. Several findings can be observed from Table 2 and Table 3. First, the biases were less than 0.01 and the

ABSEs were less than 0.05 of parameters related to the accuracy of the response (i.e., δ_{j0} and δ_{j1}) under almost all conditions. Compared with δ_{j0} and δ_{j1}, the parameters related to time (i.e., τ_i, β_j and γ_j) had slightly larger biases and ABSEs, but the values were also in the acceptable range. Second, the parameters δ_{j0} and δ_{j1} generally had smaller biases and ABSEs when attributes were taken from higher-order distribution. For other parameters about time (i.e., τ_i, β_j and γ_j), however, smaller biases and ABSEs basically occurred when attributes were uniformly distributed. Third, as expected, parameter estimation was basically consistent with the decrease of ABSEs as sample size (N) increased.

Table 2 Biases of item parameters of MS-CDM-RT model

Attribute Structure	N	δ_{j0}	δ_{j1}	τ_i	β_j	γ_j
Uniform	500	-0.0025	0.0060	-0.0281	-0.1021	-0.0361
	1000	0.0037	0.0009	-0.0015	-0.0176	-0.0482
	2000	-0.0023	0.0085	0.0163	-0.0420	0.0106
Higher-order	500	0.0020	-0.0009	-0.0013	-0.0114	0.0828
	1000	-0.0016	0.0027	-0.0185	0.0913	-0.0650
	2000	-0.0002	-0.0047	0.0053	0.0495	-0.0311

Table 3 ABSEs of item parameters of MS-CDM-RT model

Attribute Structure	N	δ_{j0}	δ_{j1}	τ_i	β_j	γ_j
Uniform	500	0.0154	0.0340	0.6540	0.2706	0.2580
	1000	0.0115	0.0233	0.6316	0.3361	0.2431
	2000	0.0076	0.0241	0.6337	0.1786	0.1493
Higher-order	500	0.0153	0.0317	0.6583	0.5228	0.3423
	1000	0.0093	0.0246	0.6813	0.4324	0.3388
	2000	0.0119	0.0234	0.6626	0.4568	0.3514

Tables 4 and 5 showed the ACCR and PCCR of MS-CDM-RT model under various conditions. The ACCR of MS-CDM-RT model was greater than 0.9 under almost all conditions. With one exception, ACCR was 0.883 when N=500 and uniform distribution. In addition, the results of ACCR and PCCR of MS-CDM-RT model both showed that compared with uniform distribution, individuals tended to be better classified under higher-order attribute distribution. This was consistent with the result of better parameter estimation accuracy under high-order attribute distribution.

Table 4 Attribute correct classification rate (ACCR) of MS-CDM-RT model

Attribute Structure	N	A1	A2	A3	A4	A5	Mean
Uniform	500	0.883	0.954	0.935	0.931	0.923	0.925
	1000	0.928	0.934	0.949	0.942	0.939	0.938
	2000	0.912	0.931	0.942	0.934	0.947	0.933
Higher-order	500	0.947	0.968	0.986	0.998	0.990	0.978
	1000	0.946	0.951	0.981	0.994	0.985	0.971
	2000	0.947	0.912	0.983	0.987	0.987	0.963

Table 5 Pattern correct classification rate (PCCR) of MS-CDM-RT model

N	Uniform	Higher-order
500	0.730	0.899
1000	0.785	0.878
2000	0.763	0.845

Simulation Study II

The second simulation study compares the newly proposed MS-CDM-RT model with the traditional MS-DINA and GMS-DINA models through Monte Carlo simulation. The goal of the second simulation study is two-fold: (1) to check whether the GMS-DINA model and MS-DINA model fit the data generated by MS-CDM-RT model can accurately classify individuals, and (2) to compare the classification accuracy of the proposed model with the conventional models (i.e., MS-DINA and GMS-DINA).

Design

The Q matrix and item parameter generation were ~~the~~ similar as the first simulation study. Sample size were fixed at $N = 500$ and 1000 and individuals' attribute profiles came from a uniform distribution or a higher-order distribution like the first study. Under the above four experimental conditions, MS-CDM-RT was used to simulate data sets, which were then fitted by the MS-DINA, GMS-DINA and MS-CDM-RT models, respectively.

Results of Study II

Table 6 and table 7 presented ACCR and PCCR related to MS-DINA, GMS-DINA and MS-CDM-RT models. Several findings can be observed. First, under various conditions, using the generating model to fit the data will always produce the most accurate attribute classification, that is, the ACCRs of the MS-CDM-RT model were always higher than those of the MSDINA and GMSDINA models. Among them, MS-DINA model had the worst attribute classification accuracy. Second, the PCCR of the proposed MS-CDM-RT model was much better than the GMS-DINA and MS-DINA models. Additionally, no matter what kind of model fit the data, the PCCR of the model whose attributes came from a higher-order distribution is higher than that of the model whose attributes come from a uniform distribution. Third, it was known from the PCCR results that the accuracy of the GMS-DINA model and MS-DINA model under uniform distribution was relatively low. The PCCR values of the GMS-DINA model were all less than 0.6, and the PCCR values of the MS-DINA model were all less than 0.5.

Table 6 ACCR of MS-CDM-RT, GMS-DINA and MS-DINA models

Attribute Structure	N	Model	A1	A2	A3	A4	A5	Mean
Uniform	500	MS-DINA	0.725	0.862	0.843	0.847	0.823	0.820
		GMS-DINA	0.780	0.906	0.891	0.891	0.861	0.866
		MS-CDM-RT	0.883	0.954	0.935	0.931	0.923	0.925
	1000	MS-DINA	0.748	0.806	0.842	0.833	0.815	0.809
		GMS-DINA	0.806	0.851	0.881	0.881	0.863	0.856
		MS-CDM-RT	0.928	0.934	0.949	0.942	0.939	0.938
Higher-order	500	MS-DINA	0.846	0.935	0.941	0.970	0.985	0.935
		GMS-DINA	0.907	0.947	0.965	0.976	0.992	0.957
		MS-CDM-RT	0.947	0.968	0.986	0.998	0.990	0.978
	1000	MS-DINA	0.821	0.836	0.935	0.985	0.978	0.911
		GMS-DINA	0.898	0.863	0.957	0.986	0.988	0.938
		MS-CDM-RT	0.946	0.951	0.981	0.994	0.985	0.971

Table 7 PCCR of MS-CDM-RT, GMS-DINA and MS-DINA models

Attribute Structure	N	MS-DINA	GMS-DINA	MS-CDM-RT
Uniform	500	0.429	0.557	0.730
	1000	0.381	0.510	0.785
Higher-order	500	0.718	0.814	0.899
	1000	0.625	0.735	0.878

Table 8 and Table 9 show the biases and ABSEs of the item parameters shared by MS-CDM-RT and GMS-DINA (i.e., δ_{j0} and δ_{j1}). Compared with the GMS-DINA model, MS-CDM-RT always had smaller biases and ABSEs under various conditions.

Specifically, the biases of item parameters of MS-CDM-RT model were less than 0.01, and the ABSEs were less than 0.05. When attributes were from a high-order distribution, biases and ABSEs will be smaller than the uniform distribution regardless of the model used. The same result can also be obtained from table2 and table3.

Table 8 Biases of item parameters of MS-CDM-RT and GMS-DINA

Attribute Structure	N	Model	δ_{j0}	δ_{j1}
Uniform	500	GMS-DINA	0.0255	-0.0396
		MS-CDM-RT	-0.0025	0.0060
	1000	GMS-DINA	0.0372	-0.0592
		MS-CDM-RT	0.0037	0.0009
Higher-order	500	GMS-DINA	0.0159	-0.0158
		MS-CDM-RT	0.0020	-0.0009
	1000	GMS-DINA	0.0177	-0.0195
		MS-CDM-RT	-0.0016	0.0027

Table 9 ABSEs of item parameters of MS-CDM-RT and GMS-DINA

Attribute Structure	N	Model	δ_{j0}	δ_{j1}
Uniform	500	GMS-DINA	0.0284	0.0570
		MS-CDM-RT	0.0154	0.0340
	1000	GMS-DINA	0.0384	0.0705
		MS-CDM-RT	0.0115	0.0233
Higher-order	500	GMS-DINA	0.0240	0.0517
		MS-CDM-RT	0.0153	0.0317
	1000	GMS-DINA	0.0223	0.0460
		MS-CDM-RT	0.0093	0.0246

Real Data Illustration

The real data includes 904 students' response data and the corresponding response time at Raven's Advanced Progressive Matrices Test (APM). This study designed the test into a computer format to record participants' response times. Many studies (Deshon, Chan, & Weissbein, 1995; Carpenter, Just, & Shell, 1990) had pointed out that individuals may use both Verbal Analytic and Visual-spatial strategies to answer APM. According to the previous studies (Deshon, Chan, & Weissbein, 1995; Carpenter, Just, & Shell, 1990), the strategy of Verbal Analytic involves attributes of pairwise progression (PP), Constant in a row (C), Distribution of two values (D2) and Distribution of three values(D3). The strategy of Visual-spatial involves attributes of Superimposition with Cancellation rule (SC), Superimposition rule (S), Movement rule (M), Figure addition or subtraction (A/S), Mental Transformation (MT) and Rotatio rule (R).

Among the items of the APM, item 15, 19, 20 and 25 cannot be coded according to the existing rules and item 18 only measures the Totatio Rotatio rule (Deshon et al., 1995). Therefore, the analysis of the APM in this study does not include the above five items. In the process of defining the Q matrix, some attributes are integrated to increase the number of measurements. Since the Superimposition with Cancellation rule is actually a special case of the Superimposition rule, and only the item 11 measures the Figure addition or subtraction, the Superimposition and Superimposition with Cancellation rule are combined with the Figure addition or subtraction, which is called A/S.

Table 10 presents the multi-strategy Q matrix used in the APM Test. QA is the Q matrix of individuals who tend to use the Verbal Analytic strategy, and QV is the Q matrix of individuals who tend to use the Visual spatial strategy. Among them, the responses to item 24, 26, 31 and 35 need to use two strategies at the same time and item 2, 3, 5, 6, 10, 14, 22, 23 and 32 can be solved through Verbal Analytic strategy or Visual spatial strategy, other items can only be solved by Verbal Analytic strategy or Visual spatial strategy.

Table 10 Multi-strategy Q matrix for APM items

Item	Strategy of Verbal Analytic (QA)						Strategy of Visual-spatial (QV)					
	C	PP	D3	D2	M	A/S	C	PP	D3	D2	M	A/S
1	1	0	1	0	0	0	1	0	1	0	0	0
2	1	1	0	0	0	0	0	0	0	0	1	0
3	0	1	0	0	0	0	0	0	0	0	1	0
4	1	1	0	0	0	0	1	1	0	0	0	0
5	1	1	0	0	0	0	0	0	0	0	1	0
6	1	1	0	0	0	0	0	0	0	0	1	0
7	0	0	0	0	0	1	0	0	0	0	0	1
8	0	0	1	0	0	0	0	0	1	0	0	0
9	0	0	0	0	0	1	0	0	0	0	0	1
10	1	1	0	0	0	0	0	0	0	0	1	0
11	0	0	0	0	0	1	0	0	0	0	0	1
12	0	0	0	0	0	1	0	0	0	0	0	1
13	1	0	1	0	0	0	1	0	1	0	0	0
14	1	1	0	0	0	0	0	0	0	0	1	0
16	0	0	0	0	0	1	0	0	0	0	0	1
17	1	0	1	0	0	0	1	0	1	0	0	0
21	1	0	1	0	0	0	1	0	1	0	0	0
22	0	0	0	1	0	0	0	0	0	0	0	1
23	0	0	0	1	0	0	0	0	0	0	0	1
24	1	0	0	0	1	0	1	0	0	0	1	0
26	0	0	1	0	1	0	0	0	1	0	1	0
27	0	0	1	0	0	0	0	0	1	0	0	0
28	0	0	1	0	0	0	0	0	1	0	0	0
29	0	0	1	0	0	0	0	0	1	0	0	0
30	0	0	0	1	0	0	0	0	0	1	0	0
31	0	0	1	1	1	0	0	0	1	1	1	0
32	0	0	1	1	0	0	0	0	0	0	1	0
33	0	0	0	0	1	1	0	0	0	0	1	1
34	0	0	1	0	0	0	0	0	1	0	0	0
35	1	0	0	1	0	1	1	0	0	1	0	1
36	1	0	0	1	0	0	1	0	0	1	0	0

The GMS-DINA and MS-CDM-RT models were both used to analyze the response data of APM Test. To compare two models, we calculated the Deviance Information Criterion statistics (DIC), which was used as an index to measure the degree of fit between the model and the data. DIC is suitable for the model with MCMC algorithm as the parameter estimation method. Study provided two chains and the length of each chain was set to 10,000 with the first 5,000 as burn-in values, then the average value of last 5,000 values were taken as the estimated value of model parameters. The improved Gelman-rubin convergence statistic $\hat{R}$ was used to evaluate the convergence of the parameters. When $\hat{R} < 1.1$, the parameter estimation reaches the convergence standard, otherwise it does not converge (Brooks & Gelman, 1998).

The results showed that the parameter convergence index $\hat{R}$ values estimated for APM data were all less than 1.1, indicating the parameter estimation of the MCMC algorithm satisfies the convergence. The DIC values of the MS-CDM-RT and GMS-DINA are 29258.5 and 29227.4 respectively, which has negligible difference, indicating that two models had similar model-fit to the APM.

Table 11 gives the classification Templin reliability (Templin & Bradshaw, 2013) of MS-CDM-RT and GMS-DINA models on each attribute. It can be seen that the two models both have high reliability on all attributes (both greater than 0.8), and under the same attribute (except attribute M), the reliability of MS-CDM-RT model is better than that of GMS-DINA model, and this result can also be obtained on the mean value of attributes.

Table 11 Classification Templin reliability of MS-CDM-RT and GMS-DINA

Model	C	PP	D3	D2	M	A/S	Mean
MS-CDM-RT	0.918	0.985	0.958	0.915	0.827	0.996	0.933
GMS-DINA	0.832	0.965	0.858	0.907	0.888	0.995	0.924

According to the above description of Advanced Raven Reasoning Test, the items that can only be solved by QA strategy are 1, 4, 8, 13, 17, 21, 27, 28, 29, 30 and 34. And items that can only be solved by QV strategy are 7, 9, 11, 12, 16 and 33.

By analyzing the time taken by the students to answer these items, it was found that the average response time of the items using only QA strategy was 58.74 seconds, while the average response time of the students using only QV strategy was 35.40 seconds. The response time of the QV strategy was significantly shorter than that of the QA strategy (see Figure 1). This indicates that Verbal Analytic strategy (QA) takes longer to answer, which is consistent with the research results of Deshon et al. (1995).

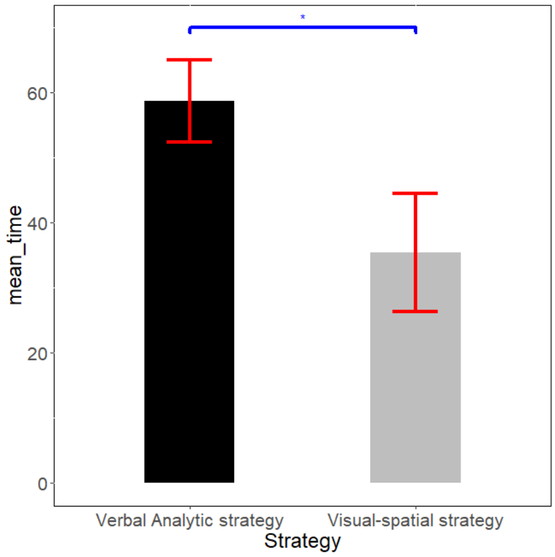

Figure 1. Average response time for each strategy (* refers to significant P <0.05 with Cohen's d=2.982)

Studies (Carpenter et al., 1990) have found that Verbal Analytic strategy involve attributes of D2 and D3 with higher difficulty, while Visual spatial strategy involve attributes of M and A/S with lower difficulty. Therefore, in theory, for items that can be solved by either Verbal Analytic strategy or Visual spatial strategy, the latter will have lower difficulty in solving problems and higher probability of correct answers. As can be seen from Table 10, items 2, 3, 5, 6, 10, 14, 22, 23 and 32 can be solved either through Verbal Analytic strategy (QA) or Visual spatial strategy (QV). Further analysis of the above items can be carried out. Both MS-CDM-RT and GMS-DINA defined the formula for a correct response using a certain strategy as: $\delta_{j0}+\delta_{jm1}$. As shown in Table 12, the item parameter estimation results of the two models both indicated that the item parameters corresponding to the Visual spatial strategy for most such items were smaller than those of the Verbal Analytic strategy ($\delta_{j21}<\delta_{j11}$). Therefore, $\delta_{j0}+\delta_{j21}<\delta_{j0}+\delta_{j11}$, which means that using the Visual spatial strategy will have a higher probability of correct response. This is consistent with the conclusions of Carpenter et al. (1990).

Table 12 Parameter estimates and its standard errors (in parenthesis) of different model of APM test

Item	MS-CDM-RT			GMS-DINA		
	δ_{j0}	δ_{j11}	δ_{j21}	δ_{j0}	δ_{j11}	δ_{j21}
2	0.093(0.032)	0.852(0.032)	0.810(0.049)	0.080(0.019)	0.865(0.018)	0.858(0.021)
3	0.086(0.029)	0.848(0.031)	0.778(0.048)	0.088(0.024)	0.849(0.023)	0.833(0.033)
5	0.152(0.038)	0.788(0.038)	0.776(0.056)	0.148(0.025)	0.798(0.025)	0.786(0.029)
6	0.091(0.036)	0.855(0.036)	0.834(0.046)	0.108(0.023)	0.839(0.023)	0.835(0.024)
10	0.130(0.032)	0.777(0.037)	0.699(0.040)	0.166(0.023)	0.666(0.026)	0.749(0.035)
14	0.108(0.036)	0.820(0.036)	0.782(0.056)	0.152(0.024)	0.787(0.024)	0.775(0.031)
22	0.112(0.029)	0.803(0.037)	0.711(0.035)	0.175(0.017)	0.693(0.040)	0.629(0.017)
23	0.173(0.020)	0.766(0.023)	0.746(0.034)	0.181(0.015)	0.755(0.020)	0.638(0.019)
32	0.091(0.018)	0.649(0.019)	0.672(0.034)	0.153(0.018)	0.655(0.019)	0.650(0.018)

Response time can obtain information about the use of the strategy, and it is also helpful to analyze the strategy selection when the response time is incorporated into the model. Taking item 5 as an example, real participants coded 008, 010 and 016 had mastered all the attributes of item 5 for two strategies. Table 13 showed their estimated probabilities of selecting each strategy under GMS-DINA model and MS-CDM-RT model ($P(m|\alpha_c)$ and w_{ijm}), respectively. For different participant, the GMS-DINA model estimates the very similar probability for each strategy selection. In more detail, the GMS-DINA model judges that the probabilities of the three participants choosing the QA strategy are all 0.503, and the probabilities of choosing the QV strategy are all 0.497. As seen, it is very difficult for GMS-DINA model to determine which strategy three participants chose to complete the item, due to that the two strategy selection probabilities are very close. However, when adding additional response time information, that is, using the proposed MS-CDM-RT model, it estimates that three participants have very different probability of choosing strategies for answering item 5. More specially, the probabilities of participant 008 and participant 010 to choose QA strategy are 0.345 and 0.388 respectively, and the probabilities of choosing QV strategy are 0.655 and 0.612 respectively. Therefore, it is easy to judge that the two participants are more inclined to adopt QV strategy. In addition, the observed response times of the two participants answering item 5 are 25.26s and 27.78s respectively, which were closer to the average response time (35.4s) of using the QV strategy than that of QA strategy. Therefore, the result also tends to judge that these two students chose the QV strategy to solve the problem. For participant 016, the probabilities of choosing QA strategy and QV strategy were 0.602 and 0.398 respectively under the MS-CDM-RT model. Therefore, he is more likely to choose QA strategy. At the same time, his observed response time when answering item 5 was 61.26s, which was closer to the average response time (58.7s) of the QA strategy than that of QV strategy.

Table 13 Probability of choosing each strategy for item 5

Participant ID	MS-CDM-RT		GMS-DINA		response time (s)
	QA strategy	QV strategy	QA strategy	QV strategy	
008	0.345	0.655	0.503	0.497	25.26
010	0.388	0.612	0.503	0.497	27.78
016	0.602	0.398	0.503	0.497	61.26

Summary and Discussion

Despite the fact that a large number of CDMs have been developed in the psychometric literature, most of them 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 may be of great concern. In this study, we introduce and incorporate the information of response time into multi-strategy cognitive diagnosis model to improve the existing CDMs. Compared with the existing models, the proposed model has several obvious features: First, based on theories and empirical findings about how strategies are chosen in problem-solving processes, the proposed model successfully integrates the additional information of the individual's answer accuracy and response time into the same framework to define strategy choices, which can more effectively analyze the participants' strategy on each task/question. Second, the model can accommodate the situation that students adopt distinct strategies for different items and allow the estimation of the probability of using each strategy for each item given the attribute profiles. Last, the model has strong expansibility and flexibility as shown in the section of Real Data Illustration. In addition, the MS-CDM-RT model is an extension of the most widely used DINA model in the single-strategy cognitive diagnosis model. Of course, it can be easily extended to CDMs to define the item response function under a specific strategy, such as G-DINA (de la Torre, 2011) or log-linear CDM (Henson, Templin, & Willse, 2009).

Results from the simulation study indicate the proposed model had acceptable parameter recovery and had higher parameter estimation accuracy than the existing multi-strategy CDMs. Compared with the conventional GMS-DINA, the real data analysis showed that the proposed MS-CDM-RT model had a prominent advantage on classification reliability and the judgement of strategy selection for participant due to that the information of response time is included and help to define strategy choices.

Despite promising results, additional researches along these lines are needed for further study. First, the simulation study only considered fixed test length and number of attributes, and only the influence of the distribution of the attribute mastering mode and the sample size on the accuracy of the MS-CDM-RT model parameter estimation were considered. However, research had shown that the length of the test affected the accuracy of the diagnosis (de la Torre, Hong, & Deng, 2010). In future studies researchers could vary these conditions to examine the parameter recovery of the MS-CDM-RT model. Second, the focus of this study is the development of the RTMS-CDMs in response to the 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 than the psychometric models. Research questions such as how to correctly define the corresponding q-vectors also need to be carefully answered to exploit the potential of the proposed models. Third, although the idea of the proposed model can be easily extended to other complex CDMs (such as the G-DINA model), its effectiveness also needs further investigate. Finally, the existing research on multi-strategy cognitive diagnosis models is still relatively weak, and more research can be done in this field. Future research can continue to consider adding some additional information in the modeling strategy selection to make it closer to the real strategy selection state. For example, when an individual deals with cognitive tasks, a variety of strategies are involved. From the perspective of cognitive psychology, some psychological processes (such as information on brain nerve activity) of the individual's choice of strategies can be obtained, and then quantitative research can be carried out.

References

Brooks, S. P., Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434-455.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological review*, *97*(3), 404-431.

Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, *74*(4), 633-665.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179-199.

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333-353.

de la Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, *73*(4), 595-624.

de la Torre, J., Hong, Y., & Deng, W. (2010). Factors Affecting the Item Parameter Estimation and Classification Accuracy of the DINA Model. *Journal of*

Educational Measurement, *47*(2), 227-249.

DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's advanced progressive matrices: Evidence for multidimensional performance determinants. *Intelligence*, *21*(2), 135-155.

Fu, J., & Li, Y. (2008). Cognitively diagnostic psychometric models: An integrative review. ETS research report.

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.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258-272.

Lovett, Marsha, C., Anderson, John, & R. (1996). History of success and current context in problem solving: combined influences on operator selection. *Cognitive Psychology*, *31*(2), 168-217.

Ma, W., & Guo, W. (2019). Cognitive diagnosis models for multiple strategies. *British Journal of Mathematical and Statal Psychology*, *72*(2), 370-392.

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review*, *118*(3), 393-437.

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

Mislevy, R. J., Wingersky, M. S., Irvine, S. H., & Dann, P. L. (1991). Resolving mixtures of strategies in spatial visualization tasks. *British Journal of Mathematical and Statistical Psychology*, *44*(2), 265-288.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select stratagies. Journal of Experi-Mintal Psychology:General,135 (2), 207-236.

Rittle-Johnson, B., & Siegler, R. S. (1999). Learning to spell: Variability, choice, and change in children's strategy use. *Child Development*, *70*(2), 332-348.

Siegler, R. S. (1989). Hazards of mental chronometry: An example from children's subtraction. *Journal of Educational Psychology*, *81*(4), 497-506.

Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of AS-CM using the choice / no-choice method. *Journal of Experimental Psychology: General*, *126*(1), 71-92.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345-354.

Tatsuoka, K. K. (1995). Architecture of knowledge structure 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. 327-361). Hillsdale, NJ: Erlbaum.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, *31*(2), 181-204.