A Motivation-based Cognitive Diagnostic Model for Disengaged Responses Detection

Yingshi Huang *

Shiyu Wang †

Yanfang Pan, Xiangyu Lu, and Ping Chen‡

1 Introduction

Cognitive diagnostic testing (CDT) aims to provide students with feedback reflecting their mastery of each skill or knowledge (de la Torre & Douglas, 2004; de la Torre, 2011). Such information is not only effective for students' learning and development but powerful for the improvement of teaching design. Given that students' test performance will not directly hurt their grades and students will have no consequence for poor performances, CDT is a typical low-stake testing scenario. Compared to high-stake testing, students tend to display lower test-taking motivation and a more significant proportion of disengaged responses, such as rapid guessing and omission behavior (Ulitzsch, Shin, & Lüdtke, 2023; Wise & Gao, 2017). Students may make no attempt to activate their knowledge in answering the test items, but rather randomly select an answer or even skip the items (Ulitzsch, von Davier, & Pohl, 2020; Zhu, Arthur, & Chang, 2022). This phenomenon undermines the validity of CDT. If observed responses contain no information about the problem-solving process, educators

*School of Education & Information Studies, University of California, Los Angeles, USA

†Department of Educational Psychology, University of Georgia, Athens, USA

‡Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, China; CONTACT: Ping Chen. Email: pchen@bnu.edu.cn

can never achieve precise estimates of students' latent knowledge profiles. For instance, if a student has mastered a knowledge attribute but answers the item incorrectly due to low test-taking motivation, educators may misclassify the student into the non-mastery group and conclude that this student requires further study. Additionally, when implementing the pre-test for item calibration, low-quality response data would impair the estimation of the Q-matrix in CDT (Hsu, Jin, & Chiu, 2020).

To detect disengaged responses, researchers have proposed various models to capture rapid guessing and omission behaviors (Lu, Wang, & Shi, 2023; Ulitzsch et al., 2020). The principle for these models is to construct different model structures for the disengaged behaviors and normal behaviors (Hsu et al., 2020; Lu, Wang, & Shi, 2021; C. Wang & Xu, 2015). However, most models were developed under the item response theory (IRT) and little attention was paid to the CDT. A classic IRT model for disengaged response detection is the hierarchical mixture model proposed by C. Wang and Xu (2015). A higher-order latent discrete variable was defined to represent whether a specific response should be labeled as rapid guessing. Further, Ulitzsch et al. (2020) considered both rapid guessing and omission behaviors by redesigning the higher-order latent variable with a continuous latent trait representing the engagement tendency of a student. These approaches are built with IRT and leverage the hierarchical modeling framework. This hierarchical structure enables researchers to jointly model the response accuracy and response time with a multivariable distribution in the second layer, but it also constrains the model into continuous variables which leads to the limited model development in CDT. Considering that the attribute profiles in CDT are discrete, it is hard to integrate the cognitive diagnostic models (CDMs) into the multivariable distribution. To the best of our knowledge, only Hsu et al. (2020) extended the mixture model proposed by C. Wang and Xu (2015) into the CDT scenario. The higher-order CDM was used to estimate a general continuous latent trait for attribute patterns. However, this approach would considerably increase the number of parameters needed to be estimated and bring computational issues. To reduce the computational burden, Hsu et al. (2020) focused

only on rapid guessing detection.

In this paper, we formulate a new motivation-based cognitive diagnostic model (MCDM), leveraging multiple-level attributes instead of the hierarchical structure to connect students' test-taking motivation and latent profiles. We further put forward a plug-and-play modeling framework that enables developers to deal with both rapid guessing and omission and to differentiate two kinds of omission behavior (i.e., failing to generate an answer or intentionally skipping). In the following sections, the traditional CDMs will be described. Next, the new MCDM model is introduced followed by the simulation and application studies. Finally, a discussion of suggestions provided for practitioners and future directions are presented.

2 Traditional Cognitive Diagnostic Models

CDMs can offer information about how well a student masters a specific knowledge or skill. The validity of such inference relies heavily on the assumption that students actively engage in the test and thus manage to trigger their knowledge to solve the items (Rupp et al., 2010). In other words, it is assumed that students possess high test-taking motivation during the problem-solving process. Hence, the probability of giving a correct answer to the item is purely contributed by students' knowledge profiles and item parameters. Specifically, CDMs assume that students can answer an item correctly only when they master or partially master all knowledge attributes measured by the item. Let K denote the number of attributes involved in a test with J items. To establish the relationship between attributes and items, a $J \times K$ Q-matrix specifying the attributes needed for each item is calibrated by domain experts. The jth row of the Q-matrix, $\boldsymbol{q}_j = (q_{j1}, q_{j2}, \ldots, q_{jK})'$, indicates the attribute vector measured by item j with $q_{jk} = 1$ if mastering attribute k is necessary to correctly answer item j and $q_{jk} = 0$ otherwise. The knowledge profiles of student $i (i = 1, 2, \ldots, N)$ can be represented as $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK})'$ and $\alpha_{ik} = 1$ signifies the mastery of the kth attribute and 0 otherwise.

CDMs can be categorized into non-compensatory and compensatory models based on whether the mastery of all attributes measured by the item is required for a correct response. Non-compensatory models, such as the deterministic inputs, noisy "and" gate model (DINA) (Junker & Sijtsma, 2016) and the noisy inputs, deterministic "and" gate model (NIDA) (de la Torre & Douglas, 2004), assume that only when mastering all the attributes involved in an item can the student answer the item correctly. For compensatory models, such as the deterministic inputs, noisy "or" gate model (DINO) (Templin & Henson, 2006), only partial mastery of the attributes is required (von Davier & Lee, 2019). Note that non-compensatory models are nested within the compensatory models, as only the joint effect of all attributes is considered in non-compensatory models. To provide a coherent line for various CDMs, generalized frameworks like generalized DINA (G-DINA) (de la Torre, 2011) and log-linear CDM (Henson, Templin, & Willse, 2009) have been developed, which hold the properties of compensatory models. Next, DINA, DINO, and G-DINA are introduced as classic models for non-compensatory, compensatory, and generalized models, respectively.

2.1 Non-compensatory Models

DINA is a typical example of non-compensatory models, restricting that a success response on an item requires the mastery of all attributes measured. The latent response for student i on item j is defined as $\eta_{ij}^{\text{DINA}} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$, that is, if the student possesses all attributes needed for the item j,$\eta_{ij}^{\text{DINA}} = 1$; otherwise, $\eta_{ij}^{\text{DINA}} = 0$. In reality, even students mastering all attributes may answer incorrectly due to slipping, while those who have not mastered all attributes could provide a correct response by guessing. Hence, slipping (s_j) and guessing (g_j) parameters are included in the model and defined as $s_j = P(Y_{ij} = 0|\eta_{ij}^{\text{DINA}} = 1)$ and $g_j = (Y_{ij} = 1|\eta_{ij}^{\text{DINA}} = 0)$. The probability of student i correctly answering item j is expressed as:

$$P(Y_{ij} = 1|\boldsymbol{\alpha}_i) = (1 - s_j)^{\eta_{ij}^{\text{DINA}}} g_j^{(1-\eta_{ij}^{\text{DINA}})}. \tag{1}$$

2.2 Compensatory Models

Differing from DINA, DINO posits that the mastery of at least one attribute measured by the item is adequate for providing a correct response. Hence, the latent response for the student i on the item j is redefined as $\eta_{ij}^{\text{DINO}} = 1 - \prod_{k=1}^{K}(1-\alpha_{ik})^{q_{jk}}$, meaning that $\eta_{ij}^{\text{DINO}} = 1$ when at least one attribute needed for the item is mastered. Similar to DINA, the slipping and guessing parameters are incorporated and defined as $s_j = P(Y_{ij} = 0|\eta_{ij}^{\text{DINO}} = 1)$ and $g_j = (Y_{ij} = 1|\eta_{ij}^{\text{DINO}} = 0)$. Accordingly, the probability of giving a correct answer by the student i on the item j is formulated as:

$$P(Y_{ij} = 1|\boldsymbol{\alpha}_i) = (1 - s_j)^{\eta_{ij}^{\text{DINO}}} g_j^{(1-\eta_{ij}^{\text{DINO}})}. \tag{2}$$

To provide a unified framework for different CDMs, the G-DINA model was developed, modeling the correct probability with main effects and interaction effects. Specifically, under the identity link, the correct probability of student i on item j is:

$$P(Y_{ij} = 1|\boldsymbol{\alpha}_{ij}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{ik} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\alpha_{ik}\alpha_{ik'} + ... + \delta_{j12,...,K_j^*} \prod_{k=1}^{K_j^*} \alpha_{ijk}, \tag{3}$$

where $K_j^* = \sum_{k=1}^{K} q_{jk}$ represents the number of attributes measured by item j ($K_j^* < K$). $\boldsymbol{\alpha}_{ij}^*$ is the reduced attribute profile for item j. δ_{j0} is the intercept, indicating the correct probability when students mastering no knowledge; δ_{jk} denotes the main effect of attribute k, representing the influence of mastering attribute k on the probability of answering correctly; $\delta_{jkk'}$ denotes the interaction between attributes k and k', and $\delta_{j12,...,K_j^*}$ represents the interaction among all K_j^* attributes. Notice that in G-DINA, the interaction effects of different attributes affect the probability of answering correctly, thus belonging to the compensatory model along with DINO.

3 Motivation-based Cognitive Diagnostic Model

Assuming that students actively engage in answering the test items, CDMs decipher students' latent knowledge profiles with their observed responses (de la Torre, 2011). Ideally, when answering an item, students with high test-taking motivation intend to activate the attribute(s) measured by the item and try their best to solve the problem. In this situation, we can safely conclude that the responses and response times generated by students contain reliable information about students' attribute profiles. Hence, our inference for the attributes involved in the item is reliable and valid. In contrast, students with low test-taking motivation may invest no effort in answering the item, rapidly guessing an answer or even skipping the item. Therefore, the responses and response times reflect no information on students' mastery of the attributes measured by the item. For instance, in a non-graded classroom arithmetic exam, the examinee with low test engagement may randomly select an option as the answer or skip the item. In this case, the inference for the attribute(s) involved in the item contains noises from the disengaged responses.

It is clear that valid inference for students' knowledge profiles is contingent on high-quality evidence (e.g., response accuracy and response time) and the active statuses of students' attribute profiles mirror their test-taking motivation levels. If a low-motivative student does not attempt to activate the attribute(s) measured by the item while answering, the conclusion about whether this student masters the attribute(s) or not becomes doubtful. Based on this principle, this study proposed a motivation-based CDM (MCDM) to detect disengaged responses by labeling the problematic attribute(s) in the problem-solving process. With MCDM, we can pinpoint the specific attribute(s) affected by disengaged responses and refine the inference of students' knowledge profiles, enhancing the quality of decision-making.

3.1 Model Specification

To detect disengaged responses, MCDM employs the idea of multiple-level attributes to connect students' test-taking motivation and latent profiles. To be more specific, when a student masters the attribute(s) measured by the item and actively engages in answering the item, the students' performance reflects their peak level; when a student does not master the attribute(s) involved in the item but makes a best effort to solve the item, the performance in this situation should be inferior to the peak. Lastly, the student's performance will reach its lowest level when the student disengages in answering the item. Therefore, instead of using a dichotomous variable to indicate students' mastery of attributes, the knowledge status for attribute k of student i is defined as: (1) $\alpha_{ik} = 1$ when the student masters the attribute k and actively leverage the knowledge to solve the item (i.e., mastery); (2) $\alpha_{ik} = 0$ when the student does not master the attribute k but still endeavors to solve the item (i.e., non-mastery); (3) $\alpha_{ik} = -1$ when the student invests no efforts in answering the item and thus shows disengaged responses (i.e., non-active). Under this definition, we can further develop response accuracy, response time, and omission models for the normal behavior and disengaged behavior.

3.1.1 Response accuracy model

To classify the normal and disengaged responses, similar to DINA, the latent response η_{ij} can be defined as:

$$\eta_{ij} = \mathbb{I}\left(\sum_{k=1}^{K} \alpha_{ik}^{q_{jk}} = \sum_{k=1}^{K} q_{jk}\right) - \mathbb{I}\left(\sum_{k=1}^{K} \alpha_{ik}^{q_{jk}} < \sum_{k=1}^{K} |\alpha_{ik}|^{q_{jk}}\right). \tag{4}$$

Here, $\mathbb{I}(\cdot)$ is an indicator function, taking the value of 1, when a condition is satisfied and 0 otherwise. Thus, when examinee i activates and masters all attribute(s) assessed by item j, η_{ij} equals 1; when examinee i activates but does not master all the attributes measured by item j, η_{ij} equals 0. Whereas, when examinee i does not activate one of the

attributes involved in item j, that is, there exists a $q_{ik} = 1$ but $\alpha_{ik} = -1$, $\eta_{ij} = -1$. The first two scenarios are considered normal responses, while the last case is identified as a disengaged response. To illustrate this definition, a bare-bones example of Test A measuring five attributes was presented. Suppose the item j in this test assesses three attributes, and there are five examinees with different knowledge profiles. The latent response η_{ij} of all students can be computed for each item (see Table 1).

Table 1: Example of normal and disengaged behaviors.

		$q_j = [10110]$
Normal behavior	$\eta_{1j} = 1$	$\alpha_1 = [10111]$
	$\eta_{2j} = 0$	$\alpha_2 = [10000]$
Disengaged behavior	$\eta_{3j} = -1$	$\alpha_3 = [-10110]$
	$\eta_{4j} = -1$	$\alpha_4 = [-10000]$
	$\eta_{5j} = -1$	$\alpha_5 = [-10-1-10]$

The probability of giving a correct answer on item j for a student i with knowledge status α_i can be formulated as:

$$P(Y_{ij} = 1|\boldsymbol{\alpha}_i) = \begin{cases} 1 - s_j & \text{if } \eta_{ij} = 1, \\ g_{1j} & \text{if } \eta_{ij} = 0, \\ g_{2j} & \text{if } \eta_{ij} = -1, \end{cases} \tag{5}$$

where s_j represents the slipping parameter, indicating the probability that a student who masters and activates all attributes required by item j answers the item incorrectly due to slip and mistake (when using the identity link, $s_j = \delta_{j0} + \delta_{j12,\ldots,K_j^*}$); g_{1j} is the guessing parameter for engaged responses, indicating the probability of a student, who has activated all attributes measured by the item but not fully mastered all of them, guessing correctly on the item ($g_{1j} = \delta_{j0}$ for the identity link); g_{2j} denotes the guessing parameter for disengaged responses, indicating the correct response probability for a student who has not activated all attributes required by the item. Furthermore, following the monotonic increasing assumption, that is, as the level of knowledge attributes increases, the student has a higher probability of

answering the item correctly, we can have: $0 < g_{2j} < g_{1j} < 1 - s_j < 1$.

3.1.2 Response time model

One main component of disengaged responses is the rapid guessing behavior. To label the rapid guessing response, students' response time spent on answering an item is a crucial indicator. As can be seen in Figure 1, the response time distribution shows a clear bimodal pattern, suggesting different means and variances exist in the response time distribution. Therefore, this study models response times separately for engaged and disengaged responses.

The response time distribution for engaged responses follows a log-normal distribution, determined by the time intensity parameter β_M, the speed τ of the student, and response time variance σ_M^2. Additionally, based on the common-guessing theory (Schnipke & Scrams, 1997), disengaged responses contain no information about students' knowledge profiles and response speed. Consequently, the response time distribution for disengaged responses follows common item time intensity parameters (β_D) and variance (σ_D^2). The response time t for student i on item j is as follows:

$$\ln(t_{ij}) \sim \begin{cases} N(\beta_{Mj} - \tau_i, \sigma_{Mj}^2) & \text{if } \eta_{ij} = 0 \text{ or } 1, \\ N(\beta_D, \sigma_D^2) & \text{if } \eta_{ij} = -1. \end{cases} \tag{6}$$

The main difference between engaged and disengaged responses is the underlying problem-solving process. The gap between engaged and disengaged responses reflects the student's encoding and cognitive processing efforts to solve the item (Y. Chen, Yang, & Lee, 2022). Compared to disengaged responses, students with high test-taking motivation spend more time on reading, understanding, and activating relevant knowledge to answer the item-what we refer to as the wake-up phrase (Ulitzsch et al., 2020; Wise & Gao, 2017). Thus, we define β_j^* to represent the wake-up time required for actively answering item j, resulting in $\beta_{Mj} = \beta_j^* + \beta_D$ where $\beta_j^* \geq 0$.

In this model, the effect of knowledge profiles on response time is embedded in η_{ij}.

Equation 6 can be expressed in a mixture modeling fashion as: $\ln(t_{ij}|\eta_{ij}) = [1 - P(\eta_{ij} = -1)]f(t_{ij}^{M}) + P(\eta_{ij} = -1)h(t_{ij}^{D})$ where $f(\cdot)$ represents the response time function for engaged responses, and $h(\cdot)$ represents the response time function for disengaged responses.

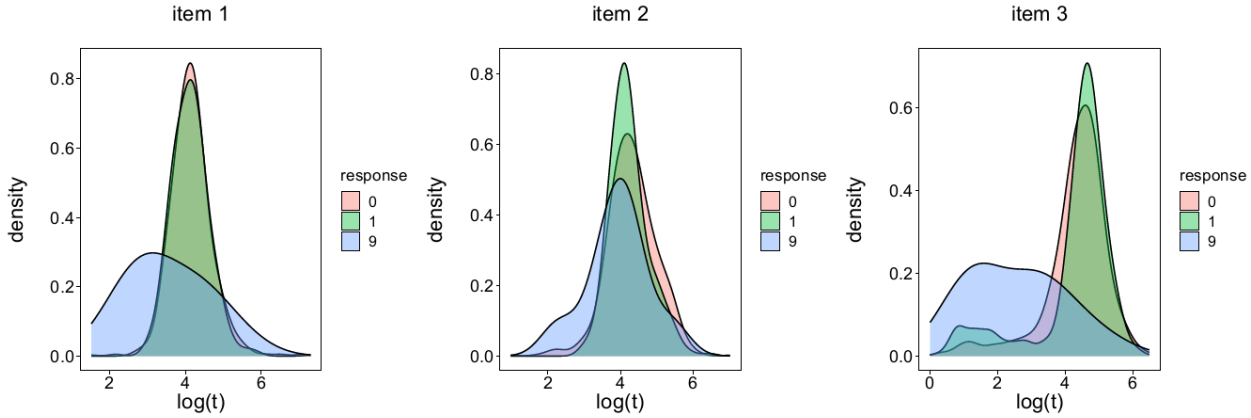


Figure 1: Response time distribution in log-scale for three items of the PISA 2015 math assessment (0 = incorrect response; 1 = correct response; 9 = missing).

3.1.3 Omission model

Another important indicator of disengaged response detection is the omission behavior. The challenge in modeling omission behavior lies in its dual nature. That is, omission may exist in both normal and disengaged responses. The question will be: How can we distinguish two types of omission in the model? Existing models for omission primarily employ a regression fashion, predicting students' omitted responses based on students' abilities and speeds (Ulitzsch et al., 2020). This approach is straightforward and intuitive, allowing researchers to clearly understand the contribution of different factors by analyzing the coefficients. However, the selection of predictors implies assumptions of the underlying mechanism of omission behavior. Specifically, building an omission model with students' abilities and speeds assumes that students' knowledge statuses and speed are the two biggest factors that account for omission. Models relying on strong assumptions inevitably encounter model-fitting issues such as missing key variables.

Differing from the regression approach, we leverage the multi-level attribute to construct

the latent response η_{ij} and separate two types of omission responses without making assumptions on the model predictors. The probability of student i omitting item j is formulated as:

$$P(O_{ij} = 1) = (1 - r_j)^{\{1-\mathbb{I}(\eta_{ij}=-1)\}} o_j^{\mathbb{I}(\eta_{ij}=-1)}, \tag{7}$$

where r_j is the responding parameter, indicating the probability of a student with high test-taking motivation giving a response on the item; o_j denotes the omitting parameter, representing the probability of a student with low test-taking motivation leaving the item blank. $\mathbb{I}(\cdot)$ is an indicator function, which has a value of 1 when $\eta_{ij} = -1$, and a value of 0 otherwise. To ensure that disengagement will introduce a higher omission rate than engaged problem-solving behavior, the constraint of $0 < 1 - r_j < o_j < 1$ is imposed. Note that in Equation 7, no assumptions of predictors are made, so it is flexible to embrace potential factors in the model by including covariates to build relationships between variables of interest and the r_j and o_j parameters.

3.2 Model Summary

MCDM identifies the disengaged responses by defining the active latent response $\eta_{ij} = 0/1$ and the resting (non-active) latent response $\eta_{ij} = -1$. The novel model allows for an understanding of how examinees extract various knowledge attributes during the exam and captures the potentially unreliable attribute. In the current study, MCDM is built upon the response accuracy, response time, and omission. Within this framework, it is straightforward for researchers to further expand or simplify the current model by adding or canceling data components. Additionally, compared to traditional CDMs with multi-attributes (J. Chen & de la Torre, 2013; Ma & Torre, 2016), the attribute level in MCDM has a unique function of disengaged response detection. Below, we offer a detailed discussion of MCDM in terms of model variation and connection with multi-attribute CDMs, aiming to provide readers with

a comprehensive summary of MCDM.

3.2.1 Model variation

The current model framework incorporates both rapid guessing and omissions as indicators of disengaged responses. It is plausible to jointly model these two behaviors, considering that, in the typical testing scenario, full control is given to the test-taker, meaning that examinees can decide whether to answer the presented item or leave it blank. Further, as both the response time and omission model can be viewed in a mixture modeling framework that enjoys a "plug-and-play" nature, researchers can easily remove the unnecessary parts according to their specific research purposes. For example, if the testing procedure forces examinees to answer the current item before moving to the next one, researchers can cancel the omission model from MCDM. In our simulation and empirical study, we investigated the recovery of parameters under the 0% omission condition and fit the reduced model with real data. In addition, it is also promising to expand the model by incorporating more indicators, such as eye-tracking data, to gain a deeper understanding of the problem-solving process.

3.2.2 Connection with multi-attribute CDMs

MCDM utilizes a multi-level attribute framework to depict students' test-taking engagement. The key distinction between this model and existing attributes-based multi-level CDM (J. Chen & de la Torre, 2013) lies in the model objective: While multi-level CDM aims to analyze the variation in the mastery status of students' knowledge attributes, MCDM focuses on identifying the specific attribute(s) affected by disengaged responses.

By employing the design of multi-level attributes, researchers utilize various types of data to explore traits of interest. For instance, S. Wang and Chen (2020) and Su and Davison (2019) defined students' knowledge attribute states or abilities into 0, 1, and 2 levels, with high accuracy and speed associated with the highest proficiency level, identifying students who can answer questions quickly and accurately. Similarly, this study utilizes multi-level

attributes to detect the rapid guessing and omission behavior, thereby delving deeper into students' test-taking motivation.

3.3 Estimation Procedure

MCDM is estimated with the Markov chain Monte Carlo (MCMC) method under the Bayesian framework. In general, MCMC obtains unbiased estimates of parameters by approximating the complex posterior distributions with the likelihood and priors. Considering that the current model we are proposing contains three parts of parameters to be estimated, to improve estimation efficiency, the Hamiltonian Monte Carlo (HMC) method is implemented in the Stan program (Betancourt, 2017). The HMC method enhances the sampling efficiency of posterior distributions by treating posterior samples as points with potential and kinetic energy and by leveraging gradients to define the sampling direction. HMC offered by the Stan program can reach the targeted distribution with fewer iterations compared to other algorithms and thus is favored in this study (Luo, De Carolis, Zeng, & Jeon, 2023).

The joint likelihood function of response accuracy, response time, and omission is as follows:

$$\mathcal{LL} = f(\tau|\mu_\tau, \sigma_\tau^2) \times \prod_{i=1}^{N}\prod_{j=1}^{J}(P(Y_{ij}|\boldsymbol{\alpha}_i, s_j, g_{1j}, g_{2j})^{(1-O_{ij})} f(\ln t_{ij}|\boldsymbol{\alpha}_i, \tau_i, \beta_j^*, \sigma_{Mj}^2, \beta_D, \sigma_D^2) P(O_{ij}|\boldsymbol{\alpha}_i, r_j, o_j)), \tag{8}$$

where $f(\tau|\mu_\tau, \sigma_\tau^2)$ is the marginalized probability density function of latent speed τ. To ensure identifiability of the log-normal distribution, we restrict $\mu_\tau = 0$ (van der Linden, 2016) and let $\sigma_\tau^2 = 1$. The formula,

$$P(Y_{ij}|\boldsymbol{\alpha}_i, s_j, g_{1j}, g_{2j}) = P(Y_{ij} = 1)^{Y_{ij}}(1 - P(Y_{ij} = 1))^{(1-Y_{ij})}, \tag{9}$$

represents the likelihood function for response accuracy. Whereas,

$$f(\ln t_{ij}|\boldsymbol{\alpha}_i, \tau_i, \beta_j^*, \sigma_{Mj}^2, \beta_D, \sigma_D^2) \tag{10}$$

is the probability density for log-scale response time. And,

$$P(O_{ij}|\boldsymbol{\alpha}_i, r_j, o_j)) = P(O_{ij}=1)^{O_{ij}}(1-P(O_{ij}=1))^{(1-O_{ij})} \tag{11}$$

denotes the likelihood function for omission. For priors, the settings used in the simulations and empirical study follow Ulitzsch et al. (2020) and C. Wang and Xu (2015) and are shown in Table 2.

As mentioned above, the estimation is implemented in the Stan program in conjunction with R software. The HMC algorithm in Stan utilizes the No-U-Turn sampler to dynamically adjust the step size, allowing for effective integration of the Hamiltonian trajectory (Betancourt, 2017). It should be noted that HMC cannot directly estimate categorical variables. Therefore, Table 2 does not directly include prior settings for the latent attribute profile $\boldsymbol{\alpha}_i$, but instead displays the probability of occurrence for each attribute pattern $\pi_c = P(\boldsymbol{\alpha} = \boldsymbol{\alpha}_c)$ where $c = 1, 2, \ldots, 3^K$.

In the sampling process, to ensure the monotonic assumption can be satisfied (e.g., $0 < g_{2j} < g_{1j} < 1 - s_j < 1$), researchers typically use use truncated distribution for the constrained parameters (Jiang & Carter, 2019; S. Wang & Chen, 2020). However, in the pilot study, we found that using truncated distributions for MCDM introduced dependencies among parameters. Specifically, the truncation boundary of a specific parameter is determined by another estimate, resulting in non-convergence in the estimation process. Therefore, in this study, instead of directly constructing posterior distributions for constrained item parameters, we built the posteriors of transformation parameters $\mathcal{R} \sim Beta(\alpha_{\mathcal{R}}, \beta_{\mathcal{R}})$ for response accuracy item parameters and $\mathcal{O} \sim Beta(\alpha_{\mathcal{O}}, \beta_{\mathcal{O}})$ for omission item parameters, where $\alpha_{\mathcal{R}} = \alpha_{\mathcal{O}} = 5$ and $\beta_{\mathcal{R}} = \beta_{\mathcal{O}} = 10$. Specifically, the transformation parameters were sampled and sorted, which ensured the constraints and independence among parameter

Table 2: The prior settings of model parameters.

Parameters	Prior setting	
	Response accuracy model	
$1 - s_j$	$Beta(\alpha_s, \beta_s)$	$\alpha_s = 5, \beta_s = 10$
g_{1j}	$Beta(\alpha_{g_1}, \beta_{g_1})$	$\alpha_{g_1} = 5, \beta_{g_1} = 10$
g_{2j}	$Beta(\alpha_{g_2}, \beta_{g_2})$	$\alpha_{g_2} = 5, \beta_{g_2} = 10$
π_c	$Dirichlet(\tilde{N}_1, ...\tilde{N}_C)$	$\tilde{N}_1 = ... = \tilde{N}_C = 1$
	Response time model	
β_j^*	$N(\mu_1, \sigma_1^2)\mathbb{I}(\beta_j^* \geq 0)$	$\mu_1 = 0, \sigma_1^2 = 5$
β_D	$N(\mu_2, \sigma_2^2)$	$\mu_2 = 0, \sigma_2^2 = 5$
σ_{Mj}^2	$InvGamma(a_1, b_1)$	$a_1 = 0.1, b_1 = 0.1$
σ_D^2	$InvGamma(a_2, b_2)$	$a_2 = 0.1, b_2 = 0.1$
τ_i	$N(0, 1)$	-
	Omission model	
$1 - r_j$	$Beta(\alpha_r, \beta_r)$	$\beta_r = 10$
o_j	$Beta(\alpha_o, \beta_o)$	$\beta_o = 10$

Note. The priors for parameters of the omission model were determined based on the omission rate in the empirical data, making the peak of the posterior align with the real data.

distributions could be met.

4 Simulation Study

In this section, the main purpose is to validate the estimation process of MCDM. Model convergence and estimation accuracy for the proposed model are reported. We investigate the impact of sample size, omission rate and disengaged response rate on the estimation process, which can offer us a clear picture of whether we can obtain robust estimators under diverse conditions.

4.1 Research Design

Three factors were manipulated: (1) Sample sizes of N = 1,000 and 1,500, which indicate medium and large sample size, respectively; (2) Omission rates = 0%, 5%, and 10%, following the omission rate observed in PISA 2015 Math across different countries (see Appendix A),

which represent no missing, low missing rate, and medium missing rate respectively; (3) Disengaged response rates = 30% and 50%, following the observed proportion of disengaged students reported in previous studies (Hoyt, 2001; OECD, 2019).

As shown in Table 3, the number of attributes were $K = 3$ corresponding to three attributes, space and shape (α_1), quantity (α_2) and uncertainty and data (α_3), being measured in the PISA 2015 Math assessment. The number of items was set to 16, and the Q matrix was set according to the PISA Technical Report (OECD, 2017b). Four identity sub-matrices were contained in the Q matrix to satisfy the identification requirement of s_j, g_{1j} and g_{2j} (Xu, 2019). Also, note that r_j and o_j can be identifiable once we ensure the identifiability of the response accuracy model and thus are free from extra identification burden.

In terms of person parameters, speed τ_i was assumed to follow $N(0, 0.5)$ with $\sigma_\tau^2 = 0.5$ (Ulitzsch et al., 2020; C. Wang & Xu, 2015). To obtain latent profiles, we first generated $p(\alpha_{ik} = 1)$ from multivariate standard normal distribution. That is, students have 50% of mastering each attribute. We then randomly selected 30% or 50% of students and generated $p(\alpha_{ik} = -1)$ for those selected attribute profiles from multivariate standard normal distribution to produce disengaged responses.

In terms of item parameters, (g_2, g_1, s) were set to $(0.1, 0.2, 0.2)$ for all items, representing high item quality; the disengaged time parameters were $\beta_D = 3$ and $\sigma_D^2 = 1.05$; the wake-up parameter was $\beta_j^* = \{0.25, 0.5, 0.75, 1, 1.25\}$, and the time variance for motivated responses was $\sigma_M^2 = 0.2$. Moreover, we set $o = 0.05$ and $r = 0.96$ to reach the 5% missing rate (with $\alpha_r = \alpha_o = 1$) and set $o = 0.1$ and $r = 0.93$ for the 10% missing rate (with $\alpha_r = \alpha_o = 2$). When the missing rate was 0%, we set $P(O_{ij} = 1) := 0$.

The HMC algorithm was used to estimate parameters. Four MCMC chains were set with 5,000 iterations each and the first 2,500 iterations were discarded.

Table 3: The Q matrix.

Item	space and shape (α_1)	quantity (α_2)	uncertainty and data (α_3)
M033Q01	1	0	0
M474Q01	0	1	0
M411Q01	0	1	0
M411Q02	0	0	1
M803Q01	0	0	1
M442Q02	0	1	0
M034Q01	1	0	0
M305Q01	1	0	0
M496Q01	0	1	0
M496Q02	0	1	0
M423Q01	0	0	1
M406Q01	1	0	0
M406Q02	1	0	0
M603Q01	0	1	0
M564Q01	0	1	0
M564Q02	0	0	1

Note. Items in shadow are selected from M02 item set while others are selected from M01.

4.2 Evaluation Criteria

To assess model convergence and estimation efficiency, the Gelman-Rubin $\hat{R}$, effective sample sizes (ESSs), and running time (t) were used. Specifically, we computed the proportion of $\hat{R}$ below 1.1 and ESSs $>$ 1,000.

For estimation accuracy, the differences between estimates and true values were reported for item parameters, and the correlation between the estimated speed $\hat{\tau}$ and true speed τ was calculated (denoted as $\rho_{\hat{\tau}\tau}$). With regard to attribute patterns, the attribute-wise agreement rate (AAR) was as follows:

$$\text{AAR} = \frac{\sum_{i=1}^{N}\sum_{k=1}^{K}\mathbb{I}(\hat{\alpha}_{ik} = \alpha_{ik})}{N \times K}. \tag{12}$$

Additionally, the sensitivity and specificity were employed to examine whether the disengaged responses could be identified correctly. The sensitivity was defined as true positive rate (TPR), while the specificity was true negative rate (TNR). TPR and TNR were as

follows:

$$\text{TPR} = \frac{\sum_{i=1}^{N}(p_i^{11}/p_i^{1+})}{N}, \tag{13}$$

$$\text{TNR} = \frac{\sum_{i=1}^{N}(p_i^{00}/p_i^{0+})}{N}, \tag{14}$$

where p_i^{11} was the probability of classifying the true engaged student i to the normal group, and p_i^{00} was the probability of classifying the true low-motivated student i to the disengaged group. p_i^{1+} represented the marginal probability of labeling student ii as normal while p_i^{0+} had the opposite meaning.

4.3 Results

In terms of model convergence, Table 4 summarizes the proportion of $\hat{R}$ less than 1.1 and ESSs greater than 1,000. Results indicated that MCDM was well convergent under all conditions, with over 99.9% of the estimated parameters having $\hat{R}$ below 1.1 and over 95.8% of parameters having ESSs exceeding 1,000. Further inspection reveals that the low ESSs primarily concern the wake-up parameter β_j^* and the speed parameter τ_i. This could be attributed to the fact that the response times of normal response are determined by three parts: β_j^*, β_D, and τ_i. When disengaged responses account for only a small fraction of all responses, the estimation of β_j^* and τ_i may be influenced by the uncertainty that exists in β_D, resulting in a dissatisfied effective sample size.

For estimation efficiency, Table 4 further shows that as the number of parameters to be estimated increased, the time required for estimation extended. Specifically, the running time for a sample of 1,500 took longer than that for a sample of 1,000. Models with omission parameters required longer running time than those without missing data, but the degree of missingness and the proportion of disengaged responses had no significant impact on the running time.

The estimation accuracy for person parameters is shown in Table 5. Under all conditions, three types of classification accuracy, AAR, TPR, and TNR, were all above 0.9. As expected, a higher proportion of disengaged responses would increase TNR. Different missing rates and sample sizes had minimal impact on the attribute classification rates. Additionally, the correlation between the estimated speed and true values were all above 0.9 under all conditions, indicating high estimation precision.

The differences between true values and estimates for item parameters and 95% confidence interval (CI) were displayed in Figures 2 and 3. For the sake of conciseness, only results under missing rates of 5%, sample sizes of 1,000, and disengaged response proportions of 30% and 50% are presented. For parameters of the response accuracy model (i.e., s, g_1, and g_2), the discrepancies between estimates and true values were below 0.1, and the width of the 95% CI was within 0.1, indicating desirable recovery. For parameters of the omission model (i.e., r and o), all parameter deviations from the true values were less than 0.08. Concerning the response time model, all estimates differed from the true values by less than 0.1. The 95% CI for β_D,σ_D^2, and σ_M^2 were within 0.06, while the 95% CI for β_j^* was wider.

Further, based on results across all conditions, as the proportion of disengaged responses increased, the estimation precision of disengaged parameters g_2, o, β_D, and σ_D^2 improved. Specifically, the point estimates of the parameters were closer to the true values, and the 95% CI became narrower. Moreover, the sample size and missing rate had no significant impact on the estimation accuracy.

5 Application

We apply the proposed model to analyze the PISA 2015 Math assessment to unveil its potential for detecting disengaged responses and facilitating reliable decisions. We include the DINA model as the baseline and the reduced MCDM models to provide a plug-and-play view. The reduced models are expressed as MCDM-RT (modeling response accuracy and

Table 4: Model convergence indices and running time across different conditions.

missing (%)	N	disengaged (%)	$\hat{R} < 1.1$	ESSs$> 1,000$	$t(h)$
0	1000	30	1	0.959	11.165
		50	0.999	0.958	11.331
	1500	30	1	0.977	25.850
		50	1	0.972	19.955
5	1000	30	1	0.969	31.537
		50	1	0.969	27.591
	1500	30	1	0.972	49.539
		50	1	0.965	42.343
10	1000	30	1	0.963	27.190
		50	1	0.970	29.864
	1500	30	1	0.969	46.449
		50	0.999	0.967	54.830

Table 5: Classification accuracy and correlation coefficients for speed across different conditions.

missing (%)	N	disengaged (%)	AAR	TPR	TNR	$\rho_{\tau\hat{\tau}}$
0	1000	30	0.929	0.993	0.928	0.945
		50	0.921	0.986	0.945	0.935
	1500	30	0.924	0.991	0.927	0.958
		50	0.929	0.99	0.958	0.941
5	1000	30	0.922	0.993	0.9	0.935
		50	0.912	0.988	0.929	0.935
	1500	30	0.919	0.993	0.94	0.963
		50	0.914	0.983	0.944	0.943
10	1000	30	0.916	0.992	0.928	0.949
		50	0.914	0.979	0.929	0.939
	1500	30	0.913	0.991	0.914	0.955
		50	0.917	0.988	0.939	0.938

Note. AAR = attribute-wise agreement rate; TPR = true positive rate; TNR = true negative rate.

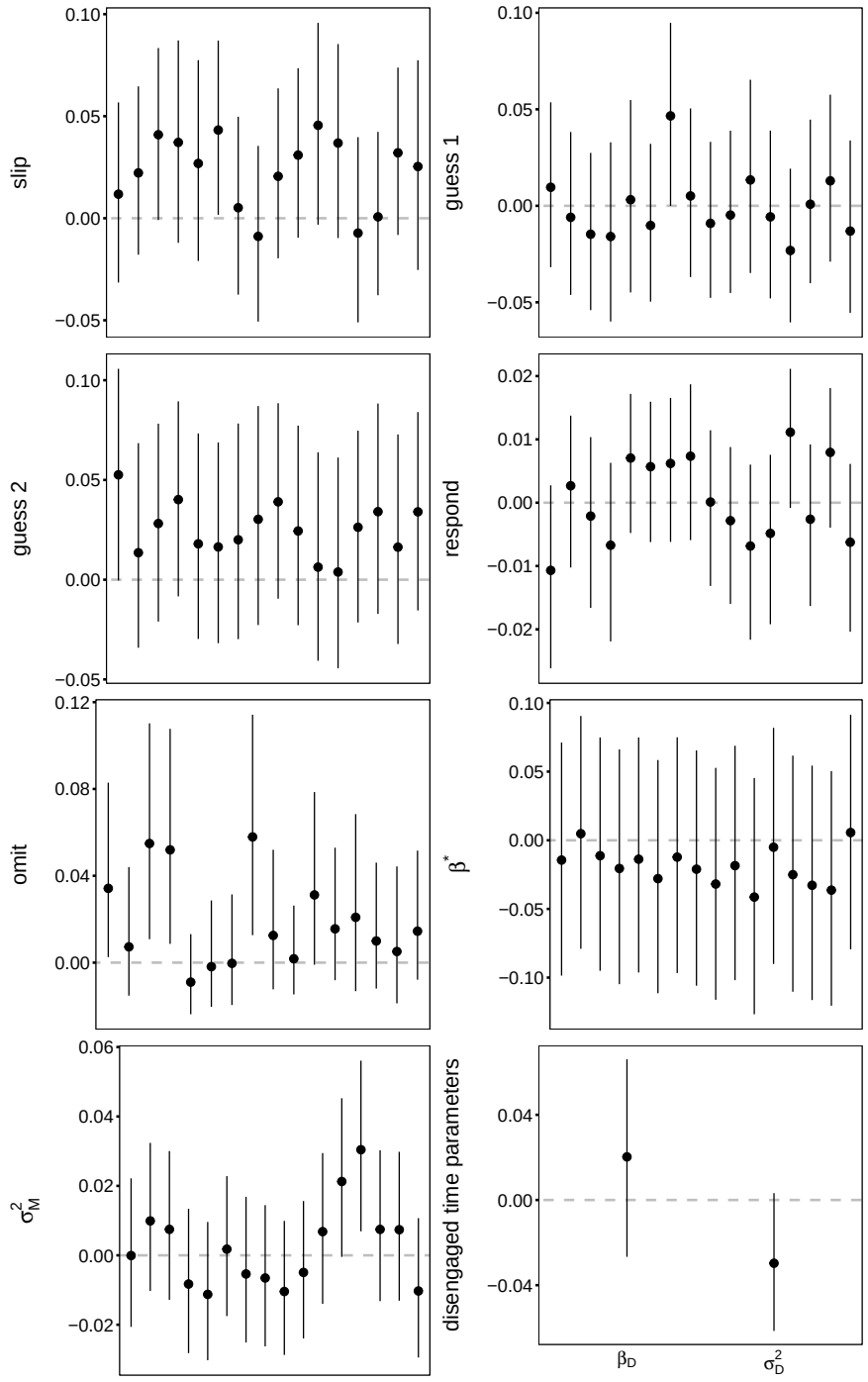


Figure 2: Discrepancy between estimated item parameters and true values for 16 items (1,000 respondents, 0.05 omission rate, 0.3 disengaged rate).

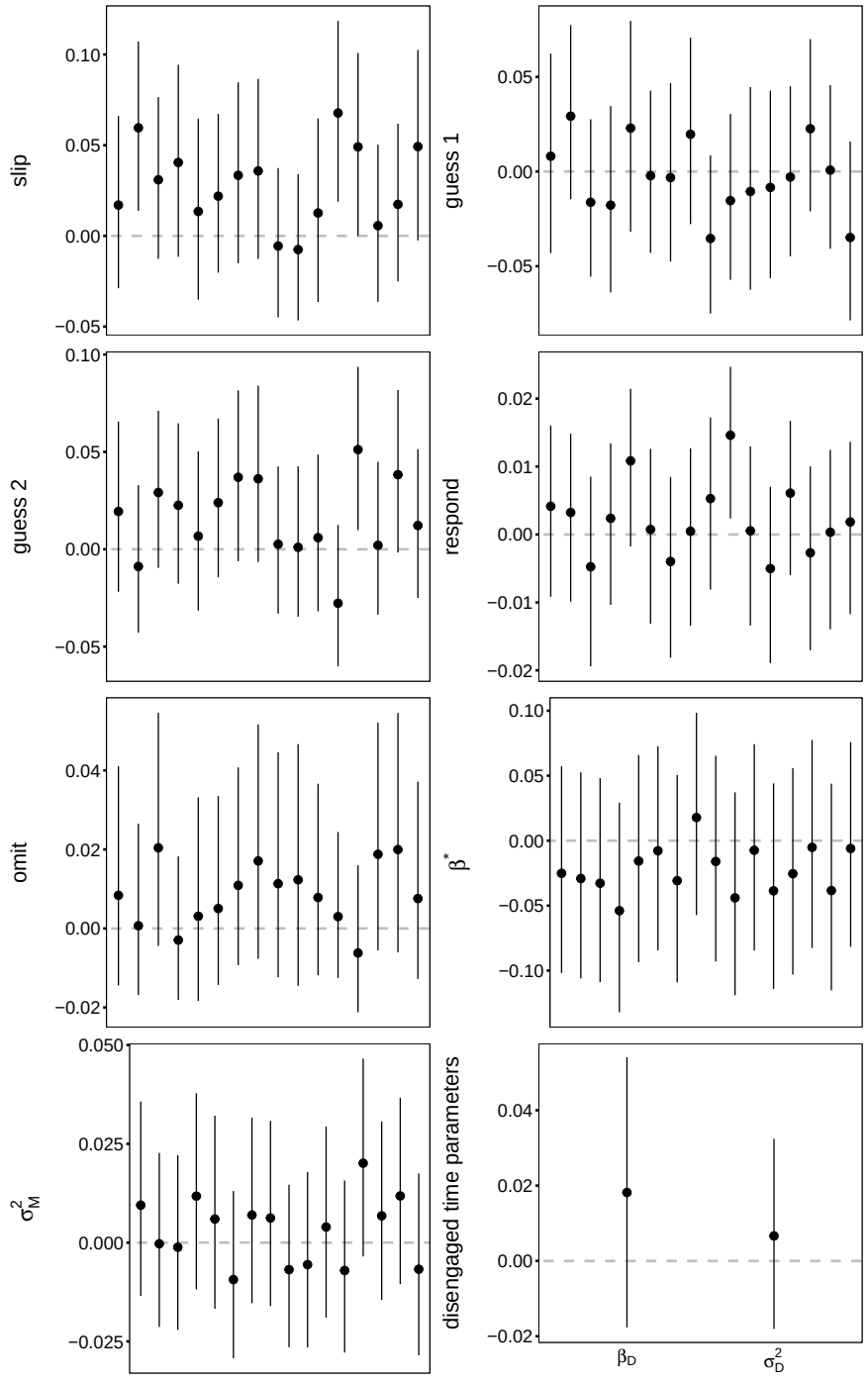


Figure 3: Discrepancy between estimated item parameters and true values for 16 items (1,000 respondents, 0.05 omission rate, 0.5 disengaged rate).

response time) and MCDM-O (modeling response accuracy and omission).

5.1 Dataset

Sixteen binary items were selected from M01 and M02 of the PISA 2015 Math assessment (OECD, 2017a, 2017b), measuring the same three attributes as the simulation study. The sample included 12,609 students who answered Form 43.

In the PISA 2015 dataset, the omissions are labeled as: 5 = Valid skip (the item was not required to be answered according to the test arrangement); 6 = Not reached (the student did not respond to the given item and subsequent items); 7 = Not applicable (the given item should be skipped or the answers could not be determined); 8 = Invalid (the answers exceeded the acceptable range); and 9 = No response (the student did not respond to the given item). Considering that not reached omissions were mainly caused by time constraints or device issues (Pohl, Ulitzsch, & von Davier, 2019), in this study, we only dealt with the label 9 to avoid the impact from compounding factors. The omission rates for each OECD country could be found in Appendix A, ranging from 0.77% to 10.78%.

Furthermore, based on the technical report (OECD, 2017b), response times greater than $4.4478 \times \text{MAD}$ were labeled as outliers, where $\text{MAD} = \text{med}_i\{|x_i - \text{med}_j(x_j)|\}$ and $\{x_i\}$ represents all sample values (Leys, Ley, Klein, Bernard, & Licata, 2013; Rousseeuw & Croux, 1993). The sample consisted of 8,180 students after the data cleaning procedure. We randomly selected 1,000 students for analysis and the omission rate was 4.46%. For models without omissions (DINA and MCDM-RT), we first removed all missing values and then randomly sampled 1,000 students. The proportion of correctness, omission rates, and the mean as well as standard deviation of response time between the full sample and the subsample can be found in Appendices B and C. Both two subsets can effectively represent the full data.

The HMC algorithm with No-U-Turn sampler was used for estimation. Four MCMC chains were set with 10,000 iterations each and the first 5,000 iterations were set as warm-

up.

5.2 Evaluation Criteria

As in the simulation study, the Gelman-Rubin $\hat{R}$ and ESSs were used to examine the model convergence. For model comparison, we computed the leave-one-out cross-validation information criterion (looic) to estimate the out-of-sample predictive accuracy for all models,

$$\text{looic} = -2 \times \sum_{s=1}^{S} \log p(y_s|y_{-s}), \tag{15}$$

where y_s represents the sth data point and y_{-s} represents the data without the sth data point. Given the negative operation in Equation 15, a small looic indicates a powerful prediction capacity. Furthermore, note that there are differences in data volume fitted by different models. For example, the data matrix size for the DINA model is 1,000 $\times$ 16, while the data matrix size for the full model is 3,000 $\times$ 16. Hence, we standardized looic to increase comparability. For instance, $\text{looic}_{\text{DINA}} = \text{looic}/1,000$ and $\text{looic}_{\text{MCDM}} = \text{looic}/3,000$.

For absolute goodness-of-fit, the posterior predictive checking was employed for each model. Specifically, we simulated 5,000 datasets with parameters generated from the posterior predictive distribution (i.e., each parameter was sampled from $5,000 \times 4$ estimates). Next, the resulting proportion of correct response, omission rate, and density distribution of response time were compared between the simulated dataset and the observed PISA dataset.

Based on the model comparison results, we proceeded with the optimal model selected for the PISA 2015 dataset and analyzed the parameter estimates from three levels. First, at the attribute level, the proportions of mastery, non-mastery and non-active status for each attribute were presented. Second, at the item level, we investigated the relationships between different parameters and examined the meaning of each parameter with the real data. Finally, at the person level, we scrutinized the response patterns for representative students to determine whether the new model can effectively detect disengaged responses

Table 6: Model fit and convergence indices.

	$\text{looic}_{\text{adjusted}}$	$\hat{R} < 1.1$	ESSs > 5,000
DINA	18.007	1	1
MCDM	17.413	1	0.928
MCDM-O	10.571	1	1
MCDM-RT	23.671	1	0.915

Note. $\text{looic}_{\text{adjusted}}$ is the model fit index after adjusting.

and to explore different types of problem-solving processes.

5.3 Results

Table 6 shows the model fit and convergence indices for all models. $\hat{R}$ for all models were close to 1, and more than 90% parameters had ESSs above 5,000, indicating that MCMC chains were well convergent.

Compared to DINA, MCDM and MCDM-O showed a better fit to the data, while MCDM-RT had worse model fit performance. This means that: (1) It is necessary to consider two types of missing responses; (2) if only rapid guessing is considered for disengaged responses detection, it may be hard to capture the full picture of the PISA 2015 data.

The posterior predictive checking results can be found in Figure 4. All three new models (MCDM, MCDM-O, and MCDM-RT) covered the actual data well, as evidenced by the median values of simulated data closely matching those of the real data in the boxplots. In comparison to DINA, the full model MCDM shows higher representativeness in predicting correct response proportions for some items (e.g., items 6 and 9). However, overall, there is no significant difference between the two models.

Combining the results of looic, we further compared MCDM and MCDM-O. For the common parts of the two models, namely the missing rate, a posterior predictive check was conducted. Figure 5 presents the results of the posterior predictive check for missing rate between MCDM and MCDM-O. It can be observed that MCDM better represented the real data compared to MCDM-O. Specifically, the actual missing rates for each item all fall within

the 25th-75th percentile of the simulated data set from MCDM, while MCDM-O tended to generate lower missing proportions for items with higher missing rates (e.g., items 12 and 13).

Additionally, we examined the absolute fit of MCDM and MCDM-RT by comparing the generated log response time with the actual log response time distribution in the PISA data. The results of the posterior predictive check are shown in Appendix. It can be observed that both MCDM and MCDM-RT effectively describe the distribution characteristics of response times for the majority of items. However, for a few items such as item 9 and item 12, there is still room for improvement in the fit of both models. Overall, the fit of MCDM is superior to MCDM-RT.

Taken together, we selected the MCDM full model for subsequent data analysis.

Attribute level. Figure 6 displays the estimated attribute profiles at three levels. It can be seen that the diagnosis of α_2 faced the highest error rate with a non-active proportion of 15.2%. Moreover, the examinee population's mastery levels of all attributes were lower than 50%.

This suggests that researchers should be cautious in the diagnosis of α_2. For instance, it cannot be simply concluded that the examinee population's mastery of α_2 is higher than that of α_1 and α_3. When examinees do not exert sufficient effort in answering items, their responses may not reflect their mastery of the knowledge points (Zhu et al., 2022). If all examinees' non-active proportion of α_2 are assumed to actually master this attribute, then the inference that "the population has the best mastery of α_2" would be reliable. However, if all non-active proportion of α_2 did not master this attribute, then the proportion of non-mastery for α_2 would reach 63.4% (higher than α_1 and α_3), leading to the opposite conclusion, namely "the population has the poorest mastery of α_2".

Item level. The estimated item parameters are shown in Figure 6. For parameters of the response accuracy model, it is clear that $1 - s_j$ is higher than g_{1j} and g_{2j}. In some items, g_{1j} is notably higher than g_{2j} (e.g., item 4), but in others, their values are close (e.g.,

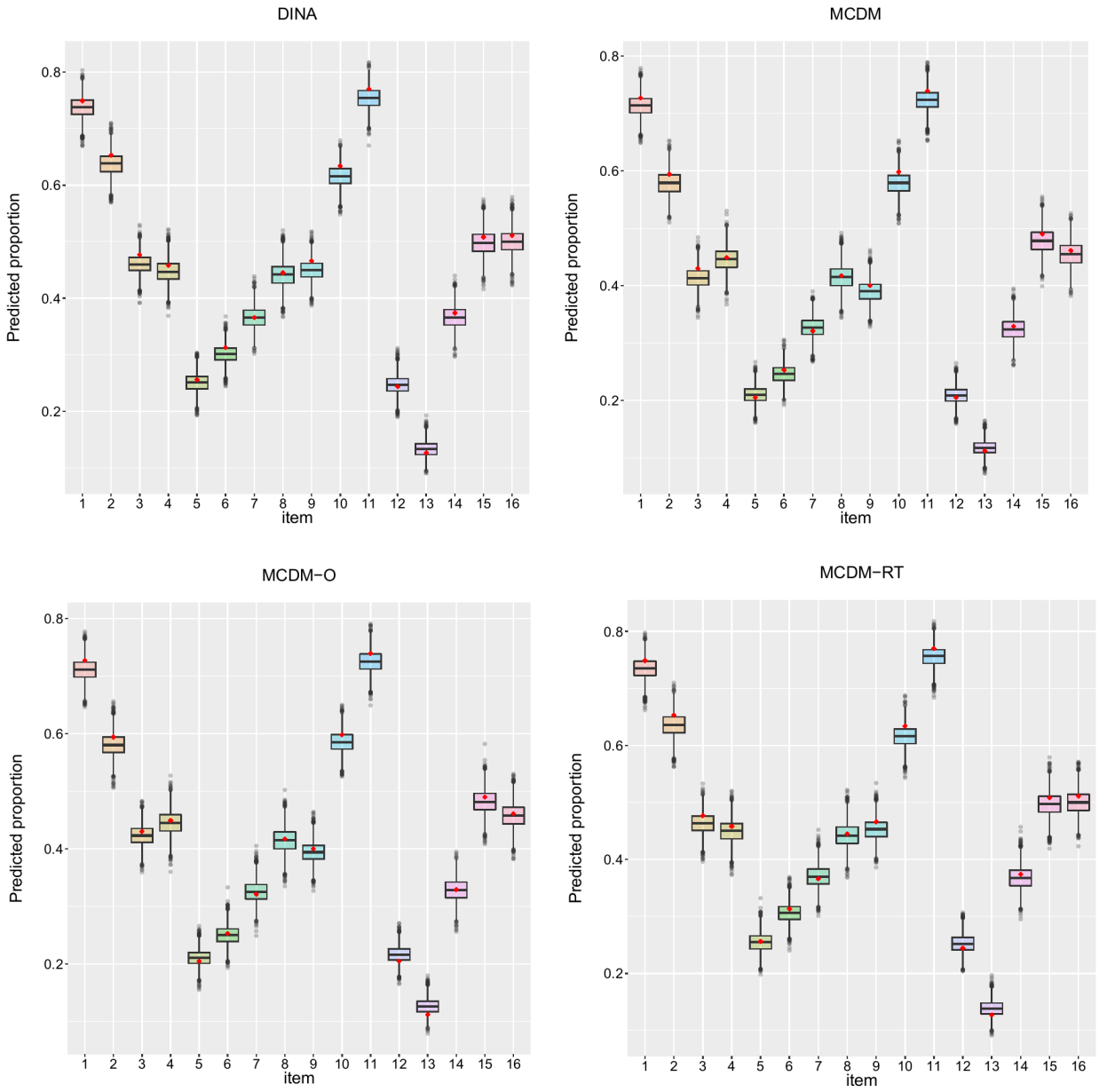


Figure 4: The boxplots of the model-based proportion of correct responses and the real proportion of correct responses (red diamond represents the actual proportion of correct responses calculated from the PISA data).

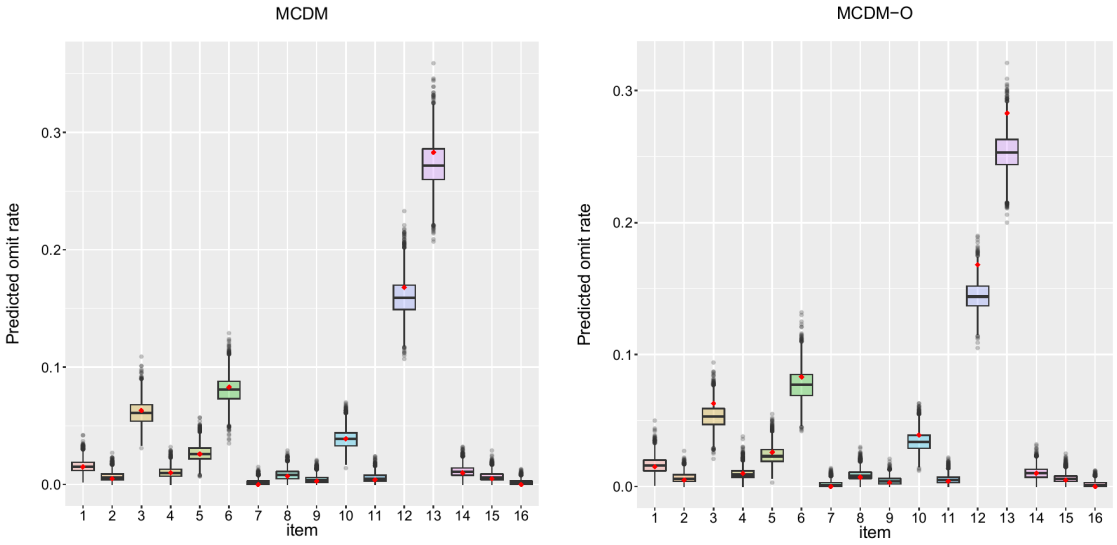


Figure 5: The boxplots of the model-based missing rates and the real missing rates (red diamond represents the actual missing rates calculated from the PISA data).

items 12 and 13). This indicates that when examinees master the required attributes, their probability of correct responses significantly increases. However, when examinees do not fully master the required attributes, the correct probabilities are similar to that of random guessing.

Additionally, it is worth noting that although there is a very small difference between g_{1j} and g_{2j} for some items (shown in the light red shaded areas in Figure 6), the time required for engaged behavior is higher than that for disengaged responses. This suggests that although the values of g_{1j} and g_{2j} are similar, they have distinct meanings with the former involving cognitive processing and the latter not. Correlation coefficients between the differences in item parameters and the wake-up parameter β_j^* are shown in Table 7. There is a significant negative correlation between the differences in g_{1j} and g_{2j} and β_j^*. Furthermore, the item easiness parameter can be calculated as $[g_{1j} + (1 - s_j)]/2$, and the correlation coefficient cor(easiness,β_j^*)=-0.575, with $p = 0.019 < 0.05$. This implies that as items become more difficult, more time is required to consider the items and fully engage the related attributes.

For the response time model, the variance of disengaged responses σ_D^2 is considerably

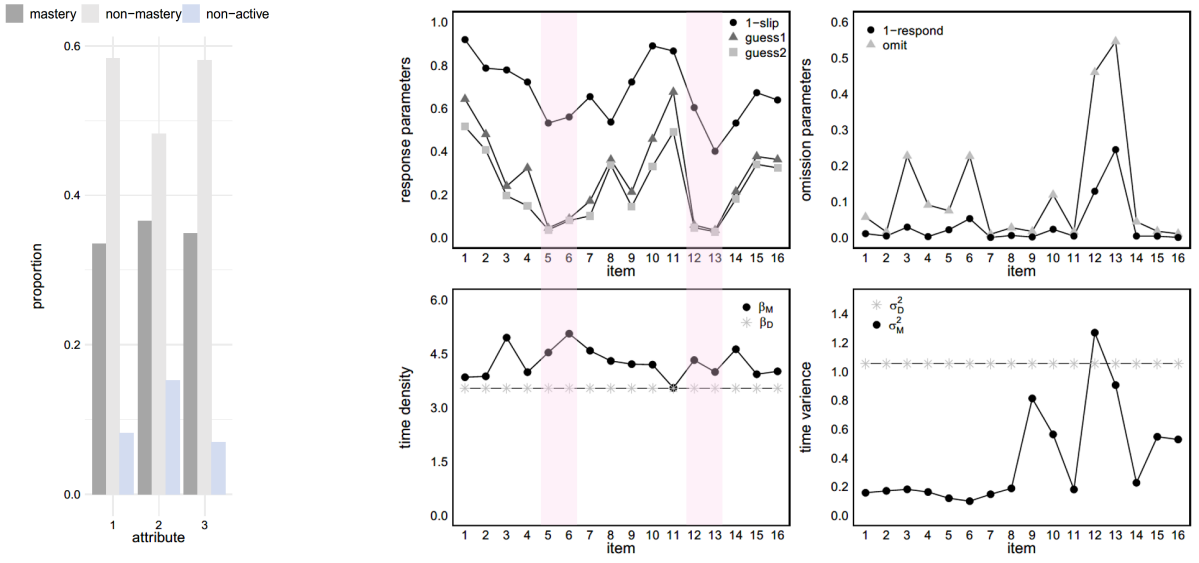


Figure 6: The estimated attribute profiles (left) and item parameters (right).

greater than that for normal responses σ_M^2 in almost all items (except for item 12). This result aligns with previous research findings (Bolsinova & Tijmstra, 2019; Domingue et al., 2022). Different examinees may exhibit different patterns of lack of effort, leading to more fluctuation in their response times compared to normal responses.

For the omission model, it can be observed that the probability of missing for engaged behavior $(1 - r_j)$ remains stable across different items, while o_j tended to exhibit more pronounced fluctuations. Further analysis of missing under normal or disengaged behaviors for each item (see Table 8) reveals that in items measuring α_1 and α_3, missing responses mainly happened in engaged responses, indicating that examinees are unable to answer these items even after investing effort, with only item 4 being an exception. Additionally, it is found that there is no significant relationship between question type (multiple-choice or open-ended) and different types of missing responses.

Examinee level. Defining examinees with $\alpha_{ik} = -1$ as the category of exhibiting disengaged responses during the test procedure, the disengaged rate of the PISA 2015 dataset is 17.3%. To determine whether MCDM can preciously capture disengaged responses, three representative examinees are presented. They are examinee 10559 with $\boldsymbol{\alpha} = (-1, -1, -1)$,

Table 7: The correlation between the differences of item parameters and the wake-up parameter.

	$\text{cor}(\text{diff}, \beta_j^*)$
$(1 - s_j, g_{1j})$	0.630**
$(1 - s_j, g_{2j})$	0.363
(g_{1j}, g_{2j})	-0.567*
$(1 - r_j, o_j)$	-0.296

$*$ indicates $p < 0.05$,
$**$ indicates $p < 0.01$;
cor = correlation, diff = the differences between two parameters.

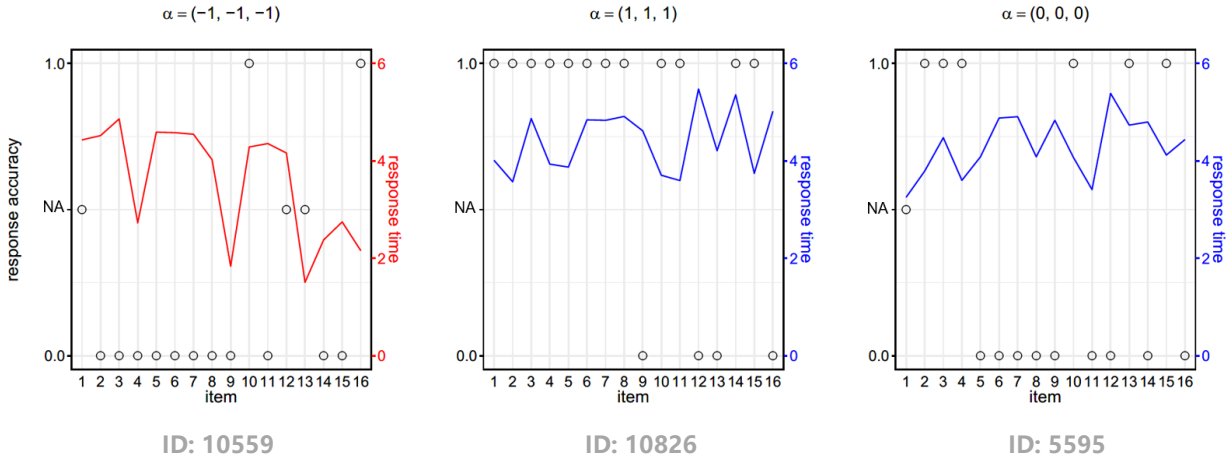


Figure 7: Response accuracy, response time, and omission behavior for representative respondents. *Note.* 0 = incorrect, 1 = correct, NA = missing; $\boldsymbol{\alpha} = (-1, -1, -1)$ represents disengaged response behavior, $\boldsymbol{\alpha} = (1, 1, 1)$ and $\boldsymbol{\alpha} = (0, 0, 0)$ represent normal behavior.

indicating disengaged responses; examinee 10826: $\boldsymbol{\alpha} = (1, 1, 1)$, indicating engaged responses and mastery of all attributes; examinee 5595: $\boldsymbol{\alpha} = (0, 0, 0)$, indicating engaged response but no mastery of any knowledge attribute.

Figure 7 presents the response patterns of the three examinees. It can be observed that compared to examinees with normal responses, those identified as having disengaged responses exhibit higher missing rates, lower correct proportions, and shorter and more fluctuating response times. The examinee took a long time to answer some items (e.g., 129 seconds for item 3), while answered others very quickly (e.g., 4 seconds for item 13).

Table 8: The omission under normal or disengaged behavior for each item.

attribute	item format	disengaged omission	engaged omission
α_1	MC	26.67	73.33
	OR	-	-
	MC	28.57	71.43
	OR	25.60	74.40
	OR	17.67	82.33
α_2	MC	40.00	60.00
	OR	63.49	36.51
	MC	46.99	53.01
	MC	66.67	33.33
	OR	48.72	51.28
	MC	70.00	30.00
	MC	40.00	60.00
α_3	MC	80.00	20.00
	OR	23.08	76.92
	MC	0.00	100.00
	MC	-	-

Note. MC = Multiple Choice; OR = Open Response.

6 Discussion

CDT aims at providing detailed feedback for students' knowledge profiles and is mainly applied in low-stakes scenarios. Given that students face no direct consequences from their test performance in low-stakes testing, disengaged behaviors are more common compared to high-stakes scenarios, which poses a substantial threat to data quality and decision validity (Hong, Steedle, & Cheng, 2020). To detect disengaged responses, this study proposed MCDM leveraging a multi-level attribute structure to link students' test-taking motivations with attribute profiles. A plug-and-play model framework was developed, allowing researchers to simultaneously consider rapid guessing and omissions, and to distinguish between omissions resulting from normal and disengaged behaviors.

To facilitate the empirical application of the new model, we offer several suggestions for the practitioners. First, MCDM can be flexibly adjusted for specific research purposes and data structures. In this study, MCDM is grounded on DINA, but it is not limited to any specific accuracy model. Therefore, researchers can establish corresponding motivation

models based on different CDMs such as DINO or G-DINA, depending on their research questions. Additionally, in our empirical study, based on the model fitting results with PISA 2015 data, the full model was ultimately chosen for parameter analysis. However, the two reduced models of MCDM (MCDM-RT and MCDM-O) also exhibit strong predictive power. Considering that the estimation time is directly related to the number of parameters, when researchers are primarily concerned with rapid guessing or different types of omissions, they can evaluate the model fitting of specific data and choose to use the reduced models to enhance estimation efficiency.

Second, the parameters obtained from MCDM can provide important insights into potential factors that cause disengaged behaviors. It would be helpful to understand why some attributes are more frequently involved in disengaged responses. Additionally, based on the results of the empirical study, for certain items, the proportion of missing responses attributable to engaged responses is considerable. While this study only analyzed the influence of item difficulty and item format on students' activation of knowledge, researchers can conduct detailed analyses with various item or person characteristics. For example, future exploration could integrate item content, item length, or students' background information to better understand factors contributing to disengaged responses, thereby enhancing the validity of diagnostic decisions and improving test design.

This study still has several limitations. First, a more efficient estimation algorithm would be beneficial. Future researchers could develop estimation procedures based on the Expectation-Maximization (EM) algorithm. This study implemented the HMC algorithm based on the Stan platform, which has improved the sampling efficiency to some extent. But when dealing with multi-level attributes, the possible attribute mastery/activation patterns can reach 3^K, which still slows down the estimation process. Another solution is developing estimation methods using reduced attributes. This study only examined scenarios involving three attributes in the test, resulting in $3^3 = 27$ possible attribute mastery patterns. If the test includes more number of attributes, addressing how to bypass the computation of 3^K

likelihood functions becomes a crucial issue. Using reduced attributes could be a potential option, where the ideal response can be represented as: $\eta_{ij} = \mathbb{I}(\boldsymbol{\alpha_{ij}}^* = 1) - \mathbb{I}(-1 \in \boldsymbol{\alpha_{ij}}^*)$.

Finally, researchers can incorporate richer data into the plug-and-play framework of MCDM. While this study has already allowed for simultaneously modeling rapid guessing and omission, identifying disengaged responses based on response accuracy, response time, and missing may still have room for improvement. It is still challenging to unleash students' problem-solving process. In the future, more detailed information such as eye-tracking data (Man & Harring, 2021) could be integrated into MCDM to precisely trace examinees' attention trajectories during the test and identify potential performances of low test-taking motivation.

References

Betancourt, M. (2017). *A conceptual introduction to hamiltonian monte carlo* [Unpublished Work]. arXiv preprint. Retrieved from https://arxiv.org/abs/1701.02434

Bolsinova, M., & Tijmstra, J. (2019). Modeling differences between response times of correct and incorrect responses. *Psychometrika*, *84*(4), 1018-1046. doi: 10.1007/s11336-019-09682-5

Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. *Applied Psychological Measurement*, *37*(6), 419-437. doi: 10.1177/0146621613479818

Chen, Y., Yang, Y., & Lee, Y. S. (2022). General cognitive diagnosis model for response time. Retrieved from https://academiccommons.columbia.edu/doi/10.7916/agge-zr74/download

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333-353. doi: 10.1007/BF02295640

de la Torre, J. (2011). The generalized dina model framework. *Psychometrika*, *76*(2), 179-199. doi: 10.1007/s11336-011-9207-7

Domingue, B. W., Kanopka, K., Stenhaug, B., Sulik, M. J., Beverly, T., Brinkhuis, M., ... Yeatman, J. (2022). Speed–accuracy trade-off? not so fast: Marginal changes in speed have inconsistent relationships with accuracy in real-world settings [Journal Article]. *Journal of Educational and Behavioral Statistics*, *47*(5), 576-602. doi: 10.3102/10769986221099906

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables [Journal Article]. *Psychometrika*, *74*(2), 191-210. doi: 10.1007/s11336-008-9089-5

Hong, M., Steedle, J. T., & Cheng, Y. (2020). Methods of detecting insufficient effort responding: Comparisons and practical recommendations [Journal Article]. *Educational and Psychological Measurement*, *80*(2), 312-345. Retrieved from https://

www.ncbi.nlm.nih.gov/pubmed/32158024 doi: 10.1177/0013164419865316

Hoyt, J. E. (2001). Performance funding in higher education: The effects of student motivation on the use of outcomes tests to measure institutional effectiveness. *Research in Higher Education*, *42*(1), 71-85. doi: 10.1023/A:1018716627932

Hsu, C. L., Jin, K. Y., & Chiu, M. M. (2020). Cognitive diagnostic models for random guessing behaviors. *Frontiers in Psychology*, *11*, 570365. doi: 10.3389/fpsyg.2020.570365

Jiang, Z., & Carter, R. (2019). Using hamiltonian monte carlo to estimate the log-linear cognitive diagnosis model via stan [Journal Article]. *Behavior Research Methods*, *51*(2), 651-662. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29949073 doi: 10.3758/s13428-018-1069-9

Junker, B. W., & Sijtsma, K. (2016). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory [Journal Article]. *Applied Psychological Measurement*, *25*(3), 258-272. doi: 10.1177/01466210122032064

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*, 764–766. doi: 10.1016/j.jesp.2013.03.013

Lu, J., Wang, C., & Shi, N. (2021). A mixture response time process model for aberrant behaviors and item nonresponses. *Multivariate Behavioral Research*, 1-19. doi: 10.1080/00273171.2021.1948815

Lu, J., Wang, C., & Shi, N. (2023). A mixture response time process model for aberrant behaviors and item nonresponses. *Multivariate Behavioral Research*, *58*(1), 71–89. doi: 10.1080/00273171.2021.1948815

Luo, J., De Carolis, L., Zeng, B., & Jeon, M. (2023). Bayesian estimation of latent space item response models with jags, stan, and nimble in r. *Psych*, *5*(2), 396–415.

Ma, W., & Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses

[Journal Article]. *British Journal of Mathematical and Statistical Psychology*, *69*(3), 253-275. doi: 10.1111/bmsp.12070

Man, K., & Harring, J. R. (2021). Assessing preknowledge cheating via innovative measures: A multiple-group analysis of jointly modeling item responses, response times, and visual fixation counts [Journal Article]. *Educational and Psychological Measurement*, *81*(3), 441-465. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/33994559 doi: 10.1177/0013164420968630

OECD. (2017a). *Pisa 2015 results (volume i): Excellence and equity in education* (Report). Retrieved from https://www.oecd.org/education/pisa-2015-results-volume-i-9789264266490-en.htm

OECD. (2017b). *Pisa 2015 technical report* (Report). Retrieved from https://www.oecd.org/pisa/data/2015-technical-report/PISA2015_TechRep_Final.pdf

OECD. (2019). *Pisa 2018 results (volume i): What students know and can do* (Report). Retrieved from https://www.oecd.org/pisa/publications/pisa-2018-results-volume-i-5f07c754-en.htm

Pohl, S., Ulitzsch, E., & von Davier, M. (2019). Using response times to model not-reached items due to time limits. *Psychometrika*, *84*(3), 892-920. doi: 10.1007/s11336-019-09669-2

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, *88*(424), 1273-1283. doi: 10.1080/01621459.1993.10476408

Schnipke, D. L., & Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness [Journal Article]. *Journal of Educational Measurement*, *34*(3), 213-232. doi: 10.1111/j.1745-3984.1997.tb00516.x

Su, S., & Davison, M. L. (2019). Improving the predictive validity of reading comprehension using response times of correct item responses [Journal Article]. *Applied Measurement in Education*, *32*(2), 166-182. doi: 10.1080/08957347.2019.1577247

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models [Journal Article]. *Psychological Methods*, *11*(3), 287-305. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/16953706 doi: 10.1037/1082-989X.11.3.287

Ulitzsch, E., Shin, H. J., & Lüdtke, O. (2023). Accounting for careless and insufficient effort responding in large-scale survey data—development, evaluation, and application of a screen-time-based weighting procedure. *Behavior Research Methods*, 1–22. doi: 10.3758/s13428-022-02053-6

Ulitzsch, E., von Davier, M., & Pohl, S. (2020). A hierarchical latent response model for inferences about examinee engagement in terms of guessing and item-level non-response. *British Journal of Mathematical and Statistical Psychology*, *73*, 83-112. doi: 10.1111/bmsp.12188

van der Linden, W. J. (2016). A lognormal model for response times on test items [Journal Article]. *Journal of Educational and Behavioral Statistics*, *31*(2), 181-204. doi: 10.3102/10769986031002181

von Davier, M., & Lee, Y.-S. (2019). *Handbook of diagnostic classification models: Models and model extensions, applications, software packages.* New York, NY: Springer. Retrieved from https://doi.org/10.1007/978-3-030-05584-4. Retrieved from https://doi.org/10.1007/978-3-030-05584-4 doi: 10.1007/978-3-030-05584-4

Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology*, *68*(3), 456-77. doi: 10.1111/bmsp.12054

Wang, S., & Chen, Y. (2020). Using response times and response accuracy to measure fluency within cognitive diagnosis models [Journal Article]. *Psychometrika*, *85*(3), 600-629. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/32816238 doi: 10.1007/s11336-020-09717-2

Wise, S. L., & Gao, L. (2017). A general approach to measuring test-taking effort on

computer-based tests. *Applied Measurement in Education*, *30*(4), 343-354. doi: 10 .1080/08957347.2017.1353992

Xu, G. (2019). Identifiability and cognitive diagnosis models. In M. von Davier & Y.-S. Lee (Eds.), *Handbook of diagnostic classification models. methodology of educational measurement and assessment* (p. 333-357). Springer, Cham. Retrieved from `https://link.springer.com/content/pdf/10.1007/978-3-030-05584-4.pdf?pdf=button\%20sticky` doi: 10.1007/978-3-030-05584-4_16

Zhu, Z., Arthur, D., & Chang, H. (2022). A new person-fit method based on machine learning in cdm in education. *British Journal of Mathematical and Statistical Psychology*, *75*(3), 616-637. doi: 10.1111/bmsp.12270

A Sample sizes and omission rates for OECD countries

OECD country	N	omission (%)	OECD country	N	omission (%)
ARE	187	3.54	KOR	147	2.72
AUS	375	4.78	LTU	169	5.29
AUT	182	5.60	LUX	137	4.38
BEL	228	4.14	LVA	134	3.45
BGR	153	5.64	MAC	111	1.46
BRA	390	6.84	MEX	171	2.96
CAN	541	2.82	MNE	127	10.78
CHE	200	4.16	NLD	135	3.19
CHL	170	7.06	NOR	151	5.30
COL	311	4.12	NZL	118	4.87
CRI	161	5.12	PER	138	5.93
CZE	170	5.74	POL	173	3.40
DEU	164	4.08	PRT	182	5.56
DNK	175	4.43	QAT	422	5.18
DOM	140	8.84	QCH	281	1.87
ESP	165	5.64	QES	819	5.20
EST	143	3.23	QUC	46	1.90
FIN	162	3.55	QUE	49	0.77
FRA	159	5.31	RUS	137	4.01
GBR	366	4.20	SGP	160	1.76
GRC	150	4.00	SVK	151	3.85
HKG	140	1.43	SVN	155	4.60
HRV	141	5.50	SWE	124	5.34
HUN	141	6.07	TAP	201	1.68
IRL	211	2.75	THA	214	2.31
ISL	90	3.13	TUN	102	8.58
ISR	171	5.48	TUR	160	5.08
ITA	313	5.31	URY	134	10.21
JPN	177	2.93	USA	144	1.87

B Data characteristics of the full sample and the subset data

Item	The proportion of correct responses		omission rates		Log response times (second)			
	full	subset	full	subset	full (Mean)	subset (Mean)	full (SD)	subset (SD)
1	0.731	0.727	0.010	0.015	3.741	3.744	0.461	0.472
2	0.602	0.594	0.005	0.005	3.799	3.778	0.508	0.509
3	0.425	0.430	0.056	0.063	4.664	4.656	0.842	0.837
4	0.434	0.449	0.014	0.010	3.796	3.778	0.687	0.672
5	0.211	0.205	0.032	0.026	4.387	4.374	0.523	0.563
6	0.250	0.253	0.094	0.083	4.818	4.807	0.635	0.658
7	0.326	0.321	-	-	4.432	4.407	0.558	0.594
8	0.419	0.417	0.007	0.007	4.118	4.127	0.622	0.640
9	0.419	0.400	0.002	0.003	4.052	4.007	0.955	1.006
10	0.595	0.598	0.035	0.039	4.048	4.059	0.835	0.834
11	0.738	0.739	0.005	0.004	3.411	3.401	0.521	0.554
12	0.187	0.205	0.169	0.168	4.132	4.127	1.144	1.170
13	0.090	0.112	0.286	0.283	3.778	3.761	1.040	1.068
14	0.331	0.329	0.011	0.010	4.383	4.387	0.747	0.746
15	0.463	0.490	0.005	0.005	3.790	3.757	0.876	0.870
16	0.442	0.461	-	-	3.867	3.815	0.858	0.873

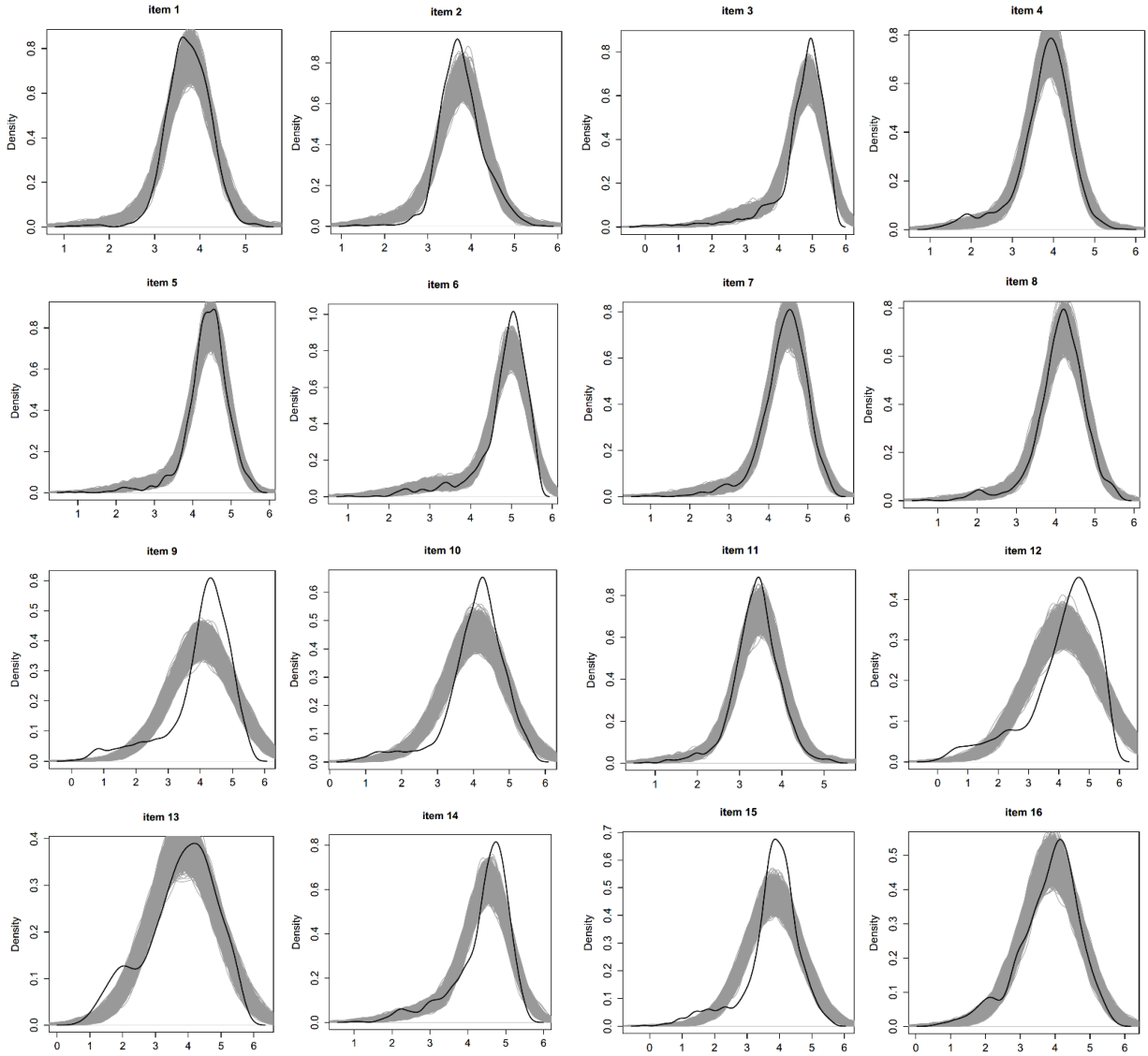
Note. Item 7 (CM034Q01S) and item 16 (CM564Q02S) are the last question of M01 and M02 respectively.

C Data characteristics of the full sample and the subset data (without omissions)

Item	The proportion of correct responses		Log response times (second)			
	full	subset	full (Mean)	subset (Mean)	full (SD)	subset (SD)
1	0.758	0.749	3.725	3.728	0.445	0.439
2	0.648	0.653	3.774	3.774	0.499	0.498
3	0.485	0.477	4.681	4.671	0.810	0.845
4	0.478	0.458	3.826	3.791	0.640	0.636
5	0.261	0.256	4.386	4.389	0.493	0.503
6	0.308	0.313	4.917	4.891	0.498	0.538
7	0.372	0.366	4.451	4.430	0.532	0.570
8	0.445	0.445	4.157	4.139	0.603	0.634
9	0.484	0.466	4.076	4.031	0.939	0.944
10	0.650	0.634	4.033	3.999	0.803	0.774
11	0.766	0.770	3.382	3.351	0.506	0.510
12	0.261	0.244	4.184	4.186	1.224	1.207
13	0.141	0.127	4.138	4.154	0.860	0.851
14	0.370	0.374	4.435	4.390	0.700	0.762
15	0.501	0.508	3.785	3.795	0.872	0.844
16	0.487	0.511	3.924	3.913	0.832	0.845

D The posterior predictive check for log response time

D.1 The model-based log response time (black line) and actual log response time distribution (gray line) of the MCDM full model



D.2 The model-based log response time (black line) and actual log response time distribution (gray line) of the reduced MCDM-RT model