Behaviormetrika (2020) 47:19–41
https://doi.org/10.1007/s41237-019-00100-9

ORIGINAL PAPER

Cognitive diagnosis models for estimation of misconceptions analyzing multiple-choice data

Koken Ozaki[1] · Shingo Sugawara[2] · Noriko Arai[2]

Received: 18 December 2018 / Accepted: 29 October 2019 / Published online: 14 November 2019

Abstract
Incorrect options for multiple-choice questions are often intentionally included so that they may be selected by an examinee who possesses a misconception. Determining whether an examinee possess a misconception is useful for educational purposes. In the present paper, two statistical models that can estimate examinees' possession of misconceptions by analyzing multiple-choice data, which are unscored data were developed. By converting multiple-choice data to binary data, which are scored data (1 = correct, 0 = incorrect), the Bug-DINO model can estimate examinees' possession of misconceptions. However, converting multiple-choice data to binary data causes a loss in information, because which incorrect option an examinee chooses is important information for an examinee's knowledge state. The three models (two developed models and the Bug-DINO model) are compared in a simulation study, and the developed models are applied to the Reading Skill Test data.

Keywords Multiple-choice item · Cognitive diagnosis model · Misconception · DINO model

1 Introduction

To date, a number of cognitive diagnosis models (CDMs) have been developed, including the deterministic input noisy output "AND" gate (DINA; Junker and Sijtsma 2001) model, the reduced reparameterized unified model (R-RUM; Hartz

Communicated by Russell George Almond.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s41237-019-00100-9) contains supplementary material, which is available to authorized users.

✉ Koken Ozaki
koken298@gmail.com

[1] Graduate School of Business Sciences, University of Tsukuba, Tokyo, Japan

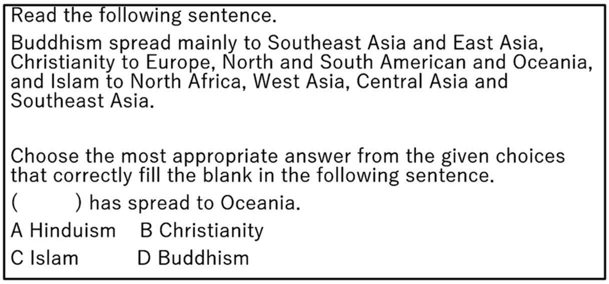
[2] National Institute of Informatics, Tokyo, Japan

2002), the deterministic input noisy output “OR” gate (DINO; Templin and Henson 2006) model, the noisy input deterministic output “AND” gate model (NIDA; Maris 1999), the additive CDM (A-CDM; de la Torre 2011), the linear logistic model (LLM; de la Torre and Douglas 2004), and the CDM for continuous response (Minchen et al. 2017). These models estimate examinees’ mastery or non-mastery of skills (referred to as “attributes”) that are needed to answer items correctly.

A Q-matrix plays an important role in CDMs. The number of rows of the matrix is the number of items in a test, and the number of columns is the total number of skills needed to correctly answer the items in a test. The elements of a Q-matrix (q_{jk}) are 0 or 1. If the element is 1, item j needs skill k to answer correctly, otherwise it is 0. The observed variable in usual CDMs is Y_{ij}, which is 1 if examinee i correctly answers item j, otherwise Y_{ij} is 0. Cognitive diagnosis models use a Q-matrix and analyze Y_{ij} and then output whether examinee i has each skill, by 1 or 0.

Recent studies (DiBello et al. 2015; Kuo et al. 2016, 2018) have focused on not only skills but also misconceptions, which means that examinees have acquired incorrect knowledge. For example, consider the item shown in Fig. 1. This item examines the skill of recognizing the dependency relations between words and phrases in a given sentence (Arai et al. 2017). This item is one of the example items in the Reading Skill Test in Japan (https://www.s4e.jp/). When two clauses are related to each other in one sentence to create the meaning of a sentence, such as the relation between a subject and a predicate, and a modifier and a modified word, the preceding clause is related to the latter. The item shown in Fig. 1 examines whether an examinee has the skill of recognizing such relations. More specifically, this item asks whether an examinee correctly recognizes the relations between “Christianity” and “Oceania”.

The correct option is B (Christianity). If an examinee selects C (Islam), he/she may possess the misconception that the nearest word is the subject, which means that the examinee misunderstands the relations between words such that the subject is the nearest word in the sentence to the word being asked about the relation (in this case Oceania). Moreover, if an examinee selects D (Buddhism), he/she may possess the misconception that the word at the beginning of a sentence is the subject. Since Japanese sentences usually start with a subject, there may be examinees who think that the word at the beginning of a sentence is always the subject



Read the following sentence.

Buddhism spread mainly to Southeast Asia and East Asia, Christianity to Europe, North and South American and Oceania, and Islam to North Africa, West Asia, Central Asia and Southeast Asia.

Choose the most appropriate answer from the given choices that correctly fill the blank in the following sentence.

(   ) has spread to Oceania.

A Hinduism B Christianity

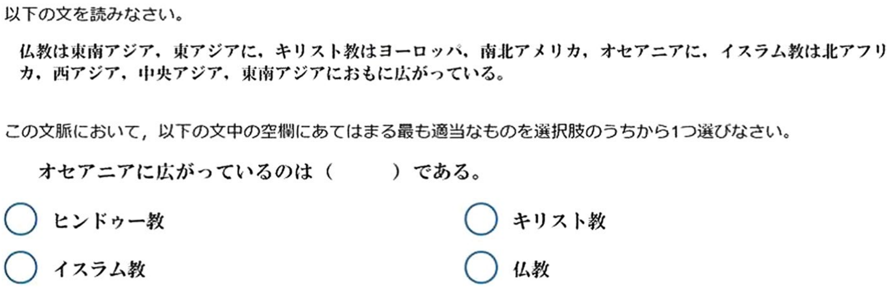
C Islam D Buddhism

Fig. 1 Example of a dependency item

of any sentence. This is a type of error analysis (Richards and Schmidt 2002) that examines the types and causes of errors on test items. Error analysis is very useful for teachers because the results of the analysis reveal why each student is stumbling in learning. Actually, as shown in Fig. 2, the item content is written in Japanese. However, Buddhism is the beginning of the sentence, and Islam is the nearest word to Oceania in the Japanese version as well. Note that if an examinee selects A (Hinduism), because “Hinduism” does not appear in the text, such an examinee appears to be not seriously working on the item and might select the option by chance.

Misconceptions have also been analyzed by CDMs. For example, Kuo et al. (2018) presented a fraction multiplication test in which students were required to write down their problem-solving process. This test measures four skills and three misconceptions, namely, turning the second fraction upside down when multiplying a fraction by a fraction, solving only the first step of a two-step problem, and performing incorrect arithmetic operations when confused about the relational terms. Moreover, DiBello et al. (2015) presented the Diagnostic Geometry Assessment for Geometric Measurement, which measures two facets using multiple-choice items: (a) a conceptual understanding of area measure and (b) a problematic facet of thinking about an area, which corresponds to a misconception.

In the present paper, two statistical models that can estimate examinees’ possession of misconceptions by analyzing multiple-choice data, which are unscored data, were developed based on the multiple-choice DINA models developed by Ozaki (2015). The Bug-DINO model can estimate examinees’ possession of misconceptions as well (Kuo et al. 2016, 2018) by converting multiple-choice data to binary data, which are scored data (1 = correct, 0 = incorrect). However, converting multiple-choice data to binary data causes a loss in information, because which incorrect option an examinee chooses is important information for an examinee’s knowledge state. The developed model is expected to tell why each student is stumbling with greater accuracy than Bug-DINO. The three models will be compared in a simulation study, and the developed models will be applied to the Reading Skill Test (Arai et al. 2017) data.


以下の文を読みなさい。

仏教は東南アジア，東アジアに，キリスト教はヨーロッパ，南北アメリカ，オセアニアに，イスラム教は北アフリカ，西アジア，中央アジア，東南アジアにおもに広がっている。

この文脈において，以下の文中の空欄にあてはまる最も適当なものを選択肢のうちから1つ選びなさい。

オセアニアに広がっているのは（　　　）である。

○ ヒンドゥー教　○ キリスト教

○ イスラム教　○ 仏教

Fig. 2 Example of a dependency item (Japanese version)

1.1 The DINO and Bug-DINO models

In all of the discussed models, let j be an item, i be an examinee, and $\boldsymbol{\alpha}_i^*$ be the knowledge state vector of examinee i. The kth element of $\boldsymbol{\alpha}_i^*$ is expressed as α_{ik}^*, which is 1 when examinee i possesses a skill $k = (1, 2, \dots, K)$ and is 0 otherwise.

The probability that examinee i correctly answers item j ($Y_{ij} = 1$) is expressed in the DINO model as follows:

$$P(Y_{ij} = 1|\boldsymbol{\alpha}_i^*) = (1 - s_j^*)^{w_{ij}} g_j^{*\,1-w_{ij}}, \tag{1}$$

where

$$w_{ij} = 1 - \prod_{k=1}^{K} (1 - \alpha_{ik}^*)^{q_{jk}}. \tag{2}$$

In this case, $\alpha_{ik}^* = 1$ means that examinee i has a skill k. Here, s_j^* is the slip parameter of item j in the DINO model, which is the probability that an examinee who has at least one of the skills needed to answer the item correctly fails to answer the item correctly, and g_j^* is the guessing parameter for item j in the DINO model, which is the probability that an examinee who has none of the skills needed to correctly answer the item nevertheless does correctly answer the item. Then, q_{jk} is the element of the Q-matrix for item j. If skill k is needed to answer item j, then $q_{jk} = 1$, and otherwise $q_{jk} = 0$.

In order to apply the DINO model to multiple-choice items and estimate examinees' misconceptions, multiple-choice data must be converted to binary data. If an examinee has selected one of the options coded by misconceptions, $Y_{ij} = 1$, otherwise $Y_{ij} = 0$. For example, for the item shown in Fig. 1, if an examinee has selected either C (Islam) or D (Buddhism), then $Y_{ij} = 1$ because they are coded by misconceptions, and if an examinee has selected either A (Hinduism) or B (Christianity), then $Y_{ij} = 0$.

Replace $\boldsymbol{\alpha}_i^*$ with $\boldsymbol{\alpha}_i$, α_{ik}^* with α_{ik}, s_j^* with s_j, and g_j^* with g_j. Here, $\boldsymbol{\alpha}_i$ is the misconception knowledge state vector of examinee i. The kth element of $\boldsymbol{\alpha}_i$ is expressed as α_{ik}, which is 1 when examinee i possesses misconception $k(= 1, 2, \dots, K)$ and is 0 otherwise. Then, s_j is the slip parameter of item j, which is the probability that an examinee who possesses at least one of the misconceptions that lead to the selection of one of the options coded by misconceptions fails to select one of these options, and g_j is the guessing parameter for item j, which is the probability that an examinee who possesses none of the misconceptions that lead to the selection of one of the options coded by misconceptions does select one of these options. Kuo et al. (2016) referred to the model as Bug-DINO. In the present paper, the model shown in Eq. (1) replacing $\boldsymbol{\alpha}_i^*$ with $\boldsymbol{\alpha}_i$, α_{ik}^* with α_{ik}, s_j^* with s_j, and g_j^* with g_j is referred to as the Bug-DINO model.

However, as noted previously, converting multiple-choice data into binary data loses information about the examinee's possession of misconceptions. Therefore, a statistical model is needed that estimates examinee's misconceptions by analyzing multiple-choice data.

1.2 Multiple-choice DINA models

Table 1 shows an example of multiple-choice items with four options, two of which are coded by misconceptions α_1 and α_2. Options 1, 2, and 3 are incorrect, and option 4 is correct. Note that options 3 and 4 (the correct option) are both coded by none of the misconceptions. Option 1 is coded by both α_1 and α_2, which means that an examinee who has both α_1 and α_2 is likely to select this option. Option 2 is coded by α_1, which means that an examinee who has only α_1 is likely to select this option.

Since incorrect options are coded by two misconceptions (α_1 and α_2), there are $2^2 = 4$ patterns of the misconception knowledge state, which are expressed as $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\alpha}_3$, and $\boldsymbol{\alpha}_4$. Therefore, $P(1|\boldsymbol{\alpha}_1) = p(Y_{ij} = 1|\boldsymbol{\alpha}_1)$ is the probability that an examinee with $\boldsymbol{\alpha}_1$ selects the first option, and $P(1|\boldsymbol{\alpha}_2)$ is the probability that an examinee with $\boldsymbol{\alpha}_2$ selects the first option. Here, $\boldsymbol{\alpha}_1 = (\alpha_1, \alpha_2) = (1, 1)$, $\boldsymbol{\alpha}_2 = (\alpha_1, \alpha_2) = (1, 0)$, $\boldsymbol{\alpha}_3 = (\alpha_1, \alpha_2) = (0, 1)$, and $\boldsymbol{\alpha}_4 = (\alpha_1, \alpha_2) = (0, 0)$.

In order to analyze multiple-choice data, the incorrect options of which are coded by misconceptions, and to estimate $\boldsymbol{\alpha}_i$, a statistical model that provides selection probabilities $P(Y_{ij} = k|\boldsymbol{\alpha}_i)$ is needed. This is the probability that examinee i, the misconception knowledge state of which is $\boldsymbol{\alpha}_i$, selects option k of item j. de la Torre (2009), DiBello et al. (2015), and Ozaki (2015) proposed models that can analyze multiple-choice data in the framework of CDMs. Among these three models, only the model proposed by DiBello et al. (2015) is a model for both skills and misconceptions, whereas the models proposed by de la Torre (2009) and Ozaki (2015) are models for skills. The model of the present paper is based on the model proposed by Ozaki (2015).

The reason why the model of the present paper is based on Ozaki (2015) is as follows. One of the features of de la Torre's (2009) model is that it uses multiple-choice items that can distinguish examinees into one of the $C_j^* + 1$ groups by means of the coded options. Here, C_j^* is the number of options of item j coded with different attribute patterns. The term "group" here has the same meaning as a class of knowledge state. Therefore, in Table 1, if option 1 is coded by the second misconception as (0, 1), i.e., an examinee who possesses only the second misconception attribute is likely to select this option, then the item cannot be analyzed using de la Torre

Table 1 Selection probabilities for an item with four options

	α_1	1	1	0	0
	α_2	1	0	1	0
Option 1	(1, 1)	$P(1\|\boldsymbol{\alpha}_1)$	$P(1\|\boldsymbol{\alpha}_2)$	$P(1\|\boldsymbol{\alpha}_3)$	$P(1\|\boldsymbol{\alpha}_4)$
Option 2	(1, 0)	$P(2\|\boldsymbol{\alpha}_1)$	$P(2\|\boldsymbol{\alpha}_2)$	$P(2\|\boldsymbol{\alpha}_3)$	$P(2\|\boldsymbol{\alpha}_4)$
Option 3	(0, 0)	$P(3\|\boldsymbol{\alpha}_1)$	$P(3\|\boldsymbol{\alpha}_2)$	$P(3\|\boldsymbol{\alpha}_3)$	$P(3\|\boldsymbol{\alpha}_4)$
Option 4	(0, 0)	$P(4\|\boldsymbol{\alpha}_1)$	$P(4\|\boldsymbol{\alpha}_2)$	$P(4\|\boldsymbol{\alpha}_3)$	$P(4\|\boldsymbol{\alpha}_4)$

α_k is misconception k, and $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_4$ are misconception knowledge states, e.g., $\boldsymbol{\alpha}_1 = (\alpha_1, \alpha_2) = (1, 1)$

(2009), because examinees who have both α_1 and α_2 can be classified into both groups, namely, the (1, 0) and (0, 1) groups. The item shown in Fig. 1 is of this type. On the other hand, Ozaki (2015) overcame this limitation. DiBello et al. (2015) developed a generalized diagnostic classification model for multiple-choice data that can estimate examinees' possession of skills and misconceptions at the same time. However, the model requires a large number of parameters, because the model imposes a cognitive diagnostic model for each $P(Y_{ij} = k|\boldsymbol{\alpha}_i)$. On the other hand, Ozaki (2015) requires few parameters and therefore is parsimonious. For example, Ozaki's (2015) first model requires one item parameter for each item. Therefore, in the present paper, a model is developed based on Ozaki (2015).

Ozaki's (2015) first model (referred to as MC-S-DINA1) is as follows:

$$P(Y_{ij} = c|\boldsymbol{\alpha}_i^*) = \gamma_{ij}^*(1-\delta_j^*)^{\eta_{ijc}}{\frac{\delta_j^*}{C_j-1}}^{1-\eta_{ijc}} + \frac{(1-\gamma_{ij}^*)}{C_j}. \tag{3}$$

Here,

$$\eta_{ijc} = \prod_{k=1}^{K_j}(2 - 2^{(\alpha_{ik}^* - q_{jkc})^2}), \tag{4}$$

$$\gamma_{ij}^* = \sum_{c=1}^{C_j}\eta_{ijc}\left(1 - \prod_{k=1}^{K_j}(1-\alpha_{ik}^*)\right), \tag{5}$$

where C_j is the number of options of item j, c is the option, $\boldsymbol{\alpha}_i$ is the knowledge state of examinee i, K_j is the number of required attributes for item j, and the attributes are ordered such that the required attributes for item j are the first K_j attributes (de la Torre 2011). In other words, the last $K - K_j$ attributes are the attributes that are not needed to answer item j correctly. In addition, $q_{jkc} = 1$ indicates that attribute k is needed in order to select option c of item j. If attribute k is not needed, then $q_{jkc} = 0$. In Eq. (4), $\eta_{ijc} = 1$ when examinee i has exactly the attributes needed for option c of item j, and $\eta_{ijc} = 0$ otherwise. In other words, $\eta_{ijc} = 1$ when examinee i's knowledge state perfectly matches the attribute vector of option c of item j.

In Eq. (5), $\gamma_{ij}^* = 1$ when $1 - \prod_{k=1}^{K_j}(1-\alpha_{ik}^*) = 1$ (examinee i has at least one attribute needed to correctly answer item j) and $\sum_{c=1}^{C}\eta_{ijc} = 1$, and otherwise $\gamma_{ij}^* = 0$. In MC-S-DINA1, when $\gamma_{ij}^* = 1$, the first term determines the selection probability, and when $\gamma_{ij}^* = 0$, the second term $(1-\gamma_{ij}^*)/C_j$ determines the selection probability. Here, $(1-\gamma_{ij}^*)/C_j$ is the guessing part of the model for an examinee with $\gamma_{ij} = 0$, which means that examinee i's knowledge state vector matches none of the attribute vectors of options of item j.

Then, δ_j^* is the probability that examinee i with $\boldsymbol{\alpha}_i^*$ selects an option other than the most likely option, given his/her knowledge state. Therefore, $1 - \delta_j^*$ is the probability that an examinee whose $\boldsymbol{\alpha}_i^*$ matches a required attribute pattern for an option of item j actually selects that option.

Table 2 Selection probabilities for MC-S-DINA1

	α_1^*	1	1	0	0
	α_2^*	1	0	1	0
Option 1	(1, 1)	0.85	0.05	0.25	0.25
Option 2	(1, 0)	0.05	0.85	0.25	0.25
Option 3	(0, 0)	0.05	0.05	0.25	0.25
Option 4	(0, 0)	0.05	0.05	0.25	0.25

α_1^* and α_2^* are skills

Table 3 Selection probabilities for MC-S-DINA2

	α_1^*	1	1	0	0
	α_2^*	1	0	1	0
Option 1	(1, 1)	0.85	0.10	0.25	0.25
Option 2	(1, 0)	0.05	0.70	0.25	0.25
Option 3	(0, 0)	0.05	0.10	0.25	0.25
Option 4	(0, 0)	0.05	0.10	0.25	0.25

α_1^* and α_2^* are skills

The selection probabilities of MC-S-DINA1 are shown in Table 2 for $\delta_j^* = 0.15$. As the table shows, an examinee who has both of the attributes selects options 1 through 4 with probabilities 0.85, 0.05, 0.05, and 0.05, respectively.

The second model (referred to as MC-S-DINA2; Ozaki 2015) relaxed the constraint on MC-S-DINA1 whereby δ_j^* is the same for all options of item j. The MC-S-DINA 2 model is then expressed as follows:

$$P(Y_{ijc} = 1|\boldsymbol{\alpha}_i^*) = \gamma_{ij}^*(1 - \delta_{jc}^*)^{\eta_{ijc}} \left(\frac{\beta_{ij}^*}{C_j - 1} \right)^{1-\eta_{ijc}} + \frac{(1 - \gamma_{ij}^*)}{C_j}. \tag{6}$$

where

$$\beta_{ij}^* = \sum_{c=1}^{C_j} \delta_{jc}^* \eta_{ijc}. \tag{7}$$

In MC-S-DINA2, δ_{jc}^* takes different parameters according to the attribute vector of option c. Then, β_{ij}^* is the probability that examinee i with α_i^* selects an option other than the option for which the attribute pattern matches α_i^*. The selection probabilities of MC-S-DINA2 are shown in Table 3 for $\delta_1^* = 0.15$ and $\delta_2^* = 0.3$.

In both MC-S-DINA1 and MC-S-DINA2, the selection probabilities for options other than the option that examinee i is most likely to select are the same for each $\boldsymbol{\alpha}_i^*$, as shown in Tables 2 and 3. In MC-S-DINA3, this restriction is relaxed by setting the selection probabilities according to the closeness between $\boldsymbol{\alpha}_i^*$ and the attribute vector for option c of item j. Closeness is measured by r_{ijc}^*, which is the inner product of $\boldsymbol{\alpha}_i^*$ and the attribute vector for option c of item j. If $r_{ijc}^* = \boldsymbol{\alpha}_i^* \boldsymbol{q}'_{jc}$ is large, examinee i is expected to select option c of item j with high probability. On the other hand, if r_{ijc}^* is small, examinee i is not expected to select option c of item j.

The MC-S-DINA3 model (Ozaki 2015) is expressed as follows:

$$P(Y_{ijc} = 1|\boldsymbol{\alpha}_i^*) = \left(1 - \prod_{k=1}^{K_j}(1 - \alpha_{ik}^*)\right)(1 - \delta_{jc}^*)^{\eta_{ijc}} \left(\omega_{ijc}^*\right)^{1-\eta_{ijc}} + \frac{\prod_{k=1}^{K_j}(1 - \alpha_{ik}^*)}{C_j}, \quad (8)$$

where ω_{ijc}^* is

$$\omega_{ijc}^* = \begin{cases} \beta_{ij}^* \frac{(1-\eta_{ijc})(1+r_{ijc}^*)}{\sum_{c=1}^{C}(1-\eta_{ijc})(1+r_{ijc}^*)} & \text{if } \gamma_{ij}^* = 1 \\ \frac{(1-\eta_{ijc})(1+r_{ijc}^*)}{\sum_{c=1}^{C}(1-\eta_{ijc})(1+r_{ijc}^*)} & \text{if } \gamma_{ij}^* = 0. \end{cases} \quad (9)$$

Equation (9) shows that β_{ij}^* is distributed to ω_{ijc}^* according to the closeness between $\boldsymbol{\alpha}_i^*$ and the attribute vector for option c of item j, where the probability is expressed as ω_{ijc}^*. If $\gamma_{ij}^* = 0$, Eq. (9) shows that the number 1 is distributed to ω_{ijc}^* according to the closeness. The selection probabilities of MC-S-DINA3 are shown in Table 4 for $\delta_1^* = 0.15$ and $\delta_2^* = 0.3$.

2 Multiple-choice models to estimate misconceptions

Let us consider α as a misconception here. Ozaki's (2015) three models shown above cannot be used for misconception cases. The reason is that when examinee i's misconception knowledge state $\boldsymbol{\alpha}_i$ is $\mathbf{0}$, namely, when the examinee possesses none of the misconceptions, all of Ozaki's (2015) models provide the probability $1/C_j$ for all of the options. However, when $\boldsymbol{\alpha}_i$ is the misconception knowledge state of an examinee i and $\boldsymbol{\alpha}_i = \mathbf{0}$, this examinee is likely to select the correct option or the option coded by none of the misconceptions. Therefore, Ozaki's (2015) models must be modified to accommodate the error analysis situation. Two models, referred to as DINO models for multiple-choice items to estimate misconceptions and denoted MC-M-DINO1 and MC-M-DINO2 are developed. In these models, "MC" refers to

Table 4 Selection probabilities for MC-S-DINA3

	α_1^*	1	1	0	0
	α_2^*	1	0	1	0
Option 1	(1, 1)	0.85	0.15	0.4	0.25
Option 2	(1, 0)	0.075	0.70	0.2	0.25
Option 3	(0, 0)	0.0375	0.075	0.2	0.25
Option 4	(0, 0)	0.0375	0.075	0.2	0.25

α_1^* and α_2^* are skills multiple choice, and “M” refers to misconception. The MC-M-DINO1 and MC-M-DINO2 models are based on MC-S-DINA1 and MC-S-DINA3, respectively.

2.1 The MC-M-DINO1 model

The MC-M-DINO1 model is expressed as follows:

$$P(Y_{ij}=c|\boldsymbol{\alpha}_i)=\gamma_{ij}\left(\frac{1-\delta_j}{\sum_{c=1}^{C_j}\eta_{ijc}}\right)^{\eta_{ijc}}\left(\frac{\delta_j}{C_j-\sum_{c=1}^{C_j}\eta_{ijc}}\right)^{1-\eta_{ijc}}+\frac{1-\gamma_{ij}}{C_j}, \tag{10}$$

where

$$\gamma_{ij}=1-0^{\sum_{c=1}^{C_j}\eta_{ijc}}, \tag{11}$$

and η_{ijc} is expressed as in Eq. (4). Note that in Eq. (4), α_{ik}^* has to be replaced with α_{ik} in the MC-M-DINO models. Therefore, $\eta_{ijc}=1$ when examinee i’s misconception knowledge state perfectly matches the misconception attribute vector of option c of item j, and $\eta_{ijc}=0$ otherwise. Table 5 shows η_{ijc} and γ_{ij} for the item shown in Table 1. Then, $\gamma_{ij}=1$ when at least one of $\eta_{ij1},\eta_{ij2},\ldots,\eta_{ijC_j}$ is 1, and $\gamma_{ij}=0$ when $\eta_{ijc}=0$ for all $c=(1,2,\ldots,C_j)$. In other words, when examinee i’s misconception knowledge state perfectly matches one or more of the misconception attribute vectors of options of item j, $\gamma_{ij}=1$, and $\gamma_{ij}=0$ otherwise. When $\gamma_{ij}=1$, the first term of Eq. (10) determines the selection probability, and when $\gamma_{ij}=0$, the second term $(1-\gamma_{ij})/C_j$ determines the selection probability. Here, $(1-\gamma_{ij})/C_j$ is the guessing part of the model for an examinee with $\gamma_{ij}=0$, which indicates that examinee i’s misconception knowledge state vector matches none of the misconception attribute vectors of options of item j.

Table 5 Terms η_{ijc}, γ_{ij}, and β_{ij} for the item shown in Table 1

	α_1	1	1	0	0
	α_2	1	0	1	0
η_{ijc} for option 1	(1, 1)	1	0	0	0
η_{ijc} for option 2	(1, 0)	0	1	0	0
η_{ijc} for option 3	(0, 0)	0	0	0	1
η_{ijc} for option 4	(0, 0)	0	0	0	1
γ_{ij}		1	1	0	1
γ_{ij}^*		1	1	0	0
$\sum\eta_{ijc}$		1	1	0	2
$C_j-\sum\eta_{ijc}$		3	3	4	2
β_{ij}		δ_{j1}	δ_{j2}	0	$(\delta_{j3}+\delta_{j4})/2$

α_1 and α_2 are misconceptions, η_{ijc} is expressed in Eq. (4), γ_{ij} is expressed in Eq. (11), γ_{ij}^* is expressed in Eq. (5), and β_{ij} is expressed in Eq. (13)

Then, δ_j is the probability that examinee i with $\boldsymbol{\alpha}_i$ selects options other than the most likely options, given his/her misconception knowledge state. Therefore, $1-\delta_j$ is the probability that an examinee whose $\boldsymbol{\alpha}_i$ matches a misconception attribute pattern for an option of item j actually selects these options. Table 5 also shows $\sum_{c=1}^{C_j}\eta_{ijc}$ and $C_j-\sum_{c=1}^{C_j}\eta_{ijc}$. Here, $\sum_{c=1}^{C_j}\eta_{ijc}$ is the number of options that examinee i with $\boldsymbol{\alpha}_i$ is most likely to select. Therefore, $1-\delta_j$ must be divided by $\sum_{c=1}^{C_j}\eta_{ijc}$. In Table 5, the examinee with $(\alpha_1,\alpha_2)=(1,1)$ is most likely to select only the first option. Therefore, $(1-\delta_j)/1=1-\delta_j$ is the selection probability. In Table 5, the examinee with $(\alpha_1,\alpha_2)=(0,0)$ is most likely to select the third or fourth option (which is why $\sum_{c=1}^{C_j}\eta_{ijc}=2$). Therefore, $(1-\delta_j)/2$ is the selection probability for both options. Then, δ_j must be divided by $C_j-\sum_{c=1}^{C_j}\eta_{ijc}$, which is the number of options that examinee i with $\boldsymbol{\alpha}_i$ is not most likely to select. Note that, when $(\alpha_1,\alpha_2)=(0,1)$, the second term of Eq. (10) determines the selection probability, because $\gamma_{ij}=0$. Note, moreover, that in Eq. (10), when $\sum_{c=1}^{C_j}\eta_{ijc}=0$, the first bracketed term cannot be calculated. However, in this case $\gamma_{ij}=0$. Therefore, the first term can be ignored.

The selection probabilities of MC-M-DINO1 are shown in Table 6 for $\delta_j=0.15$. Compared with Table 2, which shows the selection probabilities for MC-S-DINA1 for $\delta_j^*=0.15$, only the selection probabilities for examinee i with $(\alpha_1,\alpha_2)=(0,0)$ are different. For the MC-S-DINA1 case, $(\alpha_1,\alpha_2)=(0,0)$ means that such an examinee has none of the required attributes to correctly answer the item. Therefore, such an examinee is thought to randomly select an option resulting in selection probabilities of 0.25 for all options in Table 2. However, for the MC-M-DINO1 case, $(\alpha_1,\alpha_2)=(0,0)$ means that such an examinee has none of the misconception attributes to select one of the incorrect options. Therefore, such an examinee is thought to select a correct option or one of the incorrect options that are not coded by misconception attributes, resulting in the selection probabilities shown in Table 6.

The differences between MC-S-DINA1 and MC-M-DINO1 are as follows: (1) for MC-S-DINA1 in Eq. (3), $1-\delta_j^*$ is divided by 1 and δ_j^* is divided by C_j-1. However, for MC-M-DINO1 in Eq. (10), $1-\delta_j$ is divided by $\sum_{c=1}^{C_j}\eta_{ijc}$ and δ_j is divided by $C_j-\sum_{c=1}^{C_j}\eta_{ijc}$; (2) for MC-S-DINA1, γ_{ij}^* in Eq. (5) is 1 for the case in which examinee i has at least one attribute needed to correctly answer item j and $\sum_{c=1}^{C_j}\eta_{ijc}$ is 1. However, for MC-M-DINO1, γ_{ij} in Eq. (11) is 1 for the case in which at least one of $\eta_{ij1},\eta_{ij2},\ldots,\eta_{ijC_j}$ is 1. The point is that the number of correct options must be 1 for multiple-choice items. Therefore, the correct option is coded by a unique attribute vector. However, the number of incorrect options coded by the same misconception

Table 6 Selection probabilities for MC-M-DINO1

	α_1	1	1	0	0
	α_2	1	0	1	0
Option 1	(1, 1)	0.85	0.05	0.25	0.075
Option 2	(1, 0)	0.05	0.85	0.25	0.075
Option 3	(0, 0)	0.05	0.05	0.25	0.425
Option 4	(0, 0)	0.05	0.05	0.25	0.425

α_1 and α_2 are misconceptions attribute vector may be more than two for multiple-choice items. Therefore, for MC-S-DINA1, $1 - \delta_j^*$, which is the probability that an examinee who has the required attribute pattern for an option of item j actually selects that option is divided by 1 and δ_j^* is divided by $C_j - 1$. On the other hand, for MC-M-DINO1, $1 - \delta_j$ is divided by $\sum_{c=1}^{C_j} \eta_{ijc}$ and δ_j is divided by $C_j - \sum_{c=1}^{C_j} \eta_{ijc}$. Table 5 shows γ_{ij}^*. Note that γ_{ij}^* and γ_{ij} are not the same. Therefore, redefining misconception as conception and analyzing the data by MC-S-DINA1 cannot estimate misconceptions for multiple-choice items. The same is true for the differences between MC-S-DINA3 and MC-M-DINO2.

2.2 The MC-M-DINO2 model

The second model (MC-M-DINO2) has two features. The first feature is such that the model relaxes the constraint on MC-M-DINO1, whereby δ_j is the same for all options of item j. The second model is such that the selection probabilities for options other than the option that examinee i is most likely to select are different for each $\boldsymbol{\alpha}_i$, which are constrained to be the same for each $\boldsymbol{\alpha}_i$ in MC-M-DINO1.

The MC-M-DINO2 model is expressed as follows:

$$P(Y_{ij} = c|\boldsymbol{\alpha}_i) = \left(\frac{1 - \delta_{jc}}{\sum_{c=1}^{C_j} \eta_{ijc}} \right)^{\eta_{ijc}} \omega_{ijc}^{1-\eta_{ijc}}, \tag{12}$$

where

$$\beta_{ij} = \frac{\sum_{c=1}^{C_j} \delta_{jc}\eta_{ijc}}{\sum_{c=1}^{C_j} \eta_{ijc}}, \tag{13}$$

and

$$\omega_{ijc} = \begin{cases} \beta_{ij} \frac{(1-\eta_{ijc})(1+r_{ijc})}{\sum_{c=1}^{C}(1-\eta_{ijc})(1+r_{ijc})} & \text{if } \gamma_{ij} = 1 \\ \frac{(1-\eta_{ijc})(1+r_{ijc})}{\sum_{c=1}^{C}(1-\eta_{ijc})(1+r_{ijc})} & \text{if } \gamma_{ij} = 0. \end{cases} \tag{14}$$

In MC-M-DINO2, δ_{jc} takes different parameters for each option. Table 5 shows β_{ij}, which is the sum of the probabilities that examinee i with α_i selects an option other than the option having a misconception attribute vector that matches α_i.

For the second feature of MC-M-DINO2, the restriction that the selection probabilities for options other than the option that examinee i is most likely to select are the same for each $\boldsymbol{\alpha}_i$ is relaxed by setting the selection probabilities according to the closeness between $\boldsymbol{\alpha}_i$ and the attribute vector for option c of item j. Closeness is measured by r_{ijc}, as in the case of MC-S-DINA3.

Equation (14) shows that β_{ij} is distributed to ω_{ijc} according to the closeness between $\boldsymbol{\alpha}_i$ and the misconception attribute vector for option c of item j, where the probability is expressed as ω_{ijc}. If $\gamma_{ij} = 0$, then Eq. (14) shows that the number 1 is distributed to ω_{ijc} according to the closeness. The selection probabilities of MC-M-DINO2 are shown in Table 7 for $\delta_{j1} = 0.15$, $\delta_{j2} = 0.30$, and $\delta_{j3} = \delta_{j4} = 0.4$. Unlike Table 4, which shows the selection probabilities for MC-S-DINA3 for $\delta_1^* = 0.15$ and $\delta_2^* = 0.3$, only the selection probabilities for examinee i with $(\alpha_1, \alpha_2) = (0, 0)$ are different. This is the same relationship as that between MC-S-DINA1 and MC-M-DINO1. Therefore, it can be said that the major difference between MC-S-DINA and MC-M-DINO is in the selection probabilities in the case of $\boldsymbol{\alpha} = \mathbf{0}$. Table 8 shows the purposes, parameters, and the meanings of parts indicated by Greek letters for MC-S-DINA1, MC-S-DINA3, MC-M-DINO1, and MC-M-DINO2.

In practical application of the models, some goodness-of-fit indices are necessary in order to determine the best-fit model and to examine the absolute fit of the model. Appendix A of the supplemental file shows a method of model comparison and examining the absolute fit, and Appendix B of the supplemental file shows the reason why a model based on MC-S-DINA2 is not developed herein.

2.3 Parameter estimation

In estimating the parameters of the developed models, an MCMC method with the Metropolis–Hastings (M–H) algorithm (Hastings 1970) is adopted. Appendix C of the supplemental file shows the MCMC algorithms. Note that, in order to estimate parameters using the same method, the MCMC method with the M–H algorithm was also used to estimate parameters for Bug-DINO in the simulation study.

3 Simulation study

The purposes of the simulation study were the following: (1) to compare the accuracy of the estimates and the fit to data of Bug-DINO, MC-M-DINO1, and MC-M-DINO2 when data are generated from MC-M-DINO1 and MC-M-DINO2; and (2) to examine the effect of the number of items (which directly corresponds to the total number of coded options for each misconception), the number of examinees, and the rates of examinees who possess misconceptions on the estimates of the examinee parameters and item parameters. In the simulations, the estimation program was

Table 7 Selection probabilities for MC-M-DINO2

	α_1	1	1	0	0
	α_2	1	0	1	0
Option 1	(1, 1)	0.85	0.15	0.4	0.20
Option 2	(1, 0)	0.075	0.70	0.2	0.20
Option 3	(0, 0)	0.0375	0.075	0.2	0.30
Option 4	(0, 0)	0.0375	0.075	0.2	0.30

α_1 and α_2 are misconceptions

Table 8 Purposes, parameters, and meanings of parts indicated by Greek letters

Model	MC-S-DINA1	MC-S-DINA3	MC-M-DINO1	MC-M-DINO2
Purpose	To estimate skills	To estimate skills	To estimate misconceptions	To estimate misconceptions
Slip parameter	δ_j^*	δ_{jc}^*	δ_j	δ_{jc}
$\eta_{ijc}^* = 1$ or $\eta_{ijc} = 1$	When examinee i's knowledge state perfectly matches the attribute vector of option c of item j.		When examinee i's misconception knowledge state perfectly matches the misconception attribute vector of option c of item j	
$\gamma_{ij}^* = 1$ or $\gamma_{ij} = 1$	Examinee i has at least one attribute needed to correctly answer item j and one of $\eta_{ijc} = 1$ for $c(= 1, \dots, C)$		When at least one of $\eta_{ij1}, \eta_{ij2}, \dots, \eta_{ijC_j}$ is 1	
$\omega_{ijc}^* = 1$ or $\omega_{ijc} = 1$		Selection probabilities for options other than the option that examinee i is most likely to select		Selection probabilities for options other than the option that examinee i is most likely to select

written in R version 3.5.2, which can be downloaded from http://www010.upp.so-net.ne.jp/koken/cdm.html.

3.1 Simulation settings

The Q-vectors shown in Table 9 were used in the simulation study. The 1s and 2s in other than the "NC" columns indicate that the items were coded based on misconception once and twice, respectively. In the table, NC indicates the number of options coded by misconceptions. The number of options is four for all items. Therefore, for example, only one option is coded by the first attribute in item 1. The items shown in Tables 1 through 7 refer to item 16 in Table 9, because NC = 2 means that the number of coded options is two, and the first and second misconceptions appear twice and once, respectively. The Q-vectors are generated so that the total number of coded options for each misconception is the same, which is 14 in Table 9.

In the simulation study, three conditions for the number of items were examined: 10, 20, and 30 items. Items 7, 8, 11, 12, 14, and 16 through 20 were used for the 10-item case. Items 1 through 6, 9, 10, 13, 15, and 21 through 30 were used for the 20-item case. In addition, all of the items were used for the 30-item case. The reason for using these Q-vectors is that, within these conditions, the total number of coded options for each misconception for each number of items was the same: five for the 10-item case, nine for the 20-item case, and 14 for 30-item case. Furthermore, three conditions for the number of examinees were examined: 250, 500, and 1000 examinees. Then, two

Table 9 Q-vectors for the simulated data

Item	Misconception						Item	Misconception					NC
	1	2	3	4	5	NC		1	2	3	4	5	
1	1	0	0	0	0	1	16	2	1	0	0	0	2
2	0	1	0	0	0	1	17	0	2	1	0	0	2
3	0	0	1	0	0	1	18	0	0	2	1	0	2
4	0	0	0	1	0	1	19	0	0	0	2	1	2
5	0	0	0	0	1	1	20	1	0	0	0	2	2
6	1	1	0	0	0	2	21	1	1	1	0	0	3
7	1	0	1	0	0	2	22	1	1	0	1	0	3
8	1	0	0	1	0	2	23	1	1	0	0	1	3
9	1	0	0	0	1	2	24	1	0	1	1	0	3
10	0	1	1	0	0	2	25	1	0	1	0	1	3
11	0	1	0	1	0	2	26	1	0	0	1	1	3
12	0	1	0	0	1	2	27	0	1	1	1	0	3
13	0	0	1	1	0	2	28	0	1	1	0	1	3
14	0	0	1	0	1	2	29	0	1	0	1	1	3
15	0	0	0	1	1	2	30	0	0	1	1	1	3

NC is the number of coded options. The 1s and 2s in the "NC" columns indicate that the items were coded based on misconception once and twice, respectively conditions for the rates of examinees who possess each of the five misconceptions were examined: 0.2 and 0.4 (no difference between the five misconceptions). Finally, two conditions for the true models (MC-M-DINO1 and MC-M-DINO2) were examined. Therefore, the total number of conditions was 36 (= 3 × 3 × 2 × 2).

These simulation conditions were decided with reference to those of de la Torre (2009), where the number of items was 30, the number of examinees was 1000, and the number of attributes was five, and with reference to Ozaki (2015), where the number of items was 10, 20 or 30, the number of examinees was 1000, and the number of attributes was five. Both de la Torre (2009) and Ozaki (2015) used Q-vectors, in which some items are coded by a small number of attributes and other items are coded by a large number of attributes. However, the total number of coded options for each misconception was the same. Since Ozaki (2015) estimated attributes well in the 1000-examinee case, in the present paper, 250 and 500 cases were also examined.

The number of MCMC samples was 1500, and the burn-in was 500 for the two developed models and Bug-DINO. These numbers were determined using the convergence criterion of Gelman and Rubin (1992). By running five parallel chains, the criterion ($\hat{R} < 1.1$) was satisfied for all item parameters.

In both simulations, 50 repetitions were performed for each of the 36 conditions. When the true structure was MC-M-DINO1, the true δ_j were generated from Uniform (0,0.2). When the true structures were MC-M-DINO2, the true δ_{j1} were generated from Uniform (0,0.2), the true δ_{j2} were generated from Uniform (0.2,0.3), the true δ_{j3} were generated from Uniform (0.2,0.3), and the true δ_{j4} were generated from Uniform (0.2,0.3). For the purpose of model comparisons, the χ^2 values were calculated for MC-M-DINO using the method described in Appendix A1. When the number of coded options was 1, the other three options were pooled and two categories (selected or did not select the coded option) were used to calculate the χ^2 value (using 2 × 2 cross-tables comparing the observed and expected numbers of examinees) and the AIC. Similarly, when the numbers of coded options were three and four, three categories (using 2 × 3 cross-tables) and four categories (using 2 × 4 cross-tables) were used, respectively.

The results for the item parameters were examined using bias and root mean square error. Equation (15) shows the expression used to calculate Bias_{jc}^t, which is the bias for item parameter δ of item j of category c in the tth repetition, where $\hat{\delta}_{jc}^t$ is the estimate.

$$\mathrm{Bias}_{jc}^t = \hat{\delta}_{jc}^t - \delta_{jc}^t \tag{15}$$

Equation (16) shows the RMSE_j^t, which is the root mean square error of the item parameters of item j in the tth repetition.

$$\mathrm{RMSE}_j^t = \sqrt{\frac{1}{C_j}\sum_{c=1}^{C_j}(\hat{\delta}_{jc}^t - \delta_{jc}^t)^2}. \tag{16}$$

3.2 Results

The results of simulation study 1 are shown in Tables 10 through 13. In Table 10, the average correct recovery rate of α_{ik} is shown. For example, when MC-M-DINO1 was the true model, MC-M-DINO2 was fit to 10 items, and the number of examinees was 250, the average rate was 0.955. The table illustrates that, generally, the two multiple-choice models can estimate α_{ik} much more accurately than Bug-DINO. Therefore, using the information from the incorrect coded options is extremely useful for estimating α_{ik}. However, note that in this simulation, the true models were MC-M-DINO1 or MC-M-DINO2, and it therefore is reasonable

Table 10 Average recovery rate of examinee parameters for three analysis models

True model	Bug-DINO	MC1	MC2	True model	Bug-DINO	MC1	MC2
250 examinees	10 items (0.2)				10 items (0.4)		
MC1	0.275	0.970	0.955	MC1	0.412	0.972	0.970
MC2	0.303	0.879	0.855	MC2	0.444	0.892	0.905
	20 items (0.2)				20 items (0.4)		
MC1	0.627	0.998	0.998	MC1	0.716	0.992	0.992
MC2	0.618	0.970	0.982	MC2	0.701	0.953	0.968
	30 items (0.2)				30 items (0.4)		
MC1	0.834	1.000	1.000	MC1	0.757	0.999	0.999
MC2	0.612	0.990	0.994	MC2	0.660	0.985	0.991
500 examinees	10 items (0.2)				10 items (0.4)		
MC1	0.507	0.983	0.974	MC1	0.521	0.978	0.976
MC2	0.357	0.878	0.854	MC2	0.465	0.891	0.905
	20 items (0.2)				20 items (0.4)		
MC1	0.623	0.999	0.999	MC1	0.751	0.995	0.995
MC2	0.611	0.969	0.981	MC2	0.662	0.953	0.969
	30 items (0.2)				30 items (0.4)		
MC1	0.779	1.000	1.000	MC1	0.768	0.999	0.999
MC2	0.594	0.989	0.995	MC2	0.644	0.984	0.992
1000 examinees	10 items (0.2)				10 items (0.4)		
MC1	0.614	0.979	0.968	MC1	0.551	0.968	0.966
MC2	0.490	0.880	0.857	MC2	0.490	0.894	0.907
	20 items (0.2)				20 items (0.4)		
MC1	0.629	0.999	0.998	MC1	0.698	0.991	0.991
MC2	0.586	0.969	0.983	MC2	0.646	0.952	0.969
	30 items (0.2)				30 items (0.4)		
MC1	0.831	1.000	1.000	MC1	0.768	0.999	0.999
MC2	0.581	0.990	0.996	MC2	0.637	0.985	0.992

MC1 and MC2 indicate MC-M-DINO1 and MC-M-DINO2, respectively. The numbers (0.2 or 0.4) in parentheses are the rates of examinees who possess each of the five misconceptions that rather than the non-true model (Bug-DINO), the true models provided higher recovery rates.

When the analysis models were MC-M-DINO1 or MC-M-DINO2, the number of examinees had little effect on the recovery rate. Therefore, for MC-M-DINO1 and MC-M-DINO2, 250 examinees was sufficient for these settings. However, when the analysis model was Bug-DINO, as the number of examinees decreased, the recovery rate decreased. In particular, when the number of items was 10, Bug-DINO often provided recovery rates lower than 0.5. In these cases, although not shown here, the estimates of the slip and guessing parameters were both large. Therefore, Bug-DINO inversely estimated 0 and 1 for α_{ik}, which is illustrated in Table 13. This is the reason for such poor recovery rates for Bug-DINO for cases in which the number of items was 10.

The number of items and the two conditions for the rates of examinees who possess each of the five misconceptions also had little effect on the recovery rate, when the analysis models were MC-M-DINO1 or MC-M-DINO2. Therefore, although a larger number of examinees, a larger number of items, and/or larger misconception rates provide higher recovery rates, the combination of 10 items, 250 examinees, and a misconception rate of 0.2 was sufficient (the recovery rate was more than 0.85) to estimate α_{ik} by MC-M-DINO1 or MC-M-DINO2.

In general, the average recovery rate for the three models decreased as the true model became more complicated. However, the decrement of the rate is larger for Bug-DINO than for the two developed models.

In general, when the true model was MC-M-DINO1, the average recovery rate for the true model is greater than or equal to that for MC-M-DINO2. When the true model was MC-M-DINO2, in most cases, the average correct recovery rate for the true model was also larger than that for MC-M-DINO1. However, when the number of items was 10 and the misconception rate was 0.2, MC-M-DINO1, which has fewer item parameters, provided higher recovery rates, even when the true model was MC-M-DINO2.

Table 11 shows the results for the item parameters. The bias shown in Table 11 is the average of Bias_{jc}^{t} in Eq. (15) for all cases. The RMSE shown in Table 11 is the average of RMSE_{j}^{t} in Eq. (16) for all cases.

Table 11 illustrates that, for all cases, the biases are close to 0. In addition, the RMSEs are small when the number of examinees is large. However, when the number of items was 10, the true misconception rate was 0.2, and MC-M-DINO2 was fit, the biases and RMSEs were relatively large. Therefore, the item parameters can be estimated well except for this case.

Table 12 shows the results of the model comparisons. The numbers in the tables are the average number of items that fit best for each model, based on the AIC (the best-fit model was taken to be the model that gives the smallest AIC). If a model is often selected using the AIC, this means that the model is parsimonious when describing data. Generally, the AIC prefers the true model. However, although not shown in Table 11, when the BIC was used, the parsimonious MC-M-DINO1 was preferred more frequently. A better fitting model should be used to interpret the results of α_{ik} in the real-data analysis. Furthermore, the absolute fit of the best-fit model should be examined, because the best-fit model does not

Table 11 Biases and RMSEs of item parameters for two analysis models

Model	Bias	RMSE	Model	Bias	RMSE
250 examinees	10 items (0.2)			10 items (0.4)	
MC1	– 0.010	0.028	MC1	0.008	0.029
MC2	0.017	0.222	MC2	– 0.005	0.091
	20 items (0.2)			20 items (0.4)	
MC1	– 0.004	0.024	MC1	0.009	0.028
MC2	0.004	0.077	MC2	– 0.006	0.067
	30 items (0.2)			30 items (0.4)	
MC1	0.015	0.025	MC1	– 0.008	0.027
MC2	0.005	0.079	MC2	0.017	0.064
500 examinees	10 items (0.2)			10 items (0.4)	
MC1	0.008	0.017	MC1	0.009	0.019
MC2	– 0.015	0.228	MC2	0.008	0.073
	20 items (0.2)			20 items (0.4)	
MC1	0.008	0.015	MC1	– 0.008	0.017
MC2	0.010	0.052	MC2	0.012	0.045
	30 items (0.2)			30 items (0.4)	
MC1	– 0.007	0.016	MC1	0.009	0.018
MC2	0.016	0.055	MC2	0.011	0.042
1000 examinees	10 items (0.2)			10 items (0.4)	
MC1	0.004	0.012	MC1	– 0.004	0.014
MC2	0.016	0.234	MC2	0.002	0.062
	20 items (0.2)			20 items (0.4)	
MC1	– 0.003	0.010	MC1	0.003	0.012
MC2	0.012	0.034	MC2	0.007	0.030
	30 items (0.2)			30 items (0.4)	
MC1	0.004	0.010	MC1	0.004	0.011
MC2	0.004	0.036	MC2	– 0.009	0.029

MC1 and MC2 indicate MC-M-DINO1 and MC-M-DINO2, respectively. The numbers (0.2 or 0.4) in parentheses are the rates of examinees who possess each of the five misconceptions

necessarily fit the data absolutely. The absolute fit was calculated in the real-data study.

Table 13 shows two-way classified tables of examinees for the first misconception by the three models when the number of examinees was 250, the number of items was 10, the true misconception rate was 0.2, and the true model was MC-M-DINO1. For example, when the true $\alpha = 1$, an average of 45.78 examinees were classified as $\alpha = 0$ by Bug-DINO and as $\alpha = 1$ by MC-M-DINO1. For the second example, when the true $\alpha = 0$, an average of 190.08 examinees were classified as $\alpha = 0$ by MC-M-DINO1 and $\alpha = 0$ by MC-M-DINO2. Since the true misconception rate was 0.2, $250 \times 0.2 = 50$ examinees should be classified as $\alpha = 1$ and 200 examinees should

Table 12 Number of items each model fit best using the AIC

True model	MC1	MC2	True model	MC1	MC2
250 examinees	10 items (0.2)			10 items (0.4)	
MC1	8.96	1.04	MC1	9.68	0.32
MC2	6.98	3.02	MC2	7.82	2.18
	20 items (0.2)			20 items (0.4)	
MC1	18.84	1.16	MC1	19.16	0.84
MC2	15.80	4.20	MC2	12.16	7.84
	30 items (0.2)			30 items (0.4)	
MC1	28.78	1.22	MC1	28.84	1.16
MC2	23.48	6.52	MC2	17.58	12.42
500 examinees	10 items (0.2)			10 items (0.4)	
MC1	9.46	0.54	MC1	9.84	0.16
MC2	5.44	4.56	MC2	6.30	3.70
	20 items (0.2)			20 items (0.4)	
MC1	19.36	0.64	MC1	19.50	0.50
MC2	13.82	6.18	MC2	9.14	10.86
	30 items (0.2)			30 items (0.4)	
MC1	28.78	1.22	MC1	29.28	0.72
MC2	19.42	10.58	MC2	11.38	18.62
1000 examinees	10 items (0.2)			10 items (0.4)	
MC1	9.34	0.66	MC1	9.66	0.34
MC2	3.84	6.16	MC2	4.24	5.76
	20 items (0.2)			20 items (0.4)	
MC1	19.30	0.70	MC1	19.28	0.72
MC2	10.56	9.44	MC2	4.56	15.44
	30 items (0.2)			30 items (0.4)	
MC1	28.78	1.22	MC1	29.36	0.64
MC2	13.98	16.02	MC2	5.22	24.78

MC1 and MC2 indicate MC-M-DINO1 and MC-M-DINO2, respectively. The numbers (0.2 or 0.4) in parentheses are the rates of examinees who possess each of the five misconceptions

be classified as $\alpha = 0$. The table shows that, as shown in Table 10, the two developed models estimate α well. However, as noted previously, Bug-DINO tends to inversely estimate 0 and 1 for α.

4 Real-data study

In the real-data study, eight items, all of which examine the skill of recognizing the dependency relations between words and phrases in a given sentence (Arai et al. 2017), were used. These items are used in the Reading Skill Test in Japan and were

Table 13 Two-way classified tables of examinees for the first misconception

$\alpha = 1$	MC1(0)	MC1(1)	$\alpha = 0$	MC1(0)	MC1(1)
DINO(0)	0.56	45.78	DINO(0)	25.92	3.28
DINO(1)	0.46	3.28	DINO(1)	168.66	1.76
$\alpha = 1$	MC2(0)	MC2(1)	$\alpha = 0$	MC2(0)	MC2(1)
DINO(0)	0.48	45.86	DINO(0)	23.68	5.82
DINO(1)	0.48	3.26	DINO(1)	167.38	3.04
MC1(0)	0.74	0.28	MC1(0)	190.08	4.5
MC1(1)	0.22	48.84	MC1(1)	0.98	4.36

MC1 and MC2 indicate MC-M-DINO1 and MC-M-DINO2, respectively. $\alpha = 1$ and $\alpha = 0$ indicate whether the true α is 1 or 0. The numbers (0 or 1) in parentheses indicate the cases in which examinees are estimated as not possessing a misconception (0) or as possessing a misconception (1)

not originally made for the purpose of estimating misconceptions. However, estimating examinees' misconceptions is useful for knowing why each student is stumbling in learning, and illustrating that the models can also be used for items not originally constructed for the purpose of estimating misconceptions demonstrates the versatility of the models.

The number of examinees was 281 (40 junior high school students and 241 senior high school students). Two misconceptions, A, in which the nearest word is the subject, and B, in which the word beginning a sentence is the subject, were used. An example of an item that measures both misconceptions is shown in Fig. 1. The Q-vectors for the eight items are shown in Table 14, which indicates that three items measure only misconception A, three items measure only misconception B, one item measures both misconceptions A and B by the same option, and one item measures both misconceptions A and B by different options. Although the number of options

Table 14 Q-vectors, item parameters, and χ^2 tests for the real-data study

Item	Misconception		NC	δ	χ^2	*df*	*p*
1	0	1	1	0.076	1.336	3	0.721
2	1	0	1	0.060	19.376	3	0.000***
3	0	1	1	0.084	3.386	3	0.336
4	1	0	1	0.069	26.481	3	0.000***
5	0	1	1	0.047	37.589	3	0.000***
6	1	1	1	0.177	1.452	3	0.693
7	1	0	1	0.372	0.807	3	0.848
8	1	1	2	0.138	11.908	11	0.371

NC is the number of coded options, δ is the estimated item parameter, χ^2 is the χ^2 value for each item, df is the degree of freedom for each item, and p is the p value of the χ^2 test for each item. $***p < 0.001$

was four for all of the items, options that were not coded by misconception attributes were summed in the same option.

If the mean of MCMC samples for α_{ik} was greater than 0.5, examinee i was judged to possess misconception k; otherwise, the examinee was judged not to possess the misconception. The AIC and BIC for MC-M-DINO1 were 49.481 and 79.003, respectively, and those for MC-M-DINO2 were 73.550 and 136.287, respectively. Therefore, MC-M-DINO1 was the best-fit model. Furthermore, the absolute fit of MC-M-DINO1 was examined for each item using the method described in Appendix A2. Table 14 shows the χ^2 value, the degree of freedom, and the p-value for each item. The table indicates that items 2, 4, and 5 do not fit the data. However, the estimated misconception rate for misconception A was 0.027, and that for B was 0.030. The number of examinees was 281, and the expected and observed numbers of examinees for some cells were less than 5. Thus, the approximation to the chi-squared distribution was not good. Therefore, a large sample size is needed in order to examine whether items 2, 4, and 5 truly fit poorly.

The results for the item parameters (δ) for MC-M-DINO1 are shown in Table 14, which indicates that an examinee who possesses misconception A or B selects the coded options with a probability of more than 90% for items 1 through 5. The probability is more than 60% for items 6 through 8. These high probabilities show that the items can detect examinees' possession of misconceptions very well.

As stated in Appendixes A1 and A2, χ^2 tests require large sample sizes. However, the sample size of the real data is not large. Therefore, other model-fit indices may be needed, which is a subject for future study.

5 Discussion

One of the difficulties with the CDM is that it is difficult to appropriately specify the Q-matrix. Although we now have ways to examine the validation of a Q-matrix (de la Torre and Chiu 2016; Chen 2017) or its completeness, which means that the Q-matrix allows for the identification of all possible proficiency classes among examinees (Köhn and Chiu 2017), de la Torre and Chiu (2016) noted that the process of establishing the Q-matrix for a given test tends to be subjective in nature. Therefore, in Rupp and Templin (2008) and Im and Corter (2011), the effects of misspecification of the Q-matrix were examined. The difficulty in specifying the Q-matrix when possession of α is required in order to answer items correctly lies in breaking up the way of reaching the correct answer into several appropriate skills. However, if α is a misconception, specification of the Q-matrix is easier, because incorrect options are usually or sometimes intentionally designed using misconceptions. Therefore, the developed models that can detect misconceptions of examinees using data from multiple-choice items may broaden the use of the CDM. However, if the Q-matrix is misspecified in the misconception case, then the estimation accuracy for α decreases.

Although two models were developed to be used for multiple-choice items, these models can be used for tests in which examinees write how they arrive at their answers. Based on their answers, examinees' can be classified into two groups: individuals who arrive or do not arrive at incorrect answers for an item because of a misconception. If an examinee frequently falls for a misconception option, he/she probably possesses the misconception. This judgment can be performed using the two developed models.

Recently Kuo et al. (2018) developed a model that can estimate examinees' skills and misconceptions at the same time by analyzing binary data. Providing each examinee's possession state of skills and misconceptions at the same time is useful because this is the information of the entire knowledge state of each examinee. However, as de la Torre (2009) and Ozaki (2015), and the present paper revealed, in the case of multiple-choice data, models for multiple-choice data can provide more accurate estimation of α. Therefore, a model that can estimate examinees' skills and misconceptions at the same time for multiple-choice data is needed.

In MC-S-DINA1 and MC-S-DINA2, two options that have the same misconception attributes provide the same selection probabilities. For example, Table 6 shows that the selection probabilities for options 3 and 4 (the correct option) are the same, because both options are coded by none of the misconceptions. However, an examinee who has the required skills and lacks the misconceptions would normally select option 4. To treat the options that are coded by the same misconception attributes differently, not only misconception attributes but also skills have to be included in the models. A model that can estimate the examinee's skills and misconceptions at the same time for multiple-choice data is needed in this sense as well.

As stated previously, DiBello et al. (2015) developed a generalized diagnostic classification model for multiple-choice items that can estimate examinees' possession of skills and misconceptions at the same time with a large number of parameters. A large number of parameters is not necessarily a bad thing. Models that have larger numbers of parameters can capture the response behavior in detail. Moreover, if the sample size is large, these parameters may be estimated accurately. Therefore, comparing the estimation accuracy of misconceptions between the developed parsimonious models and the model of DiBello et al. (2015) by a simulation study may be interesting.

The simulation studies showed that using the information from the incorrectly coded options is extremely useful for estimating α_{ik}. However, in this simulation, in which the true models were MC-M-DINO1 or MC-M-DINO2, it would be reasonable that rather than the non-true model (Bug-DINO), the true models provided higher recovery rates. To examine the robustness of this model, an additional simulation study that examines the recovery rate when the true models and the analysis models are totally different would be necessary. Furthermore, some overall model fit indices (at the model-level rather than the item-level), such as the posterior predictive model check under a Bayesian framework, are also necessary.

In the real-data study, due to the small number of examinees and very low estimated misconception rates, the χ^2 approximation was not good. Other fit indices that can overcome this problem are needed.

The results of the model comparisons indicate that the best-fit model tends to differ according to the number of coded options. Therefore, when analyzing multiple-choice items, the best-fit model should be examined item by item. Although this requires significant time, this can be achieved by performing analysis $2^{\text{number of items}}$ times.

Acknowledgements This research was funded by Grant-in-Aid for Scientific Research(C) 18K03057.

References

Arai NH, Todo N, Arai T, Bunji K, Sugawara S, Inuzuka M, Matsuzaki T, Ozaki K (2017) Reading skill test to diagnose basic language skills in comparison to machines. In: Proceedings of the 39th annual cognitive science society meeting (CogSci 2017), pp 1556–1561

Chen J (2017) A residual-based approach to validate Q-Matrix specifications. Appl Psychol Meas 41(4):277–293

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

de la Torre J (2009) A cognitive diagnosis model for cognitively based multiple-choice options. Appl Psychol Meas 33(3):163–183

de la Torre J (2011) The generalized DINA model framework. Psychometrika 76(2):179–199

de la Torre J, Chiu C-Y (2016) A general method of empirical Q-matrix validation. Psychometrika 81(2):253–273

DiBello LV, Henson RA, Stout WF (2015) A family of generalized diagnostic classification models for multiple choice option-based scoring. Appl Psychol Meas 39(1):62–79

Gelman A, Rubin DB (1992) Inference from iterative simulation using mutiple sequences. Stat Sci 7(4):457–472

Hartz S (2002) A Bayesian framework for the unified model for assessing cognitive abilities: blending theory with practicality (Doctoral dissertation). University of Illinois, Urbana-Champaign

Hastings WK (1970) Monte carlo sampling methods using markov chains and their applications. Biometrika 57(1):97–109

Im S, Corter JE (2011) Statistical consequences of attribute misspecification in the rule spece method. Educ Psychol Meas 71(4):712–731

Junker BW, Sijtsma K (2001) Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. Appl Psychol Meas 25(3):258–272

Köhn HF, Chiu CY (2017) A procedure for assessing the completeness of the Q-Matrices of cognitively diagnostic tests. Psychometrika 82(1):112–132

Kuo B-C, Chen C-H, Yang C-W, Mok MMC (2016) Cognitive diagnostic models for tests with multiple-choice and constructed-response items. Educ Psychol 36(6):1115–1133

Kuo B-C, Chen C-H, de la Torre J (2018) A cognitive diagnosis model for identifying coexisting skills and misconceptions. Appl Psychol Meas 42(3):179–191

Maris E (1999) Estimating multiple classification latent class models. Psychometrika 64(2):187–212

Minchen ND, de la Torre J, Liu Y (2017) A cognitive diagnosis model for continuous response. J Educ Behav Stat 42(6):651–677

Ozaki K (2015) DINA models for multiple-choice items with few parameters: considering incorrect answers. Appl Psychol Meas 39(6):431–447

Richards JC, Schmidt R (2002) Dictionary of language teaching and applied linguistics, 3rd edn. Longman, London

Rupp AA, Templin JL (2008) The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. Educ Psychol Meas 68(1):78–96

Templin J, Henson R (2006) Measurement of psychological disorders using cognitive diagnosis models. Psychol Methods 11(3):287–305