



# The Q-Matrix Anchored Mixture Rasch Model

Ming-Chi Tseng<sup>1\*</sup> and Wen-Chung Wang<sup>2\*</sup>

<sup>1</sup> National University of Tainan, Tainan, Taiwan, <sup>2</sup> The Education University of Hong Kong, Tai Po, Hong Kong

Mixture item response theory (IRT) models include a mixture of latent subpopulations such that there are qualitative differences between subgroups but within each subpopulation the measure model based on a continuous latent variable holds. Under this modeling framework, students can be characterized by both their location on a continuous latent variable and by their latent class membership according to Students' responses. It is important to identify anchor items for constructing a common scale between latent classes beforehand under the mixture IRT framework. Then, all model parameters across latent classes can be estimated on the common scale. In the study, we proposed Q-matrix anchored mixture Rasch model (QAMRM), including a Q-matrix and the traditional mixture Rasch model. The Q-matrix in QAMRM can use class invariant items to place all model parameter estimates from different latent classes on a common scale regardless of the ability distribution. A simulation study was conducted, and it was found that the estimated parameters of the QAMRM recovered fairly well. A real dataset from the Certificate of Proficiency in English was analyzed with the QAMRM, LCDM. It was found the QAMRM outperformed the LCDM in terms of model fit indices.

## OPEN ACCESS

### Edited by:

Chun Wang,  
University of Washington,  
United States

### Reviewed by:

Mirka Henninger,  
University of Zurich, Switzerland  
Sedat Sen,  
Harran University, Turkey

### \*Correspondence:

Ming-Chi Tseng  
mingchi@gms.ndhu.edu.tw  
Wen-Chung Wang  
wawang@eduhk.hk

### Specialty section:

This article was submitted to  
quantitative Psychology  
and Measurement,  
a section of the journal  
Frontiers in Psychology

Received: 24 June 2020

Accepted: 12 February 2021

Published: 04 March 2021

### Citation:

Tseng M-C and Wang W-C (2021)  
The Q-Matrix Anchored Mixture  
Rasch Model.  
Front. Psychol. 12:564976.  
doi: 10.3389/fpsyg.2021.564976

**Keywords:** Q-matrix anchored mixture Rasch model, Q-matrix, anchor, mixture Rasch model, Rasch model

## INTRODUCTION

Measurement invariance is a key assumption that enables score comparison across different groups of respondents (Hambleton et al., 1991). In reality, the assumption may not hold and needs to be checked empirically. In the context of Rasch measurement, different groups of respondents may take different views on items, resulting in measurement non-invariance. Rost (1990) integrated latent class analysis (LCA; Lazarsfeld and Henry, 1968) to the Rasch model (Rasch, 1960) and derived the mixture Rasch model (MRM), which can be viewed as an extension of the Rasch model that allows different groups (latent classes) of respondents to have different item parameters and ability distributions. To the extent that these classes are substantively meaningful, the mixture Rasch model provides a potentially important means to understanding how and why examinees respond in different ways. It is assumed that a Rasch model holds in each class, but each class may have different item and ability parameters. Specifically, the probability of a correct response in the MRM can be given as:

$$P(Y_{ij} = 1) = \sum_{g=1}^G \pi_g \cdot P(Y_{ijg} = 1 | g, \theta) = \sum_{g=1}^G \pi_g \cdot \frac{\exp(\theta_{ig} - b_{jg})}{1 + \exp(\theta_{ig} - b_{jg})}, \quad (1)$$

where  $g$  is an index for the latent class,  $g = 1, \dots, G$ ,  $i = 1, \dots, N$  examinees,  $\pi_g$  is the proportion of examinees for each class,  $\theta_{ig}$  is the latent ability of examinee  $i$  in latent class  $g$ , and  $b_{jg}$  is the

difficulty parameter of item  $j$  for latent class  $g$ . The MRM can account for qualitative differences between latent classes and quantitative differences within latent classes (Rost, 1990).

An important feature of the MRM or other mixture models (von Davier and Yamamoto, 2007) is that the number of latent classes must be explored from the data, which is an exploratory approach. Usually, the Akaike information criterion (AIC), Bayesian information criterion (BIC), or deviance information criterion (DIC) are applied to determine the number of latent classes but they do not always provide the same answer. Over- or under-extraction of latent classes may occur, making the interpretation problematic (Alexeev et al., 2011). It is desirable to adopt a constrained approach to the identification of latent classes when there are substantive theories or hypotheses.

Recent developments in the Q-matrix (Tatsuoka, 1983) for diagnostic classification models (DCMs) may help with the identification. Domain-specific assessment experts encode the relationships that they believe exist between the diagnostic assessment items and the latent variables that are used to classify respondents into so-called Q-matrices. The attribute is a latent characteristic of respondents in the Q-matrices.

Choi (2010) develop the diagnostic classification mixture Rasch model (DCMixRM) which combines a Mixture Rasch model with log-linear cognitive diagnostic model (LCDM; Henson et al., 2009). In the DCMixRM, this model includes mastery states of attributes as covariates. To be more specific, in the measurement component, observed item responses are jointly regressed on latent trait and attributes through the Rasch model and the LCDM. Next, in the structural model, ability is regressed on class membership, and class membership is regressed on mastery profile to explain latent class as covariates. Besides, Bradshaw and Templin (2014) develop the Scaling Individuals and Classifying Misconceptions (SICM) model which is presented as a combination of a unidimensional IRT model and LCDM where the categorical latent variables represent misconceptions instead of skills. In the SICM, IRT, and LCDM assumed to be orthogonal as in the original bifactor model (Gibbons and Hedeker, 1992). Theoretically, SICM expected that subjects vary in ability even when they possess the same misconception pattern, meaning a significant correlation between ability and misconception pattern was not expected or modeled.

On the other hand, we developed the Q-matrix anchor mixture Rasch model (QAMRM) by incorporating the Q-matrix into the MRM. The QAMRM is constrained because the number of latent classes is specified by users rather than explored from the data. The latent traits in the QAMRM can be compensatory or non-compensatory. The Q-matrix contains a set of elements  $q_{jk}$  indicating whether attribute  $k$  is required to answer item  $j$  correctly, and  $q_{jk} = 1$  if the attribute is required, otherwise it is 0. The total number of attributes and the value of  $q_{jk}$  is assigned by content experts. Similar Q-matrices have been adopted in IRT to specify *a priori* which latent traits (components) have been measured by which items, such as the linear logistic test model (LLTM; Fischer, 1973), the multicomponent latent trait model (Whitely, 1980), the loglinear multidimensional IRT model for polytomously scored items (Kelderman and Rijkes, 1994),

the multidimensional random coefficients multinomial logic model (Adams et al., 1997), the multidimensional componential IRT model for polytomous items (Hoskens and De Boeck, 2001), and the multicomponent latent trait model for diagnosis (Embretson and Yang, 2013).

The QAMRM uses the Q-matrix to check whether different classes have different measurement characteristics. The utility of this approach lies in the fact that the numbers of latent classes, immediately observable through the Q-matrix, are defined in advance and can be used to help explain item level performance to discover how members in one class differ from another. It is these Q-matrix differences in response propensities that help explain the potential causes of these differential measurement characteristics.

The approach proposed in this study provides the Q-matrix by means of a design matrix describing the composition of the different classes. We begin below by illustrating how the QAMRM can be viewed as incorporating features from the Q-matrix, and then through the Q-matrix to establish class invariant items and allow all model parameter estimates across latent classes to be on a common scale regardless of the ability distribution.

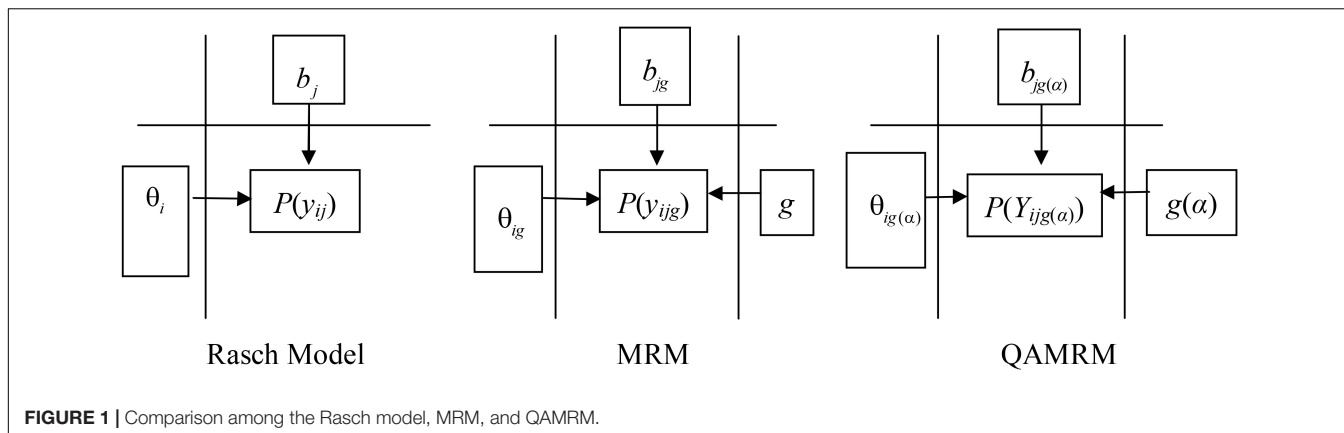
## THE Q-MATRIX ANCHOR MIXTURE RASCH MODEL

We adapted **Figure 1** from Wright and Stone (1979) about the Rasch model to express the response difference between the Rasch model, MRM, and QAMRM.

In the Rasch model diagram, person ability  $\theta_i$  and item difficulty  $b_j$  jointly determine response  $P(y_{ij})$ . In the MRM diagram, conditional on latent class  $g$ , person ability  $\theta_{ig}$  and item difficulty  $b_{jg}$  jointly determine response  $P(y_{ijg})$ . In the QAMRM diagram, conditional on class  $g(\alpha)$ , person ability  $\theta_{ig(\alpha)}$  and item difficulty  $b_{jg(\alpha)}$  jointly determine response  $P(Y_{ijg(\alpha)})$ . When there is no Q-matrix, the number of classes is estimated from data, the QAMRM simplifies to the MRM; when there is only single class, the MRM simplifies to the Rasch model.

The QAMRM has multiple classes that follow the Q-matrix design matrix. Like the MRM, it assumes that there may be heterogeneity in response patterns at different classes which should not be ignored (Mislevy and Verhelst, 1990; Rost, 1990), but should consider the Q-matrix to form the number of classes beforehand, rather than forming the number of classes during the parameter estimation. Viewed in this way, the Q-matrix inside the QAMRM captures the association between the items and classes. The probability of getting a correct response in the QAMRM can be given as follows:

$$P(Y_{ij} = 1) = \sum_{g(\alpha)=1}^G \pi_{g(\alpha)} \cdot P(Y_{ijg(\alpha)} = 1 | g(\alpha), \theta_{(\alpha)}) = \sum_{g(\alpha)=1}^G \pi_{g(\alpha)} \cdot \frac{\exp(\theta_{ig(\alpha)} - b_{jg(\alpha)})}{1 + \exp(\theta_{ig(\alpha)} - b_{jg(\alpha)})}, \quad (2)$$



where  $Y_{ijg(\alpha)}$  is the score of examinee  $i$  ( $i = 1, \dots, N$ ) on item  $j$  ( $j = 1, \dots, J$ ) in class  $g$  conditional on attribute profile  $\alpha$  ( $\alpha = \alpha_1, \dots, \alpha_K$ ),  $\theta_{ig(\alpha)}$  is the latent ability of examinee  $i$  within class  $g$  conditional on attribute  $\alpha$ , and  $b_{jg(\alpha)}$  is the difficulty parameter of item  $j$  for class  $g$  conditional on attribute  $\alpha$ . Like the exploratory MRM, there is only one mixing proportion/structural parameter/latent class membership probability in the QAMRM,  $\pi_{g(\alpha)}$ , which is the probability of being in class  $g$  conditional on attribute pattern  $\alpha$ .

## THE Q-MATRIX SETS CLASS INVARIANT ITEMS A PRIORI IN QAMRM

Paek and Cho (2015) posit four scenarios to establish a common scale across latent classes in MRM and suggest proposing the use of class-invariant items, which was also suggested by von Davier and Yamamoto (2004). Those invariant items have the same item difficulties across latent classes. Once a set of class invariant items are available, this ensures a common scale across latent classes.

The challenge in MRM is to identify class invariant items, because selecting the best measurement model regarding the numbers of item parameters, latent groups, and dimensions are mainly decided by statistical procedure *post hoc*. Some studies exist in the MRM literature (e.g., Cho et al., 2016), where a statistical procedure was applied as an attempt to locate class invariant items in real data analyses, but how to find class invariant items in the context of MRM until now remains unclear.

Paek and Cho (2015) suggest the use of class invariant items to recover the parameter differences correctly in both item profiles and ability distributions for latent classes when those differences exist simultaneously in MRM; in QAMRM we do not need to find class invariant items by statistical procedure *post hoc* to let all model parameter estimates across latent classes be on a common scale because the Q-matrix has already done that *a priori*. The invariant items in MRM are data driven, but in QAMRM, even though the latent classes follow different ability distributions in terms of their means, we still can easily set invariant items in QAMRM through the Q-matrix *a priori*. Even though there are no ability distributions in the latent class model, through the Q-matrix being set beforehand, the

invariance parameters of the DINA model (de la Torre and Lee, 2010) and LCDM model (Bradshaw and Madison, 2016) still hold when the model fit the data.

LCDM is a flexible model that allows the relationships between categorical variables to be modeled using a latent class model, because most cognitive diagnosis models are typically parameterized to define the probability of a correct response, LCDM is re-expressed in terms of the log-odds of a correct response for each item. In addition, von Davier (2005) discusses the General Diagnostic Model (GDM) as a general approach to log-linear models with latent variables, where the latent variables are both continuous and discrete in addition to focusing on ordered responses for items. As a special case, the GDM general definition easily incorporates LCDM with dichotomous latent variables for dichotomous (von Davier, 2014). Besides, Hong et al., 2015 combines DINA model and non-compensatory item response theory to form DINA-NIRT model, which tries to combine continuous and discrete latent variables in cognitive diagnosis. In contrast, the DINA-NIRT is a special case of LCDM because it does not have compensatory attributes. The QAMRM combines continuous and discrete latent variables in the same framework, which includes compensatory (disjunctive) and non-compensatory (conjunctive) models.

In compensatory QAMRM, a low value on one latent variable can be compensated for by a high value on another latent variable, so it is not necessary to master all attributes that are required by an item to produce a correct response within a class. On the contrary, in non-compensatory QAMRM, a low value on one latent variable cannot be compensated for by a high value on another latent variable, so it is necessary to master all attributes that are required by an item to produce a correct response within a class. In the QAMRM, the relationship among latent variables, either compensatory or non-compensatory, is assumed to be identical across classes.

Substantive theories can help decide the numbers of attributes and classes prior to parameter estimation, as done in the Q-matrix. As an example, let there be 14 items measuring three binary attributes in the QAMRM. In total, there will be eight ( $2^3$ ) attribute profiles, which are called classes  $g_1$ - $g_8$  as shown in column 1 in **Table 1**.

**TABLE 1** | Class, Q-matrix and item difficulty for 14-item 3 attributes QAMRM.

1	Class			Q-matrix				Non-compensatory									Compensatory								
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
	$\alpha_1$	$\alpha_2$	$\alpha_3$	Item	$\alpha_1$	$\alpha_2$	$\alpha_3$	Item	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$	$g_6$	$g_7$	$g_8$	Item	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$	$g_6$	$g_7$	$g_8$
$g_1$	0	0	0	1	1	0	0	1	2 <sup>a</sup>	2	2	2	1	1	1	1	1	2	2	2	2	1	1	1	1
$g_2$	0	0	1	2	0	1	0	2	2	2	1	1	2	2	1	1	2	2	2	1	1	2	2	1	1
$g_3$	0	1	0	3	0	0	1	3	2	1	2	1	2	1	2	1	3	2	1	2	1	2	1	2	1
$g_4$	0	1	1	4	1	1	0	4	2	2	2	2	2	2	1	1	4	3	3	2	2	2	2	1	1
$g_5$	1	0	0	5	1	0	1	5	2	2	2	2	2	1	2	1	5	3	2	3	2	2	1	2	1
$g_6$	1	0	1	6	0	1	1	6	2	2	2	1	2	2	2	1	6	3	2	2	1	3	2	2	1
$g_7$	1	1	0	7	1	1	1	7	2	2	2	2	2	2	2	1	7	4	3	3	2	3	2	2	1
$g_8$	1	1	1	8	1	0	0	8	2	2	2	2	1	1	1	1	8	2	2	2	2	1	1	1	1
				9	0	1	0	9	2	2	1	1	2	2	1	1	9	2	2	1	1	2	2	1	1
				10	0	0	1	10	2	1	2	1	2	1	2	1	10	2	1	2	1	2	1	2	1
				11	1	1	0	11	2	2	2	2	2	2	1	1	11	3	3	2	2	2	2	1	1
				12	1	0	1	12	2	2	2	2	2	1	2	1	12	3	2	3	2	2	1	2	1
				13	0	1	1	13	2	2	2	1	2	2	2	1	13	3	2	2	1	3	2	2	1
				14	1	1	1	14	2	2	2	2	2	2	2	1	14	4	3	3	2	3	2	2	1

<sup>a</sup>The 2 represents the label we are using for the threshold in Item 1. Therefore, the whole line to label the first threshold of Item 1 is 2. More importantly, each parameter with a label of 2 in Item 1 will be equated, thus providing the same value for the estimated threshold. The label 2 between each item is different; the 2 is the label, not the number.

Columns 2–4 give the attribute profiles for  $g_1$ – $g_8$ . Persons in  $g_1$  (0,0,0) have not mastered any of the three attributes; persons in  $g_2$  (0,0,1) have mastered  $\alpha_3$  but have not mastered  $\alpha_1$  and  $\alpha_2$ ; and so on for the other classes. Note that we follow substantive theories to set  $g$  classes conditional on Q-matrix ( $2^K$ ) when saturated by QAMRM *a priori*, if the class has almost no examinees in practice (de la Torre and Lee, 2010; Templin and Bradshaw, 2014), and we can thus cancel out the class.

Columns 6–8 show the Q-matrix for the 14 items. For example, item 1 measures  $\alpha_1$ , item 4 measures  $\alpha_1$  and  $\alpha_2$ , item 14 measures  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ .

Columns 10–17 list hypothetical difficulties of the 14 items for the eight classes when the attributes are non-compensatory. For example, the difficulty of item 1 is 2 for  $g_1$ – $g_4$  but 1 for  $g_5$ – $g_8$ . Because persons in  $g_1$ – $g_4$  have not mastered the attribute that item 1 measures ( $\alpha_1$ ), the item difficulty for them would be equally high. On the contrary, persons in  $g_5$ – $g_8$  have mastered  $\alpha_1$ , so the item difficulty for them would be equally low. That is, although there are eight classes, item 1 has only two difficulties, one for  $g_1$ – $g_4$  and the other for  $g_5$ – $g_8$ . Likewise, item 4 have difficulty of 2 for  $g_1$ – $g_6$  but 1 for  $g_7$ – $g_8$ . Persons in  $g_1$ – $g_6$  have not mastered all attributes that are measured by item 4 ( $\alpha_1$  and  $\alpha_2$ ) so the item difficulty for them would be equally high; on the contrary, persons in  $g_7$ – $g_8$  have mastered both  $\alpha_1$  and  $\alpha_2$  so the item difficulty for them would be equally low. The other items can be interpreted similarly.

Columns 19–26 list hypothetical difficulties of the 14 items for the eight classes when the attributes are compensatory. Item 4 have difficulty of 3 for  $g_1$  and  $g_2$ , and 2 for  $g_3$ – $g_6$ , and 1 for  $g_7$  and  $g_8$ . Persons in  $g_1$  and  $g_2$  have not mastered any of the attributes that item 4 measures ( $\alpha_1$  and  $\alpha_2$ ) so the item difficulty for them would be equally high; persons in  $g_3$ – $g_6$  have mastered

one of  $\alpha_1$  and  $\alpha_2$  so the item difficulty for them would be equally median; persons in  $g_7$  and  $g_8$  have mastered both  $\alpha_1$  and  $\alpha_2$  so the item difficulty for them would be equally low. In other words, although there are eight class, item 4 has three difficulties, one for  $g_1$  and  $g_2$ , one for  $g_3$ – $g_6$ , and the other for  $g_7$  and  $g_8$ , and the three difficulties are expected to be ordered. The other items can be interpreted similarly.

The 14 item parameters were specified according to the Q-matrix. The Q-matrix is like a bridge to connect different latent classes together and sets item parameter constraints across different latent classes *a priori*, like anchor items in different latent classes, hence the Q-matrix in QAMRM can be considered as the “anchor attribute.”

For illustrative purposes, the item difficulties in the **Table 1** are set as integers, when in reality, they can be real numbers. However, the ordinal nature of  $4 > 3 > 2 > 1$  is expected. If the Q-matrix was not adopted, there are 112 item parameters (assuming there were eight latent classes and 14 item parameters in each latent class), while according to the Q-matrix, in the non-compensatory model, each of the 14 items has two difficult parameters for the eight classes, so the total number of difficulty parameters is 28, in the compensatory model, six, six, and two items have two, three, and four difficult parameters for the eight classes, respectively, so the total number of difficulty parameters is 38.

## SIMULATION STUDY

The primary goal of this section is to demonstrate that when the QAMRM fits the item responses, through the Q-matrix setting invariant items, all model parameter estimates across latent

classes are on a common scale and will be invariant regardless of the nature of the latent ability distributions across latent classes.

In the mixture Rasch model for binary data as described by Rost (1990) who used the model constraint  $\sum \delta_{ig}=0$  in MRM, where  $\delta_{ig}$  the item difficulty of the  $i$ th item is in the  $g$ th latent class and the summation is over items at a given  $g$ , indicating the summation of group specific item difficulty parameters over items is 0 within a latent group, and compared item profiles to characterize latent groups. The constraint can be used for scale comparability only when there is no mean difference in a continuous latent variable. Because we do not know the “true” mean difference in real data sets, the constraint cannot be sufficient for all empirical data sets (Cho et al., 2016). An alternative to identifying a common scale is by making the mean of the item parameters on latent ability zero (Wu et al., 1998), and when calculating every item difficulty parameter in latent classes conditional on Q-matrix, the QAMRM uses this setting.

Several aspects of the simulated data were held constant: the number of attributes was fixed to  $K = 3$ , test length to  $J = 14$ ; the item parameters were generated as follows. For non-compensatory QAMRM (in Table 1 columns 10–17), each item had two levels of difficulty and the item parameter was set at either  $-2$  or  $2$ . For compensatory QAMRM (in Table 1 columns 19–26), an item could have 2–4 levels of difficulty. When there were two levels (items 1–3, 8–10), the item parameter was set at  $-2$  or  $2$ ; When there were three levels (items 4–6, 11–13), the item parameter was set at  $-2, 1, \text{ or } 2$ . When there were four levels (items 7, 14), the item parameter was set at  $-2, -1, 1, \text{ or } 2$ . The eight classes were uniformly distributed, meaning that the mixing proportion for each class was 12.5%. There were three sample sizes (1,000, 2,000, and 4,000) for the latent ability distributions that involve the QAMRM model, with ability distributions with means  $\mu_0=0.0$  and a common standard deviation  $\sigma_0=1.0$  been used. Note that the mixture Rasch model is employed here for discussion, where the slope parameters of items in the item response function is unity. Therefore, the difference in  $\sigma_g^2$  does not pose a problem in establishing a common scale between latent classes (Paek and Cho, 2015).

A total of 1,000 replications were generated under each condition. The item parameters were estimated via an EM implementation of the marginal maximum likelihood estimation (MMLE/EM) that was implemented using the computer program Mplus (Muthén and Muthén, 2021). In the EM estimation,

maximum likelihood optimization was done in two stages. In the initial stage, 20 random sets of initial values were generated. An optimization was carried out for 10 iterations using each of the 20 random sets of initial values. The final values from the four optimizations with the highest log-likelihoods were used as the starting values in the final stage optimizations (Muthén and Muthén, 2021). The problems with the local maximum in the QAMRM did not occur because the number of classes was

TABLE 3 | ECPE Q-Matrix and the skill profile.

Class	Skill profile ( $\alpha_1, \alpha_2, \alpha_3$ )			Item	Q-matrix		
	$\alpha_1$	$\alpha_2$	$\alpha_3$		$\alpha_1$	$\alpha_2$	$\alpha_3$
$g_1$	0	0	0	1	1	1	0
$g_2$	0	0	1	2	0	1	0
$g_3$	0	1	0	3	1	0	1
$g_4$	0	1	1	4	0	0	1
$g_5$	1	0	0	5	0	0	1
$g_6$	1	0	1	6	0	0	1
$g_7$	1	1	0	7	1	0	1
$g_8$	1	1	1	8	0	1	0
				9	0	0	1
				10	1	0	0
				11	1	0	1
				12	1	0	1
				13	1	0	0
				14	1	0	0
				15	0	0	1
				16	1	0	1
				17	0	1	1
				18	0	0	1
				19	0	0	1
				20	1	0	1
				21	1	0	1
				22	0	0	1
				23	0	1	0
				24	0	1	0
				25	1	0	0
				26	0	0	1
				27	1	0	0
				28	0	0	1

$\alpha_1$ , Morphosyntactic rules;  $\alpha_2$ , Cohesive rules;  $\alpha_3$ , Lexical rules.

TABLE 2 | Bias and 95% coverage rate for the item parameters in the simulation study.

Sample size	Non-compensatory		Compensatory	
	Bias	Coverage	Bias	Coverage
1,000	0.01	0.95	0.00	0.96
2,000	0.00	0.95	0.00	0.95
4,000	0.00	0.95	0.00	0.95

The detail simulation results about bias and coverage rate for individual parameters can be found through the link: <https://www.dropbox.com/sh/b3sb3ju4swk1pmi/AABpbXMRlXfj52PflHmNbi0Za?dl=0>.

TABLE 4 | Comparisons of model-data fit among the QAMRM, LCDM.

Model	AIC	BIC	ABIC
LCDM <sup>a</sup>	85641.43	86125.81	85868.44
QAMRM	85131.55	85568.09	85336.14
Hierarchical LCDM <sup>b</sup>	85638.63	86045.08	85829.21
Hierarchical QAMRM	85125.80	85538.42	85319.18

<sup>a</sup>LCDM = log-linear cognitive diagnostic model; More details on these models can be found in Templin and Hoffman (2013).

<sup>b</sup>Hierarchical LCDM; More details on these models can be found in Templin and Bradshaw (2014).



**TABLE 5 |** Skill profile distributions obtained for the QAMRM, LCMD.

Skill profile	QAMRM	LCMD <sup>a</sup>	Skill profile	Hierarchical LCMD <sup>b</sup>	Hierarchical QAMRM
(0,0,0)	0.44	0.30	(0,0,0)	0.34	0.44
(0,0,1)	0.11	0.13	(0,0,1)	0.11	0.08
(0,1,0)	0.00	0.01	(0,1,1)	0.18	0.13
(0,1,1)	0.10	0.18	(1,1,1)	0.38	0.34
(1,0,0)	0.00	0.01			
(1,0,1)	0.01	0.02			
(1,1,0)	0.00	0.01			
(1,1,1)	0.33	0.35			

<sup>a</sup>Templin and Hoffman (2013).

<sup>b</sup>Templin and Bradshaw (2014).

**TABLE 6 |** ECPE Q-matrix and hierarchical QAMRM item difficulty parameter estimates.

Item	Skill			Skill profile ( $\alpha_1, \alpha_2, \alpha_3$ )		
	$\alpha_1$	$\alpha_2$	$\alpha_3$	(0,0,0)	(0,0,1)	(0,1,1)
1	1	1	0	-1.98	-1.98	-2.04
2	0	1	0	-2.30	-2.30	-3.42
3	1	0	1	-0.01	-0.01	-0.01
4	0	0	1	-0.31	-2.54	-2.54
5	0	0	1	-2.27	-5.14	-5.14
6	0	0	1	-1.95	-4.18	-4.18
7	1	0	1	-0.43	-1.58	-1.58
8	0	1	0	-3.09	-3.09	-4.93
9	0	0	1	-0.68	-2.14	-2.14
10	1	0	0	-0.42	-0.42	-0.42
11	1	0	1	-0.42	-1.86	-1.86
12	1	0	1	2.68	0.30	0.30
13	1	0	0	-1.45	-1.45	-1.45
14	1	0	0	-0.68	-0.68	-0.68
15	0	0	1	-2.20	-5.01	-5.01
16	1	0	1	-0.37	-1.62	-1.62
17	0	1	1	-2.94	-2.94	-4.54
18	0	0	1	-2.11	-3.70	-3.70
19	0	0	1	-0.14	-2.83	-2.83
20	1	0	1	1.88	0.49	0.49
21	1	0	1	-0.83	-2.46	-2.46
22	0	0	1	1.06	-2.56	-2.56
23	0	1	0	-1.62	-1.62	-4.31
24	0	1	0	0.46	0.46	-1.12
25	1	0	0	-0.50	-0.50	-0.50
26	0	0	1	-0.86	-1.98	-1.98
27	1	0	0	1.09	1.09	1.09
28	0	0	1	-1.35	-3.97	-3.97

The syntax and detail analysis results can be found in: <https://www.dropbox.com/sh/b3sb3ju4swk1pmi/AABpbXMLXFji52PffHmNbi0Za?dl=0>.

For items 1, 3, 11–12, 16–17, and 20–21, by means of the Q-matrix we can see all of these items measuring two attributes, with those inside QAMRM compensatory. Take item 1 for example; item 1 measures attributes  $\alpha_1$  and  $\alpha_2$ ; if the examinee masters attribute  $\alpha_1$  and  $\alpha_2$ , he will have a high probability to answer the item 1 correctly, with the item difficulty in item 1 being -3.99, while if the examinee does not master attributes  $\alpha_1$  and  $\alpha_2$  at the same time, and just masters attribute  $\alpha_2$  compensatorily, he will have a moderate probability to answer the item correctly, with the item difficulty becoming -2.04. However, if the examinee does not master attributes  $\alpha_1$  and  $\alpha_2$ , he will have a low probability of answering the item correctly, with the item difficulty becoming -1.98. The other items can be interpreted similarly.

If the examinee doesn't master any skill, he will fall into latent class  $(\alpha_1, \alpha_2, \alpha_3) = (0, 0, 0)$  and will have the lowest probability of correctly answering all the items. If the examinee just masters the  $\alpha_3$  skill, but does not master the  $\alpha_1$  and  $\alpha_2$  skills, the examinee will fall into  $(\alpha_1, \alpha_2, \alpha_3) = (0, 0, 1)$ , and the examinee will have a higher probability of correctly answering items 4–6, 9, 15, 18–19, 22, 26, 28 than the examinee who hasn't mastered the  $\alpha_3$  skill. The details can be seen in Table 6. The other items can be interpreted similarly.

specified by the user rather than explored from the data as in finite mixture models.

The parameters in QAMRM included the mixing proportion, latent class membership conditional on the attribute profile specific item parameters, and the population parameters of latent class conditional on the attribute profile specific continuous latent variable. In the QAMRM, the Q-matrix was adopted to constrain the item parameters to be invariant across latent classes, which could reduce a large number of item parameters and improve parameter estimation (see **Table 1**). Take non-compensatory QAMRM in the simulation as an example, the total parameters to be estimated were 36, including seven mixing proportion parameters (the mixing proportion parameters should add up to 1, so only seven parameters could be estimated when there were three latent attributes and eight latent profiles), one variance parameter, and 28 item parameters for the 14 items which were specified according to the Q-matrix. If the Q-matrix was not adopted, the model became the MRM, which would estimate 120 parameters, including 112 item parameters (assuming there were eight latent classes and 14 item parameters in each latent class), seven mixing proportion parameters, and one variance parameter. Such a large number of parameters would require a large sample size, which would be a practical constraint.

To evaluate the parameter recovery, we computed the bias, the 95% coverage rate for the item parameter estimates in **Table 2**. Due to enhance readability, we do not report bias and coverage rate for individual parameters; rather, we show the mean bias and mean coverage rate across all parameters in **Table 2**.

The EM estimation method yielded very small bias. Besides, the mean coverage was very close to 95%. The results of the simulated data analysis indicate that the invariance property of the QAMRM model is absolute in that the parameter estimates were obtained using different calibration samples. By means of the Q-matrix setting invariant items *a priori* in QAMRM, all model parameter estimates across latent classes to be on a common scale, which does not require any transformation for them to be comparable.

## REAL DATA ANALYSIS

We used the Certificate of Proficiency in English (ECPE) data, which is available in the R package CDM (Robitzsch et al., 2011–2014), to demonstrate the advantages of the QAMRM over the LCDM. The ECPE data consist of responses from 2,922 test-takers to 28 items, with each item measuring one or two out of three skills. The data has been analyzed with the LCDM by Templin and Hoffman (2013) and Templin and Bradshaw (2014) and with the GDM by von Davier (2014). As shown previously, when analyze the ECPE data, the LCDM and GDM are mathematically equivalent (von Davier, 2014), hence we fit the QAMRM to the data using Mplus and compared the results with those under the LCDM (Templin and Hoffman, 2013; Templin and Bradshaw, 2014).

The Q-matrix used in ECPE example was the result of psychometric analyses on the ECPE by Buck and Tatsuoka (1998).

The analyses showed that items of the test were likely to measure three distinct skills. The left side of **Table 3** shows the skill profile and the right side of **Table 3** shows Q-matrix that maps each item to the three skills. As shown, eight items measure only one skill, seven items measure two skills, and zero items measure three skills. The morphosyntactic ( $\alpha_1$ ), cohesive ( $\alpha_2$ ), and lexical ( $\alpha_3$ ) skills were each measured by 13, 6, and 18 items, respectively.

Therefore, we only report the skill distributions with latent class pattern and model fit results with the values published by Templin and Hoffman (2013) and Templin and Bradshaw (2014) which agree with those obtained from the CDM R-package (Robitzsch et al., 2011–2014).

We use AIC, BIC, and sample-size adjusted BIC (ABIC) to select the best model (Templin and Bradshaw, 2014), the information criteria selected the best model by small value. **Table 4** presents AIC, BIC, and ABIC for the QAMRM, LCDM. It appears that the QAMRM had lower AIC (85131.55–85641.43), BIC (85568.09–86125.81), and ABIC (85336.14–85868.44). **Table 5** shows the distributions of the eight skill profiles (classes) obtained from the QAMRM, LCDM. The distributions were very similar across models and only four skill profiles were substantial: (0,0,0), (1,0,0), (1,1,0), and (1,1,1).

Templin and Hoffman (2013) analyzed a sample of 2,922 examinees who took the ECPE with the non-hierarchical LCDM. But in the ECPE example, the data and results suggest a linear attribute hierarchy: Examinees must master Attribute 3 (lexical rules) before mastering Attribute 2 (cohesive rules) before mastering Attribute 1 (morphosyntactic rules). Gierl et al. (2007) call this structure a linear hierarchy, where mastery of each attribute follows a linear progression. Therefore, Templin and Bradshaw (2014) introduce the hierarchical LCDM where attribute hierarchies are present, the model fit of hierarchical LCDM shown at the bottom of **Table 4**, the hierarchical LCDM is used to test for the presence of a suspected attribute hierarchy in ECPE, through model fit which confirming the data is more adequately represented by hierarchical attribute structure when compared to a crossed, or non-hierarchical structure.

We reanalyzed the data with hierarchical QAMRM, and compare the model fit with hierarchical LCDM. Right column of **Table 5** shows the distributions of the four hierarchical skill profiles (classes) obtained from the hierarchical QAMRM, hierarchical LCDM. It also appears that the hierarchical QAMRM had lower AIC (85125.80–85638.63), BIC (85538.42–86045.08), and ABIC (85319.18–85829.21).

**Table 6** presents Q-matrix in ECPE and hierarchical QAMRM item difficulty parameter estimates. In items 4–10, 13–15, 18–19, 22–28 only one attribute is measured, with all of these items inside hierarchical QAMRM non-compensatory, having only two different kinds of item difficulty; take item 4 for example; if the examiner masters attribute  $\alpha_3$ , he will have a high probability to answer item 4 correctly, with the item difficulty being  $-2.54$ , while, on the other hand, if the examiner does not master attribute  $\alpha_3$ , he will have a low probability to answer the item correct, with the item difficulty becoming  $-0.31$ , and the other items can be interpreted similarity.

If we only use the LCDM or Hierarchical LCDM for analysis, a second calculation is still needed to find the probability

of a correct response for each item, which may not easy for practitioners.

## CONCLUSION AND DISCUSSION

The QAMRM was used to describe for modeling the Q-matrix at the mixture Rasch model. The model developed in this study used features of a Rasch model, a restricted latent class model, and a Q-matrix. The Q-matrix of the model provides an opportunity to determine the number of latent class in advance through substantive theory and not through model fitness or parameter estimation *post hoc*. Information in the Q-matrix can be used to reveal possible differences that might be due to differences among latent classes.

A simulation study through the EM algorithm estimation was presented to investigate the performance of the model. Generated parameters were well recovered for the conditions considered. The QAMRM makes it possible to describe the differential item performance of target attributes using descriptions of Q-matrix characteristics associated with the items compared with characteristics associated with other items not in the same latent classes. This description can then be used to provide the Q-matrix with a framework within which to compare the results in their latent classes and in the other latent classes. Examiners in each of the latent class can be characterized by differences in attribute, as well as by differences in response strategies, particularly at the end of the test.

The real data comparison performed between the QAMRM, LCDM, by means of the ECPE data, shows that when a Rasch model is included inside the diagnostic classification models, QAMRM achieves a more desirable result and has better fit indices than the LCDM variants.

If we want to provide a single, continuous estimate of overall ability and classify the subjects at the same time, we should consider the QAMRM rather than mixture Rasch model or DCM. In QAMRM, the Q-matrix sets class invariant items a priori, if the Q-matrix design is not correct, then the analysis in the

QAMRM will be wrong. Kopf et al. (2015) who have discussed anchoring strategies in details which can help to correct the Q-matrix design. Future research can focus on the misspecify Q-matrix design with the model unfit of the QAMRM.

On the other hand, future research can focus on the estimation limitations of the QAMRM. Specifically, as the number of attributes included in the Q-matrix increases and as its complexity increases, the number of parameters estimated by this model will also increase. In these cases, expectations of its performance in estimated attribute mastery and item parameters must be explored. In addition, model comparisons using common indices such as the AIC, BIC, and ABIC must continue to be explored, which could result to clear guidelines for model identification. Finally, possible expansions of this model such as the addition of a continuous ability measure to imply an incomplete Q-matrix (much like what is used in the testlet IRT) will be explored.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://cran.r-project.org/web/packages/CDM/index.html>.

## AUTHOR CONTRIBUTIONS

M-CT: writing and analyzing, 80%. W-CW: correction and feedback, 20%. Both authors contributed to the article and approved the submitted version.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.564976/full#supplementary-material>

## REFERENCES

- Adams, R., Wilson, M., and Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Appl. Psychol. Meas.* 21, 1–23.
- Alexeev, N., Templin, J., and Cohen, A. S. (2011). Spurious latent classes in the mixture rasch model. *J. Edc. Meas.* 48, 313–332. doi: 10.1111/j.1745-3984.2011.00146.x
- Bradshaw, L., and Madison, M. J. (2016). Invariance properties for general diagnostic classification models. *Int. J. Testing* 16, 99–118. doi: 10.1080/15305058.2015.1107076
- Bradshaw, L., and Templin, J. (2014). Combining item response theory and diagnostic classification models: a psychometric model for scaling ability and diagnosing misconceptions. *Psychometrika* 79, 403–425. doi: 10.1007/s11336-013-9350-4
- Buck, G., and Tatsuoaka, K. K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing* 15, 119–157. doi: 10.1177/026553229801500201
- Cho, S.-J., Suh, Y., and Lee, W.-Y. (2016). An NCME instructional module on latent DIF analysis using mixture item response models. *Edc. Meas. Issues Practice* 35, 48–61. doi: 10.1111/emip.12093
- Choi, H. J. (2010). *A model that combines diagnostic classification assessment with mixture item response theory models*. Georgia: University of Georgia.
- de la Torre, J., and Lee, Y.-S. (2010). A note on the invariance of the DINA model parameters. *J. Edc. Meas.* 47, 115–127. doi: 10.1111/j.1745-3984.2009.00102.x
- Embretson, S. E., and Yang, X. (2013). A multicomponent latent trait model for diagnosis. *Psychometrika* 78, 14–36.
- Fischer, G. H. (1973). The linear logistic model as an instrument in educational research. *Acta Psychol.* 37, 359–374. doi: 10.1016/0001-6918(73)90003-6
- Gibbons, R. D., and Hedeker, D. (1992). Full-information item bi-factor analysis. *Psychometrika* 57, 423–436. doi: 10.1007/BF02295430
- Gierl, M. J., Leighton, J. P., and Hunka, S. M. (2007). "Using the attribute hierarchy method to make diagnostic inferences about respondents' cognitive skills," in *Cognitive diagnostic assessment for education: theory and applications*, eds J. P. Leighton and M. J. Gierl (Cambridge: Cambridge University Press), 242–274. doi: 10.1017/CBO9780511611186.009
- Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.
- Henson, R., Templin, J., and Willse, J. (2009). Defining a family of cognitive diagnosis models using log linear models with latent variables. *Psychometrika* 74, 191–210. doi: 10.1007/s11336-008-9089-5



- Hong, H., Wang, C., Lim, Y., and Douglas, J. (2015). Efficient models for cognitive diagnosis with continuous and mixed-type latent variables. *Appl. Psychol. Meas.* 39, 31–43.
- Hoskens, M., and De Boeck, P. (2001). Multidimensional componential item response theory models for polytomous items. *Appl. Psychol. Meas.* 25, 19–37. doi: 10.1177/01466216010251002
- Kelderman, H., and Rijkes, C. P. M. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika* 59, 149–176. doi: 10.1007/BF02295181
- Kopf, J., Zeileis, A., and Strobl, C. (2015). Anchor selection strategies for DIF analysis: review, assessment, and new approaches. *Edc. Psychol. Meas.* 75, 22–56. doi: 10.1177/0013164414529792
- Lazarsfeld, P. F., and Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Mislevy, R. J., and Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika* 55, 195–215.
- Muthén, L., and Muthén, B. (2021). *Mplus [Computer software]*. Los Angeles, CA: Author.
- Paek, L., and Cho, S. J. (2015). A note on parameter estimate comparability across latent classes in mixture IRT modeling. *Appl. Psychol. Meas.* 39, 135–143. doi: 10.1177/0146621614549651
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen and Lydiche.
- Robitzsch, A., Kiefer, T., George, A. C., and Ünlü, A. (2011–2014). *R-package documentation for CDM: Cognitive diagnosis modeling*. Available Online at: <http://cran.r-project.org/web/packages/CDM/index.html>
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Appl. Psychol. Meas.* 14, 271–282. doi: 10.1177/014662169001400305
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *J. Edc. Meas.* 20, 345–354. doi: 10.1111/j.1745-3984.1983.tb00212.x
- Templin, J., and Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika* 79, 317–339. doi: 10.1007/s11336-013-9362-0
- Templin, J., and Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus. *Edc. Meas. Issues Prac.* 32, 37–50. doi: 10.1111/emip.12010
- von Davier, M. (2005). *A general diagnostic model applied to language testing data (ETS Research Rep. No. RR-05-16)*. Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.2005.tb01993.x
- von Davier, M. (2014). *The log-linear cognitive diagnostic model (LCDM) as a special case of the general diagnostic model (GDM) (ETS Research Rep. No. RR-14-40)*. Princeton, NJ: Educational Testing Service. doi: 10.1002/ets2.12043
- von Davier, M., and Yamamoto, K. (2004). Partially observed mixtures of IRT models: an extension of the generalized partial-credit model. *Appl. Psychol. Meas.* 28, 389–406. doi: 10.1177/0146621604268734
- von Davier, M., and Yamamoto, K. (2007). "Mixture-distribution and HYBRID Rasch models," in *Multivariate and mixture distribution Rasch models*, eds M. von Davier and C. H. Carstensen (New York, NY: Springer), 99–115. doi: 10.1007/978-0-387-49839-3\_6
- Whitley, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika* 45, 479–494. doi: 10.1007/BF02293610
- Wright, B. D., and Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago, IL: MESA Press.
- Wu, M. L., Adams, R. J., and Wilson, M. R. (1998). *Conquest: Generalized item response modeling software*. Camberwell, Victoria: Australian Council for Educational Research.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Tseng and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.