International Journal of Psychology and Educational Studies, 2022, 9(4), 1187-1196

www.ijpes.com

International Journal of Psychology and Educational Studies

ISSN: 2148-9378

Performance of the Q-Matrix Validation Methods in the DINA Model

Ömür Kaya KALKAN[1], Emre TOPRAK[2]

[1] *Pamukkale University, Denizli, Turkey* *0000-0001-7088-4268*

[2] *Erciyes University, Kayseri, Turkey* *0000-0002-4131-4888*

ARTICLE INFO

Article History
Received 19.01.2022
Received in revised form 21.05.2022
Accepted 20.06.2022
Article Type: Research Article

ABSTRACT

All cognitive diagnostic models that evaluate educational test data require a Q-matrix that combines every item in a test with the required cognitive skills for each item to be answered correctly. Generally, the Q-matrix is constructed by education experts' judgment, leading to some uncertainty in its elements. Various statistical methods are suggested to validate misspecifications in the Q-matrix. This paper evaluates the performance of the Q-matrix validation methods, the sequential expectation-maximization-based δ-method (SEM δ-method), and the Q-matrix refinement (QMR) method using a study with real data and simulations. The simulation design results showed that the misspecification percentage and the length of the test had a small or no effect on the mean q-entry recovery rates (MRRs) of both methods, while the increase in sample size had an improving effect. The MRRs of both methods decreased when the number of attributes and guessing (g) - slip (s) parameters increased. According to simulation study results, the QMR method performed better than the SEM δ-method. For the q-matrix validation, it can be suggested that CDM practitioners prioritize the QMR method and use a sample size of 1,000. On the other hand, the real data results revealed that the MRRs of both methods were at the base rates. This result highlights the need for further research on method comparison, specifically for real-world data applications where the number of attributes is relatively large and the test duration is short.

Keywords:
Q-matrix validation, Q-matrix refinement method, Q-matrix misspecification, sequential EM-based δ-method, DINA model.

1. Introduction

Over the past two decades, educational testing researchers and practitioners have paid more attention to cognitive diagnostic models (CDMs) due to their positive impacts on instruction and learning (de la Torre & Lee, 2013). CDMs have a great potential to specify students' strong and weak points and provide rich feedback on educational settings (de la Torre, 2008, 2009a). These models offer feedback on whether examinees master multiple fine-grained skills needed to resolve items in a test (de la Torre, 2009a, 2009b). Therefore, CDMs can provide more in-depth information about the skills each student has or does not have, unlike the classical test theory or item response theory, where a simple overall score is derived (DeCarlo, 2012; de la Torre et al., 2010; de la Torre & Lee, 2013). This information provided by CDMs can facilitate specific improvements to students' individual needs, better design of instruction and proper measurement of student development (de la Torre, 2009b).

In the CDM framework, attributes generally refer to the basic cognitive processes, knowledge representation, or skills required to correctly solve test problems (de la Torre, 2009a, 2009b; Leighton et al., 2004). Typically, all CDMs require a Q-matrix construction that shows these attributes' relationship to each test item (de la Torre et al., 2010; Tatsuoka, 1983). A Q-matrix consisting of 1 and 0's has items in its rows and attributes in its

[1]Corresponding author's address: Faculty of Education, Erciyes University, Kayseri /Turkey
e-mail: etoprak@erciyes.edu.tr

Citation: Kurt, U. (2022). Performance of the Q-matrix validation methods in the DINA model. *International Journal of Psychology and Educational Studies*, *9*(4), 1187-1196. https://dx.doi.org/10.52380/ijpes.2022.9.4.795

columns. If the corresponding attribute is required for a test item to be answered correctly, it is equal to 1 in the Q matrix and equals 0 otherwise.

The Q-matrix is crucial in test construction as it represents the attribute blueprint of the test or cognitive specifications (Leighton, Gierl, & Hunka, 2004). However, with the type of skills required for a specific item set, the number of skills involved, and their combination, the correct specification of a Q-matrix is not a simple assignment and it can lead to some uncertainty (DeCarlo, 2012). The Q-matrix is generally constructed by education experts' judgment, but “true skills” may differ from the labels these experts give. Changing some skill labels can entirely change the Q-matrix. Therefore, in terms of most educational tests, the Q-matrix is unknown, leading to the risk of Q-matrix misspecification (Chiu, 2013; DeCarlo, 2011). Misspecification of the Q-matrix can negatively affect estimates of item parameters, classification of respondents, and latent class sizes (DeCarlo, 2011; de la Torre et al., 2010). In recent years, a limited number of methods have been proposed for determining and refining a misspecified Q-matrix, such as the sequential expectation- maximization-based δ-method (SEM δ; de la Torre, 2008), the Bayesian method (DeCarlo, 2012), the residual sum of squares-based method (Chiu, 2013), a nonparametric method (Chiu & Douglas, 2013), a discrimination index-based method (Ma & de la Torre, 2020). However, the Q-matrix validation methods used in the current study are limited to the SEM δ-method and QMR methods. The deterministic input, noisy output, “and” gate (DINA; Haertel, 1989; Junker & Sijtsma, 2001) model, SEM δ-method (de la Torre, 2008), and QMR method (Chiu, 2013) were briefly introduced below.

1.1. The DINA Model

The DINA model is one of the most widely used CDMs due to its simplicity, parsimony, and ease of interpretation (de la Torre, 2008; de la Torre & Lee, 2010; George & Robitzsch, 2014; Huang & Wang, 2014). The model requires only guessing (g) and slip (s) parameters for each test item regardless of the attributes number (de la Torre, 2008; de la Torre & Douglas, 2008; de la Torre & Lee, 2010). The DINA model presumes that examinees must have all the specified attributes for the test item in the Q-matrix to respond to a test item correctly and is considered a type of noncompensatory model (George & Robitzsch, 2014; Huang & Wang, 2014). Notably, the DINA model is most suitable for use where the conjunction of equally important attributes is required (de la Torre & Douglas, 2008).

Let $\alpha_i = \{\alpha_{ik}\}$ be the respondent i’s binary attribute vector, $k = 1,\ldots, K$, where a 1 on the kth element indicates the presence of attribute k and 0, absence of the attribute. Let Q be a J by K matrix and the element q_{jk} denoting whether skill k is required to respond to item j correctly (if the skill is required $q_{jk} = 1$, otherwise $q_{jk} = 0$). A latent response variable η_{ij} that the considered the deterministic component in the DINA model is formulated as:

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}. \tag{1}$$

In Equation 1, η_{ij} is equal to 1 if an examinee i has all the required attributes for item j and η_{ij} is equal to 0 otherwise. The item response function is formulated as:

$$P\left(Y_{ij} = 1 \middle| \alpha_i\right) = g_{\mathrm{j}}^{1-\eta_{ij}} (1 - s_j)^{\eta_{ij}}. \tag{2}$$

In Equation 2, Y_{ij} is the observed response of examinee i to item j, and the slip parameter s_j is the probability of an answer to the item j incorrectly the examinee who possesses all the required attributes for item j; conversely, the g_j is the probability of a correct response to item j the examinee who does not have at least one or more of the requisite attributes for item j. To put it differently, Equation 2 points out the probabilistic nature of the DINA model.

1.2. Sequential Expectation-Maximization-Based δ-Method

de la Torre (2008) introduced a sequential expectation-maximization-based δ-method (SEM δ-method) that can be considered as an item discrimination index, φ, to identify misspecifications in the Q-matrix for the DINA model. φ_j determines the correct q-vector by maximizing the difference between the success probabilities of the group of individuals who possess all the required attributes to answer item j correctly ($\eta_j = 1$) and the group of individuals who lack at least one of the required attributes ($\eta_j = 0$).

de la Torre (2008) comprehensively reviewed the exhaustive search algorithm (ESA) and the sequential search algorithm (SSA) for the implementation of the SEM δ-method. The author noted that computing φj for each item is burdensome when the number of attributes (K) increases because the K patterns increase exponentially. In addition, the ESA algorithm is applicable and efficient for reasonably small K values. On the other hand, the SSA is an alternative algorithm to the ESA, and it does not require computing φ_j for the 2^K - 1 possible q vectors.

1.3. Q-Matrix Refinement Method

Chiu (2013) introduced the residual sum of squares (RSS) based on (i.e., nonparametric) Q-matrix refinement (QMR) method. The proposed validation method's rationale is based on minimizing the RSS computed according to observed and ideal item responses. By recalling Y_{ij} and η_{ij} from Equations 1 and 2, the RSS of item j for respondent i is formulated as:

$$RSS_{ij} = (Y_{ij} - \eta_{ij})^2. \tag{3}$$

Then, the RSS of item j across all respondents is formulated as:

$$RSS_j = \sum_{i=1}^{N}(Y_{ij} - \eta_{ij})^2 = \sum_{m=1}^{2^K}\sum_{i\epsilon C_m}(Y_{ij} - \eta_{jm})^2. \tag{4}$$

In Equation 4, the latent proficiency-class m is denoted by C_m and N is the number of respondents. Chiu (2013) noted that the indicator of the latent response to item j was changed to "i_j" to "j_m" because ideal item responses are class-specific. This means that every respondent in the same latent class is thought to have the same ideal response to a test item. Chiu (2013) successfully applied the QMR method to the DINA model. The QMR method is fundamentally a nonparametric classification procedure introduced by Chiu and Douglas (2013) for cognitive diagnosis. Besides, de la Torre and Chiu (2016) proposed the discrimination index used with a broad class of CDMs covered by the generalized DINA (G-DINA) model. Ma and de la Torre (2020) recently introduced a stepwise Q-matrix validation method using a sequential G-DINA model for graded response data.

Although different Q-matrix validation methods have been proposed, the factors affecting these proposed Q-matrix validation methods' performances are not apparent. Studies comparing these methods' performance are minimal (e.g., Chen, 2017; Terzi & de la Torre, 2018). Notably, de la Torre (2008) noted that his work represents the first step in the empirical validation of a Q matrix, and there is still much work to be done in this area. For example, the author stated that more conditions such as the degree of Q-matrix misspecification, test length, and sample size should be investigated to determine the applicability of his method in different situations. In addition, the author indicated that real data covering more expansive areas should be analysed to obtain additional findings on how the method works in practice.

Furthermore, Terzi and de la Torre (2018) stated that research to be conducted based on the different number of attributes and other real datasets would contribute to understanding the performance of the methods. As a result, the goal of this study is to compare the performance of the SEM δ-method (de la Torre, 2008) and the QMR method (Chiu, 2013) under various study conditions, such as the number of attributes, examinees, test lengths, guessing (g) and slip (s) parameters, and the percentage of misspecified q-entries.

2. Methodology

2.1. Simulation Study

Simulation study design was crafted from six variables: (a) Q-matrix validation methods (SEM δ-method and QMR method), (b) number of attributes (K = 3, 4, 5), (c) number of examinees (N = 250, 500, 1000, 2000, 4000), (d) test lengths (J = 20, 40), (e) guessing and slip parameters (g = s = 0.1, 0.2, 0.3, 0.4, 0.5), (f) misspecified q-entries percentage (10%, 20%). The Q-matrix, in which 20 items are associated with 3, 4, and 5 attributes (Chiu, 2013), is presented in Table 1. In this study, similar to Chiu (2013), the number of attributes (K = 3, 4, 5) and the misspecified q-entries percentages (10%, 20%) were considered. For misspecified q-entries, for instance, if a Q-matrix has 10 percent misspecified q-entries for J = 20 and K = 3, 6 of 60 entries were changed randomly by producing over-specification or under-specification q-entries. Similar to Terzi and de la Torre (2018), two restrictions were imposed on changing q-vectors, a maximum of two misspecified attributes were allowed in q-entries, and at least one attribute was identified as 1. In the scope of the study, the proportion of misspecified q-entries was considered rather than the specific types of misspecified q-entries (i.e., over specification or under-specification). In addition to this, the condition $g = s = 0.1$ (de la Torre & Lee, 2010) for high item quality was added to Chiu's item quality conditions ($g = s = 0.2, 0.3, 0.4, 0.5$). Furthermore, the conditions mentioned in de la Torre and Lee's (2010) study were considered in determining the test length ($J = 20, 40$). DINA model simulation studies in which different sample sizes are used can be encountered, for example, $N = 100, 500, 1000$ (Chiu, 2013), $N = 1000, 2000, 4000$ (de la Torre, Hong & Deng, 2010), and $N = 2000$ (de la Torre & Chiu, 2016; de la Torre & Douglas, 2008) etc. The present study set the sample size to $N = 250, 500, 1000, 2000$, and 4000, considering other specified research conditions.

Table 1. *The Q-Matrices for 20-Item Tests*

	Number of attributes											
Item	3			4				5				
1	1	0	0	1	0	0	0	1	0	0	0	0
2	0	1	0	0	1	0	0	0	1	0	0	0
3	0	0	1	0	0	1	0	0	0	1	0	0
4	1	1	0	0	0	0	1	0	0	0	1	0
5	1	0	1	1	0	0	0	0	0	0	0	1
6	0	1	1	0	1	0	0	1	1	0	0	0
7	1	0	0	0	0	1	0	0	1	0	1	0
8	0	1	0	0	0	0	1	1	0	0	0	1
9	0	0	1	1	1	0	0	0	0	0	1	1
10	1	1	0	1	0	1	0	0	0	1	1	0
11	1	0	1	1	0	0	1	0	1	1	1	0
12	0	1	1	0	1	1	0	1	0	1	0	1
13	1	0	0	0	1	0	1	0	1	0	1	1
14	0	1	0	0	0	1	1	0	0	1	1	1
15	0	0	1	1	1	1	0	1	1	1	0	0
16	1	1	0	1	1	0	1	1	1	1	0	1
17	1	0	1	1	0	1	1	0	1	1	1	1
18	0	1	1	0	1	1	1	1	1	0	1	1
19	1	1	1	1	1	1	1	1	0	1	1	1
20	1	1	1	1	1	1	1	1	1	1	1	1

The 20-item Q-matrices were doubled to construct Q-matrices for the 40-item tests. Then, the original Q-matrices were used for generating simulated item responses. The original Q-matrix's 10% or 20% of q-vectors were changed randomly to create misspecified Q-matrices. 100 data sets were generated for each of the 3 (number of attributes) × 5 (number of examinees) × 2 (test lengths) × 5 (g and s parameters) × 2 (misspecified q-entries percentage) = 300 design conditions. Then, each data set was analysed using the SEM δ-method (de la Torre, 2008) and the QMR method (Chiu, 2013). Analyses were done using R (R Core Team, 2020), the CDM (George, Robitzsch, Kiefer, Groß, & Ünlü, 2016) and the NPCD (Zheng & Chiu, 2019) packages. The mean q-entry recovery rates (MRRs) were computed by comparing the Q-matrices proposed by both methods with the original Q-matrices for one hundred data sets in each design condition.

The MRR is the ratio of correct q-entries in the Q-matrix proposed by the model to the total number of q-entries in the original Q-matrix. The recovery rate of the correct q-entries equals one if all elements of the Q-matrix proposed by the model and the original Q-matrix are the same (i.e., if the matrices are equal). Additionally, for design conditions where Q-matrix entries were defined with 10% and 20% misspecification, the base rates (BRs) were 0.90 and 0.80, respectively. Whereas a higher MRR than the BR is more informative, MRR close to or below the BR is less informative (Chiu, 2013).

2.2. Real Data

The present study used a mathematics achievement test comprising 11 items measuring four attributes and was administered to 2,918 6th grade students (Başokcu et al., 2018). This achievement test was constructed based on cognitive diagnostic models and required four attributes: communication and association, mathematization, reasoning, and strategy development, and use of symbolic and technical language. Due to the fact that the s parameter of item 9 was 0.97 based on the preliminary analysis, this item was excluded from the real data set and was not included in the subsequent analyses. Real data analyses on ten items were conducted, and item parameters are shown in Table 4.

3. Findings

3.1. Simulated Data

The MRRs obtained for 20 and 40 items by SEM δ-method and QMR method were presented in Tables 2 and 3, respectively. In general, MRRs obtained for the two misspecification percentages showed that the misspecification percentage had a small or no effect on the efficacy of both Q-matrix validation methods. In addition to this, the MRRs of both methods decreased when the number of attributes and g and s parameters increased. The performance of MRRs of the QMR method indicated perfect recovery for all design conditions when the g and s parameters were 0.1. However, SEM δ-method MRRs values for the same g and s values started to deteriorate generally when the number of attributes increased.

Table 2. *The MRRs for 20 Items*

		SEM δ-method					QMR method				
		N					*N*				
K	*g* and *s*	250	500	1000	2000	4000	250	500	1000	2000	4000
10% misspecification of the Q-matrix											
3	0.1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	0.3	0.96	1.00	0.97	0.97	0.97	0.97	1.00	1.00	0.99	1.00
	0.4	0.77	0.85	0.94	0.94	0.97	0.94	0.94	0.97	0.93	0.97
	0.5	0.67	0.69	0.68	0.71	0.71	0.91	0.93	0.90	0.93	0.90
4	0.1	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	0.97	0.98	0.98	0.98	0.96	1.00	1.00	1.00	1.00	1.00
	0.3	0.91	0.96	0.96	0.96	0.96	0.93	0.97	0.98	1.00	0.99
	0.4	0.75	0.82	0. 90	0.93	0.97	0.93	0.95	0.98	0.95	0.98
	0.5	0.70	0.71	0.73	0.76	0.73	0.90	0.93	0.94	0.97	0.93
5	0.1	0.97	0.96	0.98	0.98	0.98	1.00	1.00	1.00	1.00	1.00
	0.2	0.91	0.96	0.93	0.97	0.96	0.97	0.99	0.96	1.00	1.00
	0.3	0.88	0.94	0.96	0.96	0.97	0.97	0.98	0.98	0.96	0.98
	0.4	0.74	0.78	0.85	0.86	0.92	0.90	0.95	0.98	0.94	0.96
	0.5	0.70	0.72	0.70	0.73	0.71	0.88	0.92	0.93	0.94	0.92
20% misspecification of the Q-matrix											
3	0.1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.3	0.92	0.93	0.94	0.93	0.95	0.96	0.95	0.99	0.95	0.98
	0.4	0.81	0.82	0.89	0.94	0.93	0.96	0.91	0.90	0.94	0.93
	0.5	0.65	0.66	0.68	0.66	0.68	0.87	0.87	0.87	0.87	0.87
4	0.1	0.98	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	0.2	0.94	0.95	0.95	0.97	0.96	1.00	1.00	1.00	1.00	1.00
	0.3	0.89	0.94	0.95	0.96	0.94	0.92	0.95	0.94	0.95	0.97
	0.4	0.75	0.76	0.83	0.86	0.95	0.91	0.89	0.90	0.88	0.95
	0.5	0.66	0.66	0.68	0.70	0.71	0.82	0.84	0.84	0.88	0.88
5	0.1	0.92	0.94	0.92	0.96	0.93	1.00	1.00	1.00	1.00	1.00
	0.2	0.92	0.92	0.92	0.93	0.95	0.98	0.95	0.98	0.95	0.98
	0.3	0.83	0.87	0.89	0.93	0.94	0.86	0.90	0.90	0.92	0.92
	0.4	0.70	0.76	0.79	0.82	0.86	0.87	0.89	0.90	0.88	0.88
	0.5	0.68	0.69	0.67	0.69	0.72	0.82	0.87	0.88	0.86	0.94

Notably, when the g and s parameters were 0.5, the QMR method generally provided values slightly higher than the base rates. In contrast, the SEM δ-method provided values below the base rates under all design conditions. In study conditions where g and s parameters were 0.4, the MRRs of the SEM δ-method was punished severely, but the increase in sample size mitigated this effect. However, under study conditions where g and s parameters were 0.5, the increase in sample size did not show this remedial effect for the SEM

δ-method prominently. The MRRs of the SEM δ-method were more severely affected by the increase of g and s parameters than the QMR method.

Table 3. *The MRRs for 40 Items*

		SEM δ-method					QMR method				
		N					*N*				
K	*g* and *s*	250	500	1000	2000	4000	250	500	1000	2000	4000
10% misspecification of the Q-matrix											
3	0.1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.3	0.97	0.98	0.99	0.99	1.00	0.99	1.00	1.00	1.00	1.00
	0.4	0.82	0.92	0.96	0.98	0.96	0.91	0.97	0.95	0.97	0.93
	0.5	0.65	0.64	0.68	0.70	0.71	0.84	0.88	0.94	0.98	0.97
4	0.1	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	0.97	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
	0.3	0.93	0.96	0.97	0.96	0.98	0.98	0.99	1.00	1.00	1.00
	0.4	0.75	0.80	0.90	0.96	0.96	0.90	0.92	0.95	0.96	0.94
	0.5	0.66	0.67	0.69	0.69	0.72	0.83	0.88	0.93	0.92	0.95
5	0.1	0.97	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	0.91	0.95	0.96	0.97	0.98	0.99	1.00	1.00	1.00	1.00
	0.3	0.86	0.93	0.95	0.95	0.95	0.93	0.99	0.99	0.99	0.95
	0.4	0.73	0.78	0.83	0.91	0.92	0.87	0.95	0.94	0.94	0.93
	0.5	0.68	0.67	0.67	0.68	0.71	0.82	0.87	0.88	0.92	0.95
20% misspecification of the Q-matrix											
3	0.1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.3	0.95	0.97	0.98	0.96	0.97	0.99	1.00	1.00	1.00	1.00
	0.4	0.82	0.91	0.96	0.92	0.91	0.92	0.94	0.95	0.91	0.88
	0.5	0.60	0.63	0.67	0.66	0.67	0.76	0.82	0.93	0.88	0.87
4	0.1	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	0.96	0.98	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00
	0.3	0.90	0.93	0.94	0.95	0.97	0.95	0.99	1.00	1.00	1.00
	0.4	0.71	0.79	0.87	0.92	0.89	0.86	0.91	0.92	0.93	0.89
	0.5	0.65	0.65	0.67	0.68	0.68	0.81	0.84	0.87	0.88	0.90
5	0.1	0.96	0.97	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00
	0.2	0.89	0.92	0.94	0.97	0.97	0.99	1.00	1.00	1.00	1.00
	0.3	0.83	0.89	0.92	0.89	0.92	0.90	0.96	0.94	0.90	0.97
	0.4	0.73	0.73	0.79	0.88	0.88	0.87	0.88	0.85	0.91	0.89
	0.5	0.65	0.65	0.67	0.67	0.66	0.77	0.84	0.89	0.87	0.88

In general, while the other design conditions were fixed, notably under conditions where g and s were 0.5, the increase in the test length slightly increased the deterioration in SEM δ-method's MRRs. In contrast, this situation was less observed in the QMR method MRRs. Test lengths had little or no effect on the MRRs of both Q-matrix validation method. When other conditions were fixed, increasing the percentage of misspecification slightly worsened the MRR of both methods (maximum 0.07 for the SEM δ-method and 0.11 for the QMR-method).

3.2. Real Data

The fraction-subtraction test data (Tatsuoka, 1984) are widely used in cognitive diagnosis applications. This data set comprised dichotomous responses of 536 students to the 20 test items associated with eight attributes requiring the subtraction of fractions. Additionally, its arrangements consisting of 15 items associated with five attributes (Tatsuoka, 1990) were also analysed (DeCarlo, 2012; de la Torre, 2008; de la Torre & Douglas, 2008; de la Torre & Lee, 2010; Huang & Wang, 2014). However, DeCarlo (2012) points to uncertainties in the correct specification of some components in the Q-matrix of Tatsuoka's test data. In addition, Chiu (2013) stated that this Q matrix which involves eight attributes is not complete because not all possible attribute patterns are allowed to be described, and not every attribute is associated with at least one single-attribute item. The present study used a mathematics achievement test comprising ten items measuring four attributes and was administered to 2,918 6th grade students (Başokcu et al., 2018). It should also be noted that this Q-matrix is also incomplete, but the number of potential attribute patterns that cannot be distinguished by the items (i.e., 9) is much smaller than Tatsuoka's fraction subtraction data (i.e., 198). The Q-matrix, *g*, and *s* parameters are shown in Table 4.

Table 4. *Q-Matrix and Item Parameter Estimations of Mathematics Achievement Test*

	Attribute				Estimates	
Item	1	2	3	4	$\hat{g}$	$\hat{s}$
1	0	0	0	1	0.60	0.08
2	1	0	1	0	0.19	0.75
3	1	1	1	0	0.12	0.17
4	0	1	0	1	0.23	0.49
5	0	1	0	1	0.12	0.32
6	0	0	1	0	0.26	0.42
7	0	0	0	1	0.37	0.18
8	0	0	1	1	0.26	0.15
9	0	0	0	1	0.23	0.36
10	0	1	1	1	0.23	0.34

Initially, the analyses were performed using the original Q-matrix for Q-matrix validation. However, both Q-matrix validation methods did not suggest a modification to the original Q-matrix. Subsequently, the entries that indistinguishable response patterns by the original Q-matrix (e.g., 0000, 1000, 0100, 1100, etc.) were randomly changed by 10% and 20% to obtain misspecified Q-matrices. Both the Q-matrix validation methods' MRRs were at the base rate level in the analyses performed with misspecified Q-matrices.

4. Conclusion and Discussion

A Q-matrix reflecting the attributes and item design in cognitive diagnosis is the essential element determining the quality of the diagnostic feedback for the measuring tool. Therefore, Q-matrix has a crucial role in test development (de la Torre et al., 2010; Rupp & Templin, 2008). A Q-matrix constructed with the expert opinion is generally assumed to be correct. However, most problems with item parameter estimates, classification of respondents, and latent class sizes are related to Q-matrix design (DeCarlo, 2011; de la Torre et al., 2010). Therefore, instead of assuming that the Q-matrix is correct, it should be investigated by empirical scrutiny (de la Torre, 2008; de la Torre & Douglas, 2008). Empirical scrutiny has suggested some statistical methods to identify and refine the misspecifications in the Q matrix. This article evaluated the performance of the SEM δ-method (de la Torre, 2008) and QMR method (Chiu, 2013) under various study conditions. The conditions are the number of attributes, the number of examinees, the test lengths, the percentage of misspecified q-entries, and the guessing and slip parameters.

The simulation design results showed that the percentage of misspecification and the test lengths generally had a small or no effect on the MRRs of both Q-matrix validation methods. Chiu (2013) reported similar results for the percentage of misspecification on the effectiveness of the QMR method. In addition, the author noted that the effects of various factors on MRRs were the same for tests of different lengths (J = 20, 40, and 80), but MRRs were higher than 20 items for tests of 40 and 80 items. Q-matrix validation is expected to be more difficult with a long test, as the number of misspecified Q-matrix entries increases with the number of items (or test length; Chiu, 2013).

In addition, simulation design results showed that overall, the larger sample sizes (or the number of examinees) improved the MRRs values of both methods. However, the remedial effect of sample size is minimal when *g* and *s* equal 0.1. In general, an increase in sample size is more effective on MRRs of the SEM δ-method. Chiu (2013) reported that a small sample size was sufficient (e.g., N = 100) for a high MRR in the QMR method. On the other hand, de la Torre (2008) fixed the sample size to 5,000 for the SEM δ-method. However, de la Torre et al. (2010) stated that a larger sample size provided a less biased estimation, but the improvement was not considerable. The authors reported that a sample size of 1,000 was adequate in affording true parameter estimation for the DINA model. Similarly, according to the MRRs obtained for both methods in the current study, it can be stated that a sample size of 1,000 is optimal.

The MRRs of both methods are affected by the number of attributes and *g* and *s* parameters. Chiu (2013) reported similar results on the number of attributes and *g* and *s* parameters. Notably, the deterioration in the *g* and *s* parameters severely reduced the MRRs of both models. Among the study design conditions, the most crucial factor in the MRRs was the *g* and *s* parameter values. However, the MRRs of the QMR method generally were higher than the base rates for each study design condition. Therefore, it can be stated that the RSS-based Q-matrix validation method outperformed the SEM δ-method. Dai et al. (2018) also reported similar results. This advantage of the QMR method is mainly due to the nature of it being a nonparametric method. The QMR method does not require the assumption of dependence on a possible suspicious parameter structure to determine test performance, nor does it require large sample sizes (Chiu, 2013).

On the other hand, the SEM δ-method that can be considered an item discrimination index (IDI = 1- s_j - g_j) considers both the *g* and *s* parameters. The increase in g and s parameters dramatically causes a decrease in MRRs. de la Torre (2008) stated that since the *g* and *s* parameters are utilized in calculating the posterior distribution, the misspecifications of the q-vector may reduce the approximation quality to this distribution.

The real data analysis results showed that both the Q-matrix validation methods' MRRs were at the base rate level. The Q-matrix used in the real data analysis is incomplete since not every attribute is represented by at least one single-attribute test item. The ideal item response patterns of the real data Q-matrix for the four attributes showed that only 7 of the $2^4 = 16$ potential attribute patterns could be recognized by items, but nine could not. Chiu (2013) stated that in real data analysis (K = 8, J = 20; Tatsuoka, 1984), the QMR method recovered 28.75% of misspecified q-entries. The author stated that this poor performance of the QMR method might be due to the fact that a relatively large number of attributes are measured with short tests.

On the other hand, de la Torre (2008) analysed fraction subtraction data (K = 5, J = 15; Tatsuoka, 1990) and 2003 NAEP 8th grade mathematics data (K = 9, J = 90) via the SEM δ-method and reported a reasonable model-data fit and misfit, respectively. In addition, de la Torre (2008) stated that even if the correct q-vector is used in the real data application for the SEM δ-method, there may not be a clear separation between the $\eta_j = 0$ and $\eta_j = 1$ groups because the correct q-vector and the estimated posterior distribution are used. The author stated that this limitation can be overcome if the q-vector contains more attributes than required. However, the practical implications for proposing to add more attributes to the Q-matrix than necessary are not yet clear in real test applications. Therefore, more research results are required under various conditions compatible with real data to have more in-depth information about Q-matrix validation methods’ performance.

As a result, under simulation design conditions, the MRRs of the QMR method are generally higher than the SEM δ-method. Therefore, it can be said that the QMR method performs better than the SEM δ-method. Based on the study results, it can be recommended that CDM practitioners prioritize the QMR method in q-matrix validation. In addition, it may be suggested to use a sample size of 1,000 to validate Q-matrix with the DINA model. Since *g* and *s* parameters are effective in the performance of the MRRs of both methods, it is proposed to consider *g* and *s* values in Q-matrix validation. However, a large number of real data applications are needed to deepen our knowledge of the real data performance of the Q-matrix validation methods. Therefore, it can be suggested to compare the performances of the Q-matrix validation methods, specifically in real data sets where the number of attributes is relatively large, and the test length is short. The current research results are limited to the DINA model and dichotomous data sets, which show great interest in CDMs. It may be worthwhile to consider other CDM models (e.g., DINO model [Templin & Henson, 2006], reparametrized DINA [RDINA; DeCarlo, 2011], etc.), Q-matrix validation methods (e.g., Bayesian approach, etc.), and data sets (e.g., graded response data) in future research.

5. Acknowledgments

The authors declare that a part of this study was presented as an oral abstract presentation at the 2nd International Congresses on New Horizons in Education and Social Sciences (ICES, 2019) held on 18-19 June 2019 in Istanbul, Turkey.

The authors thank Tahsin Oğuz BAŞOKÇU for generously sharing the CDM data.

6. References

Başokçu, T. O., Bardakçı, V., Öğretmen, T., Çakıroğlu, E., Yurdakul, B., & Akyüz, G. (2018). *Uluslararası geniş ölçekli sınavlarda Türkiye'nin matematik başarısını arttırabilmek için bir model önerisi: Bilişsel tanıya dayalı izleme modelinin etkililiği* [A model proposal to increase Turkey's large-scale success in international math exams: the effectiveness of the monitoring model based on cognitive diagnosis] (Research Report of TÜBİTAK, Report No. 115K531). https://app.trdizin.gov.tr/publication/project/detail/TVRnMU56SXk

Chen, C. (2017). *Q-matrix optimization for cognitive diagnostic assessment* [Doctoral dissertation]. University of Illinois at Urbana-Champaign, Illinois Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/98150/CHEN-DISSERTATION-2017.pdf?sequence=1

Chiu, C. Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement*, *37*(8), 598-618. https://doi.org/10.1177/0146621613488436

Chiu, C. Y., & Douglas, J. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns. *Journal of Classification*, *30*(2), 225-250. https://doi.org/10.1007/s00357-013-9132-9

Dai, S., Svetina, D., & Chen, C. (2018). Investigation of missing responses in Q-matrix validation. *Applied psychological measurement*, *42*(8), 660-676. https://doi.org/10.1177/0146621618762742

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, *35*(1), 8-26. https://doi.org/10.1177/0146621610377081

DeCarlo, L. T. (2012). Recognizing uncertainty in the Q-matrix via a Bayesian extension of the DINA model. *Applied Psychological Measurement*, *36*(6), 447-468. https://doi.org/10.1177/0146621612449069

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of educational measurement*, *45*(4), 343-362. https://doi.org/10.1111/j.1745-3984.2008.00069.x

de la Torre, J. (2009a). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, *34*(1), 115-130. https://doi.org/10.3102/1076998607309474

de la Torre, J. (2009b). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, *33*(3), 163-183. https://doi.org/10.1177/0146621608320523

de la Torre, J., & Chiu, C. Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika*, *81*(2), 253-273. https://doi.org/10.1007/s11336-015-9467-8

de la Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, *73*(4), 595-624. https://doi.org/10.1007/s11336-008-9063-2

de la Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement*, *47*(2), 227-249. https://doi.org/10.1111/j.1745-3984.2010.00110.x

de la Torre, J., & Lee, Y. S. (2010). A note on the invariance of the DINA model parameters. *Journal of Educational Measurement*, *47*(1), 115-127. https://doi.org/10.1111/j.1745-3984.2009.00102.x

de la Torre, J., & Lee, Y. S. (2013). Evaluating the Wald test for item-level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, *50*(4), 355-373. https://doi.org/10.1111/jedm.12022

George, A. C., & Robitzsch, A. (2014). Multiple group cognitive diagnosis models, with an emphasis on differential item functioning. *Psychological Test and Assessment Modeling*, *56*(4), 405.

George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software*, *74*(2), 1-24. https://doi.org/10.18637/jss.v074.i02

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, *26*(4), 301 321. https://doi.org/10.1111/j.1745-3984.1989.tb00336.x

Huang, H. Y., & Wang, W. C. (2014). The random-effect DINA model. *Journal of Educational Measurement*, *51*(1), 75-97. https://doi.org/10.1111/jedm.12035

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258-272. https://doi.org/10.1177/01466210122032064

Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement*, 41, 205–236. https://doi.org/10.1111/j.1745-3984.2004.tb01163.x

Ma, W., & de la Torre, J. (2020). An empirical Q-matrix validation method for the sequential generalized DINA model. *British Journal of Mathematical and Statistical Psychology*, *73*(1), 142-163. https://doi.org/10.1111/bmsp.12156

R Core Team. (2020). *R: A language and environment for statistical computing* [Computer Software]. Vienna, Austria: R Foundation for Statistical Computing.

Rupp, A. A., & Templin, J. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, 68, 78-96. https://doi.org/10.1177/0013164407301545

Tatsuoka, K. K. (1983). Rule-space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345–354. https://doi.org/10.1111/j.1745-3984.1983.tb00212.x

Tatsuoka, K. K. (1984). *Analysis of errors in fraction addition and subtraction problems* (Report No. NIE- G-81–0002). Urbana: Computer-based Education Research Laboratory, University of Illinois.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453-488). Erlbaum.

Templin, J., & Henson, R. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287–305. https://doi.org/10.1037/1082-989X.11.3.287

Terzi, R., & de la Torre, J. (2018). An iterative method for empirically-based Q-matrix validation. *International Journal of Assessment Tools in Education*, *5*(2), 248-262.

Zheng, Y., & Chiu, C. Y. (2019). NPCD: Nonparametric methods for cognitive diagnosis. R package version 1.0-11. https://CRAN.R-project.org/package=NPCD