PSYCHOMETRIKA—VOL. 81, NO. 2, 253–273
JUNE 2016
DOI: 10.1007/s11336-015-9467-8

A GENERAL METHOD OF EMPIRICAL Q-MATRIX VALIDATION

JIMMY DE LA TORRE AND CHIA-YI CHIU

RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY

In contrast to unidimensional item response models that postulate a single underlying proficiency, cognitive diagnosis models (CDMs) posit multiple, discrete skills or attributes, thus allowing CDMs to provide a finer-grained assessment of examinees' test performance. A common component of CDMs for specifying the attributes required for each item is the Q-matrix. Although construction of Q-matrix is typically performed by domain experts, it nonetheless, to a large extent, remains a subjective process, and misspecifications in the Q-matrix, if left unchecked, can have important practical implications. To address this concern, this paper proposes a discrimination index that can be used with a wide class of CDM subsumed by the generalized deterministic input, noisy "and" gate model to empirically validate the Q-matrix specifications by identifying and replacing misspecified entries in the Q-matrix. The rationale for using the index as the basis for a proposed validation method is provided in the form of mathematical proofs to several relevant lemmas and a theorem. The feasibility of the proposed method was examined using simulated data generated under various conditions. The proposed method is illustrated using fraction subtraction data.

Key words: cognitive diagnosis, G-DINA, Q-matrix, validation, MMLE.

1. Introduction

Traditional models for measuring proficiency, be they in the framework of classical test theory or item response theory, conceptualize examinee's proficiency as a unidimensional latent construct. Assessment of this proficiency typically involves a set of items varying in psychometric properties (e.g., difficulty, discrimination), but homogenous with respect to the targeted proficiency domain. Test performance is interpreted as the result of a complex interplay between a test taker's position on the latent proficiency continuum and the characteristics of the items administered. These models provide information that are most useful in determining the position of one examinee relative to other examinees and the test items.

In contrast, cognitive diagnosis models (CDMs) offer an alternative psychometric framework for comprehensively assessing examinees' proficiency. Instead of postulating a single proficiency continuum, CDMs construe proficiency as a set of interrelated but separable knowledge within a domain, allowing for a finer-grained assessment of examinees' test performance. As such, CDMs are restricted latent class models, where the latent constructs are referred to as *attributes*, and as a generic term, an attribute can correspond to a highly concrete item-solving skill, or refer to a more abstract psychological construct. Alternatively, one can also view CDMs as analogous confirmatory factor models albeit involving discrete, rather than continuous, latent variables. Performance on a set of test items is explained as a function of the match, or lack thereof, between a test taker's attribute profile and the attribute requirements for correctly answering the test items. A defining feature that sets CDMs apart from other psychometric approaches lies in its goal of providing finer-grained inferences regarding examinees' mastery or nonmastery of the different attributes in a particular domain. It should be noted that although educational applications have dominated many of the CDM developments, these models are general diagnostic tools that can be

Correspondence should be made to Jimmy de la Torre, Department of Educational Psychology, Rutgers, The State University of New Jersey, 10 Seminary Place, New Brunswick, NJ 08901, USA. Email: j.delatorre@rutgers.edu

© 2015 The Psychometric Society

applied outside education. For example, CDMs have been used as a tool for providing diagnosis of psychological disorders (Jaeger, Tatsuoka, & Berns, 2003; Templin & Henson, 2006b). As such, *diagnostic classification models* have also been recently used to refer to CDMs (Rupp, Templin, & Henson, 2010).

Several CDMs that range in generality have been proposed in the literature (DiBello, Roussos, & Stout, 2007; Fu & Li, 2007; Haberman & von Davier, 2007; Rupp & Templin, 2008b). Examples of constrained models with more limited generality are the *deterministic input, noisy and gate* (DINA; Haertel, 1984; Junker & Sijtsma, 2001) model, *noisy input, deterministic output and gate* (NIDA; Maris, 1999; Junker & Sijtsma, 2001) model, *deterministic input, noisy or gate* (DINO; Templin & Henson, 2006b) model, *compensatory and reduced reparameterized unified model* (C-RUM and R-RUM; Hartz & Roussos, 2008), the additive CDM (*A*-CDM; de la Torre, 2011), and the linear logistic model (LLM; de la Torre & Douglas, 2004). The different formulations of these specific models reflect the assumptions they invoke about the processes involved in problem solving in a particular domain. Although highly interpretable, a direct comparison of the structures of these models in their current formulation is highly problematic. To address this issue, more general CDMs that employ a more flexible parameterization have also been proposed. Examples of these models are the general diagnostic model (GDM; von Davier, 2005), the log-linear CDM (LCDM; Henson, Templin, & Willse, 2009), and the generalized DINA (G-DINA; de la Torre, 2011) model. The general CDMs each represents a class of models, and some of the aforementioned specific models can be converted as special cases of the general models.

Regardless of the generality of their formulation, all the above models share a general perspective of relating each item to single or multiple latent attributes, as specified in the Q-matrix (Tatsuoka, 1983). In a test that targets a total of K latent attributes, each of the J items requires a distinct subset of relevant attributes to be answered correctly. These specific item-attribute associations are collected into a binary $J \times K$ matrix, $\mathbf{Q} = \{q_{jk}\}$, where $k = 1, \ldots, K$ and $j = 1, \ldots, J$. The q_{jk} entries indicate whether or not the jth item requires the kth attribute. Possession of a particular attribute is construed as a binary event. Hence, for K attributes, a set of 2^K distinct attribute profiles can be constructed constituting the universe of latent attribute classes. In most, if not all CDM applications, the process of establishing the Q-matrix for a given test tends to be subjective in nature and has raised serious validity concerns among researchers (see de la Torre, 2008; Rupp & Templin, 2008a). Misspecifications in the Q-matrix can severely affect estimation of model parameters and, ultimately, correct classification of examinees.

Although the consequence of misspecified Q-matrices is recognized, at present, only few well-developed methods are available to detect misspecifications in the Q-matrix. Barnes (2010) and Liu, Xu, and Ying (2012) have proposed methods of deriving the Q-matrix without a provisional Q-matrix. That is, the Q-matrix is derived solely based on students' responses without any expert input. However, among the limitations, these methods have been used only with specific CDMs such as the DINA model, DINO model, and a restricted version of the R-RUM. It remains to be seen whether these methods can work when the underlying process is more complex. A model-based approach that can be used when the possible misspecified entries in a provisional Q-matrix can be identified in advance was proposed by DeCarlo (2012), which is the same approach proposed by Templin and Henson (2006a). In this approach, the potentially misspecified q-entries are treated as random variables and estimated with the rest of the model parameters. However, the robustness and generalizability of the proposed method need to be further investigated, particularly in situations where all the misspecified q-entries cannot be identified and included in the analysis. Chiu (2013) develops a Q-matrix refinement method based on the nonparametric classification method proposed by Chiu and Douglas (2013), and comparisons of the residual sum of squares computed from the observed and the ideal item responses. Results indicate that the method can identify and correct misspecified entries in the Q-matrix effectively and efficiently. However,

at present, it has only been shown to work for a limited number of CDMs, and the underlying process, either conjunctive or disjunctive, needs to be specified a priori.

Another related method for provisional Q-matrix validation was proposed by de la Torre (2008). In his paper, an item discrimination index for item j, denoted in this paper as φ_j, was proposed for identifying misspecifications in the Q-matrix when the data are assumed to conform to the DINA model. The rationale for the φ_j rests on comparing the two groups of examinees: group $\eta_j = 1$ composed of individuals who possess all attributes required for item j versus group $\eta_j = 0$ composed of individuals who lack one or more required attributes for the item. Correctly specifying a q-vector should maximize the difference between success probabilities for the two groups. More specifically, let $\boldsymbol{\alpha}_{l'}$, $1 \leq l' \leq 2^K$, denote the attribute pattern of an arbitrary examinee and $\boldsymbol{\alpha}_l$ a candidate q-vector from the set of 2^K possible patterns; then, for item j, $1 \leq j \leq J$, φ_{jl} is defined as

$$\varphi_{jl} = P(X_j = 1|\eta_{ll'} = 1) - P(X_j = 1|\eta_{ll'} = 0),$$

where $\eta_{ll'} = \prod_{k=1}^{K} \alpha_{l'k}^{\alpha_{lk}}$. The correct q-vector for the jth item, $\boldsymbol{q}_j$, is found by maximizing φ_{jl}, that is,

$$\boldsymbol{q}_j = \underset{\boldsymbol{\alpha}_l}{\operatorname{argmax}}(\varphi_{jl}).$$

Identifying $\boldsymbol{q}_j$ can be done through exhaustive or sequential search. Both procedures rely on calculating the difference between the success probabilities of the two groups. If the data follow the DINA model, then a correctly identified q-vector yields a probability of correct response of $1 - s_j$ for examinees who mastered all the required skills (where s_j represents the slip probability for the item j) and g_j for those who lack one or more required skills (g_j is the probability of guessing the correct answer for item j). A misspecified q-vector, however, results in probability of success that is less than $1 - s_j$ for the former group, or a probability of success that is greater than g_j for the latter group.

The performance of the φ_j has been studied with simulated as well as real datasets with encouraging results (de la Torre, 2008). Results from the simulation studies indicated that the proposed method is viable in that the method was able to identify and correctly replace inappropriate q-vectors, while at the same time retain those which were correctly specified. The method recognized and retained appropriately specified q-vector when applied to fraction subtraction data and provided information that was useful in establishing or repudiating q-vector specifications, suggesting alternative specifications, and confirming noninformative items when applied to the 2003 NAEP 8th grade mathematics assessment data. Despite the promising results, it remains an open question whether or not this technique can be applied beyond the DINA model, particularly when the CDMs are of more general formulations.

Recall that the chosen CDM parameterization reflects the researchers' specific conceptualization of the hypothesized processes underlying test performance. In the DINA model, there are only two expected probabilities of success per item across all possible 2^K latent classes; this is the specific property that φ_j exploits. Unfortunately, most CDMs produce more than two expected probabilities of success for the different latent classes. Thus, due to its definition and formulation, more general models severely limit the applicability and generality of φ_j.

The goal of this paper is to extend the idea of empirically validating attribute specifications in a provisional Q-matrix to make it applicable to a wider class of CDMs. Specifically, we propose a discrimination index ς_j^2 that can be used in conjunction with the G-DINA model. This index would not require making an assumption about which specific CDMs are involved. Like φ_j, in addition

to being a measure of the discrimination power of item j, ς_j^2 can also be used for identifying incorrectly specified q-vectors, and more importantly, for suggesting the most appropriate q-vector for the item under inspection. This paper provides a rigorous proof on how ς_j^2 can be used as a method for Q-matrix validation purposes when the underlying cognitive processes are more complex. The viability of the method is examined using simulated data that involved a wide range of CDMs, Q-matrix with multiple misspecifications, varying item qualities, and different attribute structures. It is also illustrated using fraction subtraction data.

2. Overview of the G-DINA Model

By relaxing the strong conjunctive condition, the G-DINA model represents a generalization of the DINA model (de la Torre, 2011). Like most CDMs, the G-DINA model relies on the Q-matrix to specify the associations between the J items and the K attributes. We used $K_j^* = \sum_{k=1}^{K} q_{jk}$ to denote the number of required attributes for item j, and $\boldsymbol{\alpha}_{lj}^*$ to denote the lth reduced attribute pattern among the $L = 2^{K_j^*}$ possible reduced attribute patterns. The reduced attribute patterns with respect to item j contain only the attributes required for the item. For example, if item j in a test that measures $K = 5$ attributes requires attributes 1, 4, and 5 (i.e., $K_j^* = 3$), the original attribute pattern $\boldsymbol{\alpha}_l = (\alpha_{l1}, \alpha_{l2}, \alpha_{l3}, \alpha_{l4}, \alpha_{l5})'$ reduces to $\boldsymbol{\alpha}_{lj}^* = (\alpha_{l1}, \alpha_{l4}, \alpha_{l5})'$. For notational convenience, we can assume that the first K_j^* attributes are required for item j.

The item response function of the G-DINA model can be expressed by relating the probability of a correct response and the model specifications through one of the several possible link functions such as the identity, logit, or log link. The term *link function* is to be understood as having the same meaning as the link functions in the generalized linear models (McCullough & Nelder, 1999). The canonical form of the G-DINA model is formulated using the identity link, and its response function can be written as

$$P_j(\boldsymbol{\alpha}_{lj}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk} + \sum_{k'>k}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\alpha_{lk}\alpha_{lk'} + \cdots + \delta_{j(12\cdots K_j^*)} \prod_{k=1}^{K_j^*} \alpha_{lk}, \tag{1}$$

where δ_{j0} denotes the intercept of item j, δ_{jk} the main effect due to α_{lk}, $\delta_{jkk'}$ the interaction effect of α_{lk} and $\alpha_{lk'}$, and $\delta_{j(1,2,\ldots,K_j^*)}$ the interaction effects of $\alpha_{l1}, \ldots, \alpha_{lK_j^*}$. As noted by De la Torre (2011), the saturated form of the G-DINA model in the identity link is equivalent to the saturated form of the LCDM when the logit link is used, which are both equal to the saturated form under the log link.

To understand how probabilities of success for an item are computed using (1), take the case where $K_j^* = 3$. Two individuals with the reduced attribute patterns $(101)'$ and $(111)'$ will have the following probabilities of success:

$$P_j(101) = \delta_{j0} + \delta_{j1} + \delta_{j3} + \delta_{j13} \tag{2}$$

and

$$P_j(111) = \delta_{j0} + \delta_{j1} + \delta_{j2} + \delta_{j3} + \delta_{j12} + \delta_{j13} + \delta_{j23} + \delta_{j123}, \tag{3}$$

respectively.

De la Torre (2011) showed that with careful parameterization, the saturated models can be used for specifying a host of reduced models. This includes the DINA model, DINO model, and the additive models under the identity (i.e., A-CDM), logit (i.e., C-RUM and LLM), and log (i.e.,

R-RUM) links. To obtain the DINA model, DINO model, R-RUM, and C-RUM by constraining and reparameterizing the saturated model under the logit link, see (Henson et al., 2009). We would like to note that unlike the saturated models, the additive models under the different link functions are not necessarily identical to each other. For example, the additive model under the log link when $K_j^* = 2$, as in,

$$\log P(\boldsymbol{\alpha}_{lj}) = \nu_{j0} + \nu_{j1}\alpha_{l1} + \nu_{j2}\alpha_{l2},$$

will become

$$P(\boldsymbol{\alpha}_{lj}) = \delta_{j0} + \delta_{j1}\alpha_{l1} + \delta_{j2}\alpha_{l2} + \delta_{j12}\alpha_{l1}\alpha_{l2},$$

under the identity link, where $\delta_{j0} = \exp(\nu_{j0})$, $\delta_{j1} = \exp(\nu_{j0})[\exp(\nu_{j1}) - 1]$, $\delta_{j2} = \exp(\nu_{j0})[\exp(\nu_{j2}) - 1]$, and $\delta_{j12} = \exp(\nu_{j0})[\exp(\nu_{j1}) - 1][\exp(\nu_{j2}) - 1]$, which is no longer an additive model.

Going back to the reduced attribute patterns $(101)'$ and $(111)'$, (1) can be simplified when constrained CDMs, such as the DINA model, DINO model, or *A*-CDM, are used instead of the saturated form of the G-DINA model. For the the constrained CDMs, the probabilities of success for $(101)'$ are

$$P_j(101)_{\text{DINA}} = \delta_{j0} \tag{4}$$
$$P_j(101)_{\text{DINO}} = \delta_{j0} + \delta_j^*, \tag{5}$$

and

$$P_j(101)_{A\text{-CDM}} = \delta_{j0} + \delta_{j1} + \delta_{j3}, \tag{6}$$

whereas the probabilities of success for $(111)'$ are

$$P_j(111)_{\text{DINA}} = \delta_{j0} + \delta_{j123} \tag{7}$$
$$P_j(111)_{\text{DINO}} = \delta_{j0} + \delta_j^*, \tag{8}$$

and

$$P_j(111)_{A\text{-CDM}} = \delta_{j0} + \delta_{j1} + \delta_{j2} + \delta_{j3}, \tag{9}$$

where $\delta_j^* = \delta_{j1} = \delta_{j2} = \delta_{j3} = -\delta_{j12} = -\delta_{j13} = -\delta_{j23} = \delta_{j123}$. Compared to the probabilities based on the saturated form of the G-DINA model given in (2) and (3), the probabilities of success associated with constrained CDMs given in (4) to (9) are more concise.

3. Discrimination Index

In this section, we present the theoretical proof to show that under the G-DINA model framework, the proposed discrimination index ς_j^2 (to be defined later) can precisely identify and correct the q-vector for item j if it has been misspecified. For notational convenience, the item subscript j will be dispensed with, but it should be noted that ς^2 is an item-specific index. The method is established on the rationale that a correct q-vector will yield homogeneous latent groups in terms of the probability of success, and therefore will result in groups with the highest variability of probabilities of success given a parsimonious subset of attributes. For example, the groups $\eta_j = 0$ and $\eta_j = 1$ for the DINA model are said to be homogeneous if all the attribute

patterns classified as $\eta_j = 0$ have identical probability of success (i.e., g_j). The same can be said of all the attribute patterns classified as $\eta_j = 1$—their probability of success is $1 - s_j$.

In the proposed method, the q-vector with fewer attribute specifications, but a ς^2 that approximates the the maximum ς^2 is chosen. One major theorem is proposed to explicate the property of ς^2 as a basis for q-vector validation, preceded by the definition of ς^2 and two lemmas which will serve as the foundations for proving the theorem.

The lemmas and theorem below are established on two assumptions, namely, (a) the number of required attributes, K, is known, and (b) the Q-matrix is correctly specified. Although domain experts can be relied on to determine the appropriate number of relevant attributes, it would be difficult, if not impossible, for the same experts to correctly specify all the entries of the Q-matrix, particularly when the test is long. Consequently, (b) can be expected to always be violated. However, such a violation does not automatically invalidate the viability of the proposed method. Many statistical procedures, although derived based on strict assumptions, remain useful despite some degree of violations. For example, the F test in one-way analysis of variance is derived based on the normality and homoscedasticity assumptions. Even if the variances are unequal, the test remains robust provided that the sample sizes are equal (Box, 1954). As the simulation studies will demonstrate, the proposed method appears to be robust when the misspecifications in the Q-matrix is controlled at a reasonable rate, which justifies the usefulness of the method in practice.

As before, let K be the number of attributes in the domain, and K^* the number of required attributes for the item. Without loss of generality, let the first K^* attributes be the required attributes. Also let $\boldsymbol{\alpha}_{1:K} = (\alpha_1, \ldots, \alpha_K)'$ and $\boldsymbol{\alpha}_{K':K''} = (\alpha_{K'}, \ldots, \alpha_{K''})'$ be a subset of $\boldsymbol{\alpha}_{1:K}$. A q-vector $\boldsymbol{q}$ that requires $\alpha_{K'}$ to $\alpha_{K''}$ is denoted as $\boldsymbol{q}_{K':K''}$. We first need to clarify the following conditions. Because $\boldsymbol{q}_{1:K^*}$ is the correctly specified q-vector, a q-vector $\boldsymbol{q}_{1:K'}$ is underspecified if $K' \leq K^*$; $\boldsymbol{q}_{1:K''}$ is overspecified if $K'' \geq K^*$; $\boldsymbol{q}_{K':K''}$ is both underspecified and overspecified if $K' \leq K^* \leq K''$. From here on, we assume that $K' \leq K^* \leq K''$. It should be noted that $\boldsymbol{\alpha}_{K':K''}$ is a $(K'' - K' + 1)$-dimensional vector where the entries could be 0 or 1, whereas a q-vector is always K-dimensional and $\boldsymbol{q}_{K':K''}$ denotes the q-vector where the K'th to the K''th entries are 1 and the others are 0.

The proof starts from defining the two key components of ς^2, w and p, followed by a formal definition of ς^2.

Definition 1. Let $w(\boldsymbol{\alpha}_{1:K''})$ be the posterior probability of examinees in class $\boldsymbol{\alpha}_{1:K''}$ for some $K'' \leq K$, and $p(\boldsymbol{\alpha}_{1:K''})$ be the probability of a correct answer for examinees in this class. Because we assumed $K' < K''$, then

$$w(\boldsymbol{\alpha}_{K':K''}) = \sum_{\alpha_1=0}^{1} \cdots \sum_{\alpha_{K'-1}=0}^{1} w(\boldsymbol{\alpha}_{1:K''}),$$

and

$$\begin{aligned} p(\boldsymbol{\alpha}_{K':K''}) &= \frac{\sum_{\alpha_1=0}^{1} \cdots \sum_{\alpha_{K'-1}=0}^{1} w(\boldsymbol{\alpha}_{1:K''}) p(\boldsymbol{\alpha}_{1:K''})}{\sum_{\alpha_1=0}^{1} \cdots \sum_{\alpha_{K'-1}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})} \\ &= \frac{\sum_{\alpha_1=0}^{1} \cdots \sum_{\alpha_{K'-1}=0}^{1} w(\boldsymbol{\alpha}_{1:K''}) p(\boldsymbol{\alpha}_{1:K''})}{w(\boldsymbol{\alpha}_{K':K''})}. \end{aligned}$$

TABLE 1.
The w and p of $\boldsymbol{\alpha}_{1:4}$, $\boldsymbol{\alpha}_{1:3}$, and $\boldsymbol{\alpha}_{2:4}$, respectively.

$\boldsymbol{\alpha}_{1:4}$		(0000)	(1000)	(0100)	(0010)	(1100)	(1010)	(0110)	(1110)
	w	0.053	0.076	0.039	0.057	0.069	0.047	0.068	0.078
	p	0.225	0.225	0.225	0.225	0.225	0.225	0.225	0.725
		(0001)	(1001)	(0101)	(0011)	(1101)	(1011)	(0111)	(1111)
	w	0.037	0.081	0.073	0.055	0.056	0.083	0.069	0.059
	p	0.225	0.225	0.225	0.225	0.225	0.225	0.225	0.725
$\boldsymbol{\alpha}_{1:3}$		(000−)	(100−)	(010−)	(001−)	(110−)	(101−)	(011−)	(111−)
	w	0.090	0.157	0.112	0.112	0.125	0.130	0.137	0.137
	p	0.225	0.225	0.225	0.225	0.225	0.225	0.225	0.725
$\boldsymbol{\alpha}_{2:4}$		(−000)	(−100)	(−010)	(−001)	(−110)	(−101)	(−011)	(−111)
	w	0.129	0.108	0.104	0.118	0.146	0.129	0.138	0.128
	p	0.225	0.225	0.225	0.225	0.492	0.225	0.225	0.456

The G-DINA model discrimination index of an item with the specification $\boldsymbol{q}_{K':K''}$ is defined as

$$\begin{aligned}\varsigma^2 = \varsigma^2_{K':K''} &= \sum_{\alpha_{K'}=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{K':K''}) \Big[p(\boldsymbol{\alpha}_{K':K''}) - \bar{p}(\boldsymbol{\alpha}_{K':K''}) \Big]^2 \\ &= \sum_{\alpha_{K'}=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{K':K''}) p^2(\boldsymbol{\alpha}_{K':K''}) - \bar{p}^2(\boldsymbol{\alpha}_{K':K''}), \end{aligned} \tag{10}$$

where $\bar{p}(\boldsymbol{\alpha}_{K':K''}) = \sum_{\alpha_{K'}=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{K':K''}) p(\boldsymbol{\alpha}_{K':K''})$ is the weighted probability of success across all the $2^{K''-K'+1}$ possible patterns of $p(\boldsymbol{\alpha}_{K':K''})$.

An example, as shown in Table 1, is given here to illustrate how these two important components of ς^2 are computed. In this example, $K = 4$ and the targeted item requires the first three attributes (i.e., $K^* = 3$). Table 1 includes all possible $\boldsymbol{\alpha}_{1:4}$ and their corresponding $w(\boldsymbol{\alpha}_{1:4})$ and $p(\boldsymbol{\alpha}_{1:4})$.

We next demonstrate how to compute the $w(\boldsymbol{\alpha}_{K':K''})$ and $p(\boldsymbol{\alpha}_{K':K''})$ of the collapsed latent classes. Taking $\boldsymbol{\alpha}_{1:3}$ as an example, based on Definition 1, $w(\boldsymbol{\alpha}_{1:3}) = \sum_{\alpha_4=0}^{1} w(\boldsymbol{\alpha}_{1:4})$ and $p(\boldsymbol{\alpha}_{1:3}) = \sum_{\alpha_4=0}^{1} w(\boldsymbol{\alpha}_{1:4}) p(\boldsymbol{\alpha}_{1:4}) / \sum_{\alpha_4=0}^{1} w(\boldsymbol{\alpha}_{1:4})$. Therefore, for instance, $w[(000-)] = w[(0000)] + w[(0001)] = 0.090$ and

$$\begin{aligned} p[(000-)] &= \frac{w[(0000)]p[(0000)] + w[(0001)]p[(0001)]}{w[(0000)] + w[(0001)]} \\ &= \frac{0.053 \times 0.225 + 0.037 \times 0.225}{0.225 + 0.225} \\ &= 0.225. \end{aligned}$$

The remaining weights and probabilities of correct response associated with $\boldsymbol{\alpha}_{1:3}$ are also given in Table 1. From these weights and probabilities, we can compute $\varsigma^2_{1:3}$ as

$$\varsigma^2_{1:3} = \sum_{\alpha_1=0}^{1}\sum_{\alpha_2=0}^{1}\sum_{\alpha_3=0}^{1} w(\boldsymbol{\alpha}_{1:3})p^2(\boldsymbol{\alpha}_{1:3}) - \bar{p}^2(\boldsymbol{\alpha}_{1:3})$$
$$= 0.116 - 0.294^2 = 0.030.$$

Two lemmas and one theorem are to be provided next to support the use of the ς^2 for validating a Q-matrix. For conciseness, the associated proofs are presented in Appendix A. The first lemma claims that the average probabilities of the subsets of $\boldsymbol{\alpha}_{1:K}$ (e.g., $\boldsymbol{\alpha}_{1:K'}$, $\boldsymbol{\alpha}_{1:K''}$, $\boldsymbol{\alpha}_{K':K''}$, and $\boldsymbol{\alpha}_{1:K^*}$) are identical.

Lemma 1. $\bar{p}(\boldsymbol{\alpha}_{k:K''}) = \bar{p}(\boldsymbol{\alpha}_{1:K''})$ *for all* $k < K''$.

This useful relation can easily be verified using the above example and the information in Table 1. Lemma 1 states that no matter how the latent classes are collapsed, the mean of all the corresponding probabilities of success remains the same. With this lemma, the second term on the RHS of (10) (refer to the proof for Lemma 1 in Appendix A) is a constant across all the possible ς^2, and noting this will greatly simplify the other proofs. Next, two types of q-vectors that both result in homogeneous within-group probabilities of success need to be defined and clarified before the proof of Lemma 2 is provided.

Definition 2. A q-vector is said to be *appropriate* for an item if it results in latent groups with homogeneous within-group probabilities of success. The *correct* q-vector for the item is the appropriate q-vector which contains the minimum number of attribute specifications. □

Thus, the correct q-vector results in the minimum number of latent groups necessary to produce homogeneous within-group probabilities of success. In other words, the probabilities of success are conditionally independent given the latent groups derived from the correct q-vectors. The next lemma claims that given two q-vectors that are both under- and over-specified, one being a subset of the other, the vector with fewer underspecifications will have ς^2 that is at least as large as that of the other vector.

Lemma 2. *Suppose* $K' + 1 \leq K^*$.

$$\sum_{\alpha_{K'+1}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{(K'+1):K''})p^2(\boldsymbol{\alpha}_{(K'+1):K''}) \leq \sum_{\alpha_{K'}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{K':K''})p^2(\boldsymbol{\alpha}_{K':K''}). \tag{11}$$

In the previous example, K' could be 1 or 2. To illustrate Lemma 2, K' is assumed to be 1 and $K'' = 4$. We need to show that

$$\sum_{\alpha_2=0}^{1}\sum_{\alpha_3=0}^{1}\sum_{\alpha_4=0}^{1} w(\boldsymbol{\alpha}_{2:4})p^2(\boldsymbol{\alpha}_{2:4}) \leq \sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_4=0}^{1} w(\boldsymbol{\alpha}_{1:4})p^2(\boldsymbol{\alpha}_{1:4}).$$

Using the information in Table 1, it can be shown that $\sum_{\alpha_2=0}^{1}\sum_{\alpha_3=0}^{1}\sum_{\alpha_4=0}^{1} w(\boldsymbol{\alpha}_{2:4})p^2(\boldsymbol{\alpha}_{2:4}) = 0.099$ and $\sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_4=0}^{1} w(\boldsymbol{\alpha}_{1:4})p^2(\boldsymbol{\alpha}_{1:4}) = 0.116$, which shows that the inequality in Lemma 2 holds.

The theorem states that the index ς^2 is maximized when the correct q-vector is identified. This can be shown by thoroughly comparing the correct q-vector with the remaining q-vectors, which are either underspecified, overspecified, or both.

Theorem 1. $\varsigma^2_{K':K''} \leq \varsigma^2_{1:K^*}$.

Continuing the example above, Theorem 1 states that $\varsigma^2_{1:4} \leq \varsigma^2_{1:3}$ when $K' = 1$ and $\varsigma^2_{2:4} \leq \varsigma^2_{1:3}$ when $K' = 2$. When $K' = 1$, it has been computed in the earlier example that $\varsigma^2_{1:3} = 0.030$. Now, using the information in Table 1, it can be shown that $\varsigma^2_{1:4}$ is also 0.030, which is equal to $\varsigma^2_{1:3}$. When $K' = 2$, $\varsigma^2_{2:4} = 0.013$, which is less than $\varsigma^2_{1:3}$.

The theorem also holds for $K'' = K^*$, when the provisional vector is *strictly* underspecified (i.e., $\boldsymbol{q} = \boldsymbol{q}_{K':K^*}$). In summary, the theorem suggests that when the correct q-vector is not known, we choose the provisional q-vector $\boldsymbol{q}^*$ over $\boldsymbol{q}^{**}$ when (1) $\hat{\varsigma}^{*2} > \hat{\varsigma}^{**2}$, or (2) $K^* < K^{**}$ and $\hat{\varsigma}^{*2} = \hat{\varsigma}^{**2}$.

4. Search Algorithm

It is important to note that the above lemmas and theorem are valid when the provisional Q-matrix used for calibration is indeed the correct Q-matrix, and the probabilities and weights are estimated without error. However, in practice, in addition to the noise introduced by real data, misspecifications in the Q-matrix can degrade the quality of estimation, which in turn can affect the accuracy of estimated probabilities and weights. This will result in $\hat{\varsigma}^{**2}$ being greater, if slightly, than $\hat{\varsigma}^{*2}$. Consequently, the algorithm will not stop at K^*, but continue until all K attributes are specified. Therefore, some stopping rule other than simply selecting q-vector with highest $\hat{\varsigma}^2$ has to be implemented as part of the algorithm. In this study, the criterion, which is described in the next section, will be based on the proportion of variance accounted for (PVAF) by a particular q-vector relative to the maximum $\hat{\varsigma}^2$, which is achieved when all the attributes are specified.

The search algorithm to identify the suggested q-vector begins by finding $\hat{\varsigma}^2$ for each of the possible $2^K - 1$ q-vectors. As expected, $\hat{\varsigma}^2_{1:K}$ will be the maximum, and each $\hat{\varsigma}^2$ is measured against this maximum. Specifically, q-vectors with $\hat{\varsigma}^2/\hat{\varsigma}^2_{1:K} \geq \epsilon$, where ϵ is a predetermined PVAF, are deemed appropriate. From this subset, the q-vector with the least number of required attributes is chosen as the suggested specification for the item. When more than one q-vector can be suggested, the decision is based on the size of $\hat{\varsigma}^2$. It should be noted that the use of ϵ to determine the suggested q-vector is similar to the use of amount of variance accounted for in exploratory factor or principal component analysis to determine the number of factors or components to be retained (cf., Stevens, 2009).

As with the original procedure (i.e., de la Torre, 2008), the validation procedure based on the G-DINA model discrimination index is implemented as a two-step process. In the first step, the provisional Q-matrix is used to obtain estimates of the item parameters in the identity link and posterior distribution of the attribute patterns. The estimates were based on an empirical Bayesian implementation of the EM algorithm. In the second step, the search algorithm based on ς^2 was implemented, and the correct q-vectors were identified one item at a time. In this paper, both steps of the analysis were carried out using a code written in Ox (Doornik, 2007). Readers interested in the code can contact the first author. The real data analysis provides an example of how the search algorithm is carried out for two particular items.

5. Simulation Studies

Two simulation studies consisting of various conditions were conducted to investigate the viability of ς^2 to identify Q-matrix misspecifications. Study 1 investigated the performance of the validation method for data with underlying processes that conformed to one of the reduced models;

TABLE 2.
Summary of the simulation factors.

Factor	Study 1	Study 2
Misspecified Q-entries	Random	
(N, J, K) - fixed	(2000, 30, 5)	
Attribute structure	Higher-order	
Generating model	DINA, A-CDM, DINO, DINA/A-CDM, DINO/A-CDM	G-DINA
(p_0, p_1)	(0.1, 0.9), (0.2, 0.8)	[Unif(0.1,0.3), Unif(0.7,0.9)]

in contrast, Study 2 examined data with more general underlying processes (i.e., conforming to the G-DINA model). The factors considered in the simulation studies are summarized in Table 2; the designs of the studies are elaborated thereafter. The primary goal of the studies was to examine the impact of Q-matrix misspecifications on the performance of the validation method as measured by correct Q-matrix recovery for different CDMs.

5.1. Study 1: Q-Matrix Validation with the Five Reduced Models

5.1.1. Design The number of attributes, the number of items, and sample size were fixed to $K = 5$, $J = 30$, and $N = 2000$, respectively. Examinees' attribute patterns were generated using the higher-order model (de la Torre & Douglas, 2004). The higher-order model assumes that examinee i's mastery or nonmastery attribute k is related to a general unidimensional latent ability, θ_i, through a logit link. Specifically, the probability of mastering α_{ik} given θ_i is

$$P(\alpha_{ik}|\theta_i) = \frac{\exp(\lambda_{1k}(\theta_i - \lambda_{0k}))}{1 + \exp(\lambda_{1k}(\theta_i - \lambda_{0k}))},$$

where λ_{0k} and λ_{1k} are structural parameters, and $\lambda_{1k} > 0$. The K attributes are assumed to be independent conditional on θ_i. Therefore, the joint probability of an attribute pattern conditional on θ_i i is given by

$$P(\boldsymbol{\alpha}_i|\theta_i) = \prod_{k=1}^{K} P(\alpha_{ik}|\theta_i).$$

In these simulation studies, θ_i was drawn from $N(0, 1)$, $\boldsymbol{\lambda}_0 = (-1, -0.5, 0, 0.5, 1)$, and $\lambda_{1k} = 1.5$ for all values of k.

To examine the generality and viability of the proposed method across different underlying response processes, the data were simulated using five different CDMs: DINA model, DINO model, A-CDM, DINA/A-CDM, and DINO/A-CDM. The DINA/A-CDM is a CDM whose IRF represents the average of the IRFs of the DINA model and the A-CDM; similarly, the DINO/A-CDM is a CDM whose IRF represents the average of the IRFs of the DINO model and the A-CDM.

For all attribute structures, number of required attributes, and CDMs, the probabilities of success for groups mastering none and all of the required attributes (i.e., the zero and one vectors) were set to either $(p_0, p_1) = (0.1, 0.9)$ or $(0.2, 0.8)$. These probabilities represented the amount of error perturbation in the data and were expected to have an impact on the correct Q-matrix recovery rate.

For the DINA model, the probability of success was p_1 for examinees possessing all the required attributes, and p_0 for examinees lacking at least one of the required attributes. For the DINO model, the probability of success was p_0 for examinees lacking all the required attributes, and p_1 for examinees possessing at least one of the required attributes. For the *A*-CDM, the baseline probability was p_0, and the probability of success increased by $(p_1 - p_0)/K_j^*$ for each required attribute mastered. The probabilities of success for the DINA/*A*-CDM were obtained by averaging the DINA and *A*-CDM probabilities of success associated with the same latent group for the five generating models when $K_j^* \geq 2$. The same procedure was applied for the DINO/*A*-CDM. In addition to the probabilities of success of each latent group for the five CDMs, Figure 1 also gives the corresponding parameters for these probabilities. For each of the 10 conditions resulting from combining five generating models, and two item qualities, 100 datasets were generated to minimize the impact of the Monte Carlo error.

The correct Q-matrix used to generate the data is given in Table 3, and was constructed to have equal number of 1-, 2-, and 3-attribute items. Considering that in practice, the misspecifications, if there are any, in a given Q-matrix are usually not known, the misspecifications were assumed to occur at random. An average of 5 % of the q-entries were altered to focus on the impact of the number rather than the nature of the Q-matrix misspecifications. Specifically, for each replication, a different Q-matrix misspecification was used, and in each condition, the first 50 and the last 50 out of the 100 misspecified Q-matrices consisted of 7 and 8 misspecified q-entries, respectively. As such, up to 27 % of the 30 q-vectors can be misspecified at a time. The cutoff value for PVAF is set at $\epsilon = .95$.

A 2×2 contingency table, reporting the true-positive, true-negative, false-positive, and false-negative rates, was created to present the validation results for each generating model across the 100 replications. In the context of the study, the true-negative rate, which is also known as *sensitivity*, indicates the proportion of misspecified q-entries or q-vectors that was identified and changed; the true-positive rate, which is also known as *specificity*, indicates the proportion of correctly specified q-entries or q-vectors that was retained; the false-negative and false-positive rates, which are analogous to Type I and Type II errors, represent the validation errors where correctly specified q-entries or q-vectors were modified, and misspecified q-entries or q-vectors were retained, respectively.

5.1.2. Results Table 4 cross-tabulates the true statuses of the q-entries and q-vectors against their eventual classifications for each of the generating models based on the best true-positive and true-negative rates for data analyzed with 5 % random Q-matrix misspecifications. The results indicate that the validation method could effectively retain correctly specified q-entries, regardless of the generating model. In addition, the method identified and corrected misspecified q-entries and q-vectors with reasonably high rates. The true-positive rates were all above 97 %, and the true negative rates ranged from 78 to 89 % at the attribute level and 75– 85 % at the vector level. Correcting misspecified q-entries was very effective when data were generated from the DINA or DINO model. This is because the DINA and DINO models are relatively less complex compared to the other CDMs under investigation. The correction became more difficult as the complexity of the generating model increased, the true-negative rates remained high for DINA/*A*-CDM, *A*-CDM, and DINO/*A*-CDM. These findings indicate that even with high proportions of misclassified q-entries, the validation method can improve the Q-matrix specification with high accuracy provided at least moderate quality items were used.

It should be noted that, at the q-entry level, the worst condition (i.e., DINA/A-CDM and DINO/A-CDM) had an overall correct recovery rate of 98.6 %, which represented a chance-adjusted improvement of 72.1 % from the baseline rate; at the q-vector level, the worst condition (i.e., DINA/A-CDM) had an overall correct recovery rate of 93.6 %, which represented an chance-adjusted improvement of 71.1 % from the baseline rate. Moreover, there was no substantial

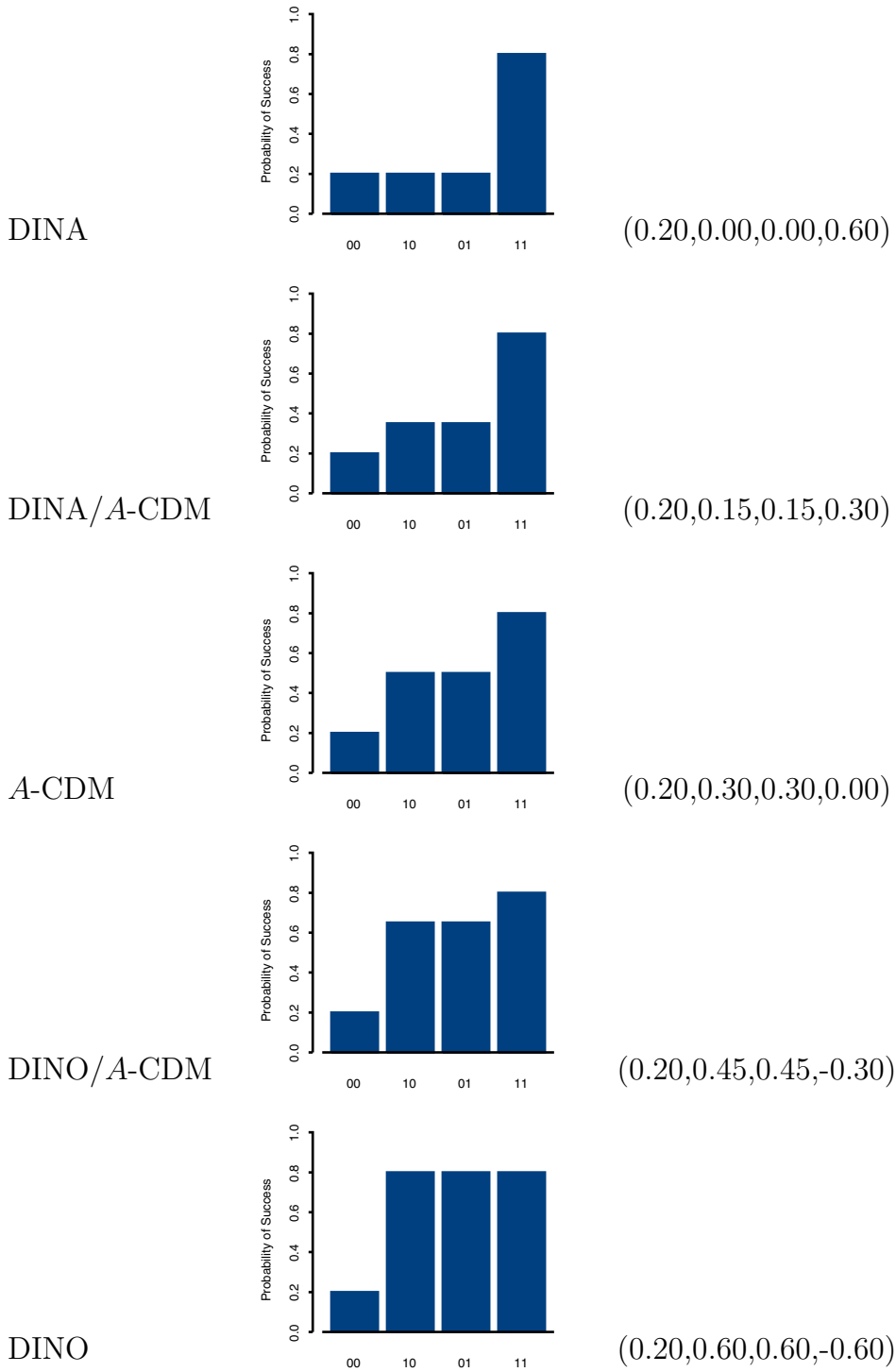

FIGURE 1.
Success probabilities for various instances of the G-DINA model when $K_j^* = 2$.

TABLE 3.
Q-matrix for the simulated data.

Item	α_1	α_2	α_3	α_4	α_5	Item	α_1	α_2	α_3	α_4	α_5
1	1	0	0	0	0	16	0	1	0	1	0
2	0	1	0	0	0	17	0	1	0	0	1
3	0	0	1	0	0	18	0	0	1	1	0
4	0	0	0	1	0	19	0	0	1	0	1
5	0	0	0	0	1	20	0	0	0	1	1
6	1	0	0	0	0	21	1	1	1	0	0
7	0	1	0	0	0	22	1	1	0	1	0
8	0	0	1	0	0	23	1	1	0	0	1
9	0	0	0	1	0	24	1	0	1	1	0
10	0	0	0	0	1	25	1	0	1	0	1
11	1	1	0	0	0	26	1	0	0	1	1
12	1	0	1	0	0	27	0	1	1	1	0
13	1	0	0	1	0	28	0	1	1	0	1
14	1	0	0	0	1	29	0	1	0	1	1
15	0	1	1	0	0	30	0	0	1	1	1

TABLE 4.
Contingency tables for the five reduced models when $p_0 = 0.2$ at both the attribute and vector levels with random misspecified q-vectors in %.

Model	Q-matrix	Attribute level		Vector level	
	Entry	Corrected	Retained	Corrected	Retained
DINA	Misspecified	88.4	11.6	84.3	15.7
	Correct	0.8	99.2	2.1	97.9
DINA/ *A*-CDM	Misspecified	79.9	20.1	76.8	23.2
	Correct	0.4	99.6	1.6	98.4
A-CDM	Misspecified	81.9	18.1	80.1	19.9
	Correct	0.2	99.8	0.8	99.2
DINO/ *A*-CDM	Misspecified	78.4	21.6	75.8	24.2
	Correct	0.3	99.7	0.9	99.1
DINO	Misspecified	88.4	11.6	83.1	16.9
	Correct	0.8	99.2	2.0	98.0

difference between the rates at the attribute and the vector levels. This is because most misspecified vectors only contained one or two misspecified q-entries. Consequently, not much additional information can be drawn by studying the results separately at q-entry and q-vector levels.

5.2. Study 2: Q-Matrix Validation with the G-DINA Model

The first study was designed to examine how the proposed validation method performs across five specific constrained versions of the G-DINA model. In contrast, this following study was designed to provide detailed evaluation of how the method performs when data conform to the unconstrained G-DINA model.

5.2.1. Design As shown in Table 2, the only difference between the design of the current study and that of the previous one was the way the probabilities of success for the latent classes

TABLE 5.
Contingency table for the G-DINA model at both the attribute and vector levels with random misspecified q-vectors in %.

Q-matrix	Attribute level		Vector level	
Entry	Corrected	Retained	Corrected	Retained
Misspecified	80.4	19.6	74.4	25.6
Correct	2.0	98.0	9.7	90.3

given the correct Q-matrix, or equivalently the item parameters, were determined. The probabilities of the latent classes possessing no skill and all skills (i.e., p_0 and p_1) were generated from Unif(0.1, 0.3) and Unif(0.7, 0.9), keeping the average probabilities to 0.2 and 0.8, respectively. The probabilities of the other latent classes are generated from the distribution of Unif(p_0, p_1) with the constraint that $P(\boldsymbol{\alpha}) \geq P(\boldsymbol{\alpha}')$ whenever $\alpha_k \geq \alpha'_k \; \forall k$.

5.2.2. Results Table 5 gives the success and error rates of the validation method at the q-entry and q-vector levels when the data conform to the G-DINA model. The high true-positive rates indicate that the method preserved almost all the correctly specified q-entries and q-vectors during the validation process. The results of the true-negative rates show that the method was able to correct a fairly high proportion of misspecified q-entries and q-vectors. However, the relatively lower numbers indicate that validating the Q-matrix involving the unconstrained G-DINA model was more challenging compared to the reduced models analyzed in Study 1. Nonetheless, the overall chance-adjusted improvements from the baseline rates were 42.8 % at the q-entry level and 39.4 % at the q-vector level.

6. Real Data Example

6.1. Data and Analysis

The data for this example represent a subset of the data originally described and used by Tatsuoka (1990) and by De la Torre (2011). The data were responses from 536 middle school students to 11 fraction subtraction problems with the following four attributes: (1) performing basic fraction subtraction operation, (2) simplifying/reducing, (3) separating whole number from fraction, and (4) borrowing one from whole number to fraction. Table 6 gives the items used in this example, and the corresponding attribute specifications (i.e., Q-matrix).

6.2. Results

The last column of Table 6 gives $\hat{\varsigma}^2_{\max}$ for the fraction subtraction items, which represents the $\hat{\varsigma}^2$ for an item when all the attributes are specified. The proposed algorithm determined for each item the $\boldsymbol{\alpha}_l$ with the minimum number of required attributes such that $\hat{\varsigma}^2_l / \hat{\varsigma}^2_{\max} \geq \epsilon = .95$. Based on the proposed algorithm, the original attribute specifications for eight of the eleven items were retained. For three items (i.e., 4, 5, and 11), different attribute specifications were suggested. Incidentally, the original and suggested attribute specifications for these three items were identical, as in, 1111 and 1011, respectively. It should be noted that the original attribute specifications items 1 and 10, which was also 1111, were retained. In comparing items 1 and 10 with items 4, 5, and 11, it is not obvious what features of the first two items make them different from the last three items. In contrast, in comparing item 9, which has an original attribute specification of 1011, to item 4, 5 and 11, it can be seen that the subtrahend of the the former is not a mixed fraction (i.e., it does not have a whole number part), so it is not clear why the four items should have the

TABLE 6.
Q-matrix for the fraction subtraction data.

Item		Attribute				$\hat{\varsigma}^2_{\max}$
		α_1	α_2	α_3	α_4	
1	$3\frac{1}{2} - 2\frac{3}{2}$	1	1	1	1	0.1463
2	$\frac{6}{7} - \frac{4}{7}$	1	0	0	0	0.0916
3	$3\frac{7}{8} - 2$	1	0	1	0	0.0351
4	$4\frac{4}{12} - 2\frac{7}{12}$	1	1	1	1	0.1383
5	$4\frac{1}{3} - 2\frac{4}{3}$	1	1	1	1	0.1922
6	$\frac{11}{8} - \frac{1}{8}$	1	1	0	0	0.1146
7	$3\frac{4}{5} - 3\frac{2}{5}$	1	0	1	0	0.1392
8	$4\frac{5}{7} - 1\frac{4}{7}$	1	0	1	0	0.1213
9	$7\frac{3}{5} - \frac{4}{5}$	1	0	1	1	0.1715
10	$4\frac{1}{10} - 2\frac{8}{10}$	1	1	1	1	0.1496
11	$4\frac{1}{3} - 1\frac{5}{3}$	1	1	1	1	0.1681

TABLE 7.
Largest $\hat{\varsigma}^2$ and PVAF of items 8 and 11 for different numbers of attribute specifications.

Item	Attribute specification	$\hat{\varsigma}^2$	PVAF
8	1000	0.1067	0.8799
	1010[a,b]	0.1206	0.9947
	1110	0.1211	0.9989
	1111	0.1213	1.0000
11	0001	0.0836	0.4973
	0011	0.1363	0.8112
	1011[a]	0.1665	0.9905
	1111[b]	0.1681	1.0000

[a] Suggested, [b] Original.

same attribute specification. Resolving these inconsistencies between the original and suggested attribute specifications would require the involvement of subject matter, specifically, mathematics education, experts.

Table 7 shows several of the steps in computing $\hat{\varsigma}^2$ of items 8 and 11. These items represent items with identical and different original and suggested attribute specifications, respectively. It should be noted that for a number of required attributes, only the specification with the highest PVAF was presented. As can be seen from the table, for item 8, the minimum number of attributes needed to account for at least 95 % of $\hat{\varsigma}^2_{\max}$ was two, and the largest PVAF of .9957 can be obtained by specifying α_1 and α_3; for item 11, the minimum number of attributes was three, and the largest PVAF of .9901 can be obtained by specifying α_1, α_3, and α_4.

7. Discussion

The impact of Q-matrix misspecification is often overlooked in the context of model fit evaluation because, after construction, Q-matrices are usually assumed to be correct. As a result, model misfit due to an inappropriate Q-matrix cannot be detected or remedied. To deal with this issue, we proposed a discrimination index as a basis for a method of validating Q-matrix specifications in association with a wide class of CDMs. The proposed index, ς^2, can be used without assuming the specific CDMs involved, only that they are subsumed by the G-DINA model. Thus, ς^2 represents a generalization of the discrimination index, φ, proposed by de la Torre (2008) specifically for the DINA model. As such, ς^2 has greater applicability and generality.

The rationale for the proposed validation method was theoretically undergirded by proving a theorem related to the properties of ς^2. The viability of the method was examined in simulation studies that involved several CDMs and multiple Q-matrix misspecifications. Results of the studies showed that the method can accurately identify and correct misspecified q-entries without altering correct entries, particularly when high-quality items are involved. The real data example showed that the proposed method can identify q-vectors that are possibly misspecified.

Despite the promising results, additional research along this line is required before the proposed method can be incorporated as routine part of CDM analysis. First, because the simulation study fixed potentially relevant factors such as test length, the number of attributes, and sample size, the findings of the study have limited generality. Additional studies that involve a greater number of conditions are needed to understand the strengths and limitations of the method in practical applications. For example, it would be interesting to know how the method will perform when K is large or N is small. Second, it should be noted that Q-matrix misspecification is only one of many factors contributing to the model-data misfit. Other factors such as an incomplete set of the attributes or inappropriateness of discrete latent traits can contribute to model misfit. It would be useful to examine to what extent the proposed index can be applied in conjunction with other model fit indices (e.g., fit indices based on expected and observed moments; Chen, de la Torre, & Zhang, 2013; de la Torre & Douglas, 2004) to provide a more holistic approach to diagnosis of model misfit.

Third, even with insights from the simulation studies, an unequivocal rule for determining a single optimal value of ϵ is not at hand. It would be interesting to examine to what extent the different Q-matrices resulting from using different values of ϵ will contain the Q-matrix one would derive from using more formal and, typically, more exhaustive procedures. If the use of ϵ can effectively narrow down the number of potential Q-matrices to be considered, then the determination of the final Q-matrix can be carried out more formally (e.g., via hypothesis testing).

Fourth, it should be noted that the proposed procedure assumes that the number of attributes K has been correctly identified. In some situations, the correct K may not be known, or different numbers of attributes may be under consideration. A procedure, perhaps an extension of the proposed method, that can also identify the correct number of attributes would have immense value in many practical testing situations. However, it should be noted that statistical optimality is only one of the considerations that needs to be taken into account when determining the most appropriate size of K. Practical constraints, such as the length of test that can be reasonably administered within a specific classroom time period, should also be considered in making such decisions.

Finally, it is important to emphasize that the proposed method is not intended to replace the current methods of constructing and validating Q-matrices. Rather it is designed to provide supplemental information for improving model-data fit, and consequently, increasing the validity of inferences from cognitive diagnostic assessments. However, in many applied situations, Q-matrix recommendations based on the proposed method can differ, sometimes markedly, from those based on expert opinions. For this reason, it would be of great value, particularly to practitioners, to examine how such discrepancies can be resolved.

Acknowledgments

This research was supported in part by National Science Foundation Grant DRL-0744486.

Appendix : Proofs of the Lemmas and the Theorem

Lemma 1. $\bar{p}(\boldsymbol{\alpha}_{k:K''}) = \bar{p}(\boldsymbol{\alpha}_{1:K''})$ *for all* $k < K''$.

Proof. According to Definition 1,

$$\begin{aligned}
\bar{p}(\boldsymbol{\alpha}_{k:K''}) &= \sum_{\alpha_k=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{k:K''}) p(\boldsymbol{\alpha}_{k:K''}) \\
&= \sum_{\alpha_k=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{k:K''}) \frac{\sum_{\alpha_1=0}^{1} \cdots \sum_{\alpha_{k-1}=0}^{1} w(\boldsymbol{\alpha}_{1:K''}) p(\boldsymbol{\alpha}_{1:K''})}{w(\boldsymbol{\alpha}_{k:K''})} \\
&= \sum_{\alpha_k=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} \sum_{\alpha_1=0}^{1} \cdots \sum_{\alpha_{k-1}=0}^{1} w(\boldsymbol{\alpha}_{1:K''}) p(\boldsymbol{\alpha}_{1:K''}) \\
&= \sum_{\alpha_1=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''}) p(\boldsymbol{\alpha}_{1:K''}) \\
&= \bar{p}(\boldsymbol{\alpha}_{1:K''}) \qquad (12)
\end{aligned}$$

Because (12) holds for all $k < K''$, $\bar{p}(\boldsymbol{\alpha}_{K':K''}) = \bar{p}(\boldsymbol{\alpha}_{1:K'}) = \bar{p}(\boldsymbol{\alpha}_{1:K^*}) = \bar{p}(\boldsymbol{\alpha}_{1:K''})$. □

Lemma 2. *Suppose* $K' + 1 \leq K^*$.

$$\sum_{\alpha_{K'+1}=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{(K'+1):K''}) p^2(\boldsymbol{\alpha}_{(K'+1):K''}) \leq \sum_{\alpha_{K'}=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{K':K''}) p^2(\boldsymbol{\alpha}_{K':K''}). \qquad (13)$$

Proof. The LHS of (13) equals

$$\begin{aligned}
&\sum_{\alpha_{K'+1}=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} \sum_{\alpha_{K'}=0}^{1} w(\boldsymbol{\alpha}_{K':K''}) \left[\frac{\sum_{\alpha_{K'}=0}^{1} w(\boldsymbol{\alpha}_{K':K''}) p(\boldsymbol{\alpha}_{K':K''})}{\sum_{\alpha_{K'}=0}^{1} w(\boldsymbol{\alpha}_{K':K''})} \right]^2 \\
&= \sum_{\alpha_{K'+1}=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} \frac{\left[\sum_{\alpha_{K'}=0}^{1} w(\boldsymbol{\alpha}_{K':K''}) p(\boldsymbol{\alpha}_{K':K''}) \right]^2}{\sum_{\alpha_{K'}=0}^{1} w(\boldsymbol{\alpha}_{K':K''})} \\
&= \sum_{\alpha_{K'+1}=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} \left[w(0, \boldsymbol{\alpha}_{(K'+1):K''}) + w(1, \boldsymbol{\alpha}_{(K'+1):K''}) \right]^{-1} \\
&\qquad \Big[w^2(0, \boldsymbol{\alpha}_{(K'+1):K''}) p^2(0, \boldsymbol{\alpha}_{(K'+1):K''}) + w^2(1, \boldsymbol{\alpha}_{(K'+1):K''}) p^2(1, \boldsymbol{\alpha}_{(K'+1):K''}) \\
&\qquad\quad 2w(0, \boldsymbol{\alpha}_{(K'+1):K''}) w(1, \boldsymbol{\alpha}_{(K'+1):K''}) p(0, \boldsymbol{\alpha}_{(K'+1):K''}) p(1, \boldsymbol{\alpha}_{(K'+1):K''}) \Big]. \qquad (14)
\end{aligned}$$

The RHS of (13) is

$$\sum_{\alpha_{K'+1}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1}\Big[w(0,\boldsymbol{\alpha}_{(K'+1):K''})p^2(0,\boldsymbol{\alpha}_{(K'+1):K''})+w(1,\boldsymbol{\alpha}_{(K'+1):K''})p^2(1,\boldsymbol{\alpha}_{(K'+1):K''})\Big]. \tag{15}$$

Subtracting (14) from (15), and simplifying,

$$\sum_{\alpha_{K'+1}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1}\left[\frac{w(0,\boldsymbol{\alpha}_{(K'+1):K''})w(1,\boldsymbol{\alpha}_{(K'+1):K''})}{w(0,\boldsymbol{\alpha}_{(K'+1):K''})+w(1,\boldsymbol{\alpha}_{(K'+1):K''})}\right]\Big[p(0,\boldsymbol{\alpha}_{(K'+1):K''})-p(1,\boldsymbol{\alpha}_{(K'+1):K''})\Big]^2$$
$$\geq 0.$$

Therefore, (13) holds. □

Theorem 1. $\varsigma^2_{K':K''} \leq \varsigma^2_{1:K^*}$.

Proof. Case 1: The provisional q-vector is *strictly* overspecified, that is, $K' = K^* < K''$ resulting in $\boldsymbol{q} = \boldsymbol{q}_{1:K''}$ and $\varsigma^2 = \varsigma^2_{1:K''}$.
By Lemma 1, the theorem for this case can be proved by showing that

$$\sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})p^2(\boldsymbol{\alpha}_{1:K''}) = \sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_{K^*}=0}^{1} w(\boldsymbol{\alpha}_{1:K^*})p^2(\boldsymbol{\alpha}_{1:K^*}). \tag{16}$$

Based on Definition 1,

$$w(\boldsymbol{\alpha}_{1:K^*}) = \sum_{\alpha_{K^*+1}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''}),$$

and

$$p(\boldsymbol{\alpha}_{1:K^*}) = \frac{\sum_{\alpha_{K^*+1}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})p(\boldsymbol{\alpha}_{1:K''})}{\sum_{\alpha_{K^*+1}=0}^{1},\cdots,\sum_{\alpha_{1:K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})}.$$

The RHS of (16) can be expressed as

$$\sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_{K^*}=0}^{1} w(\boldsymbol{\alpha}_{1:K^*})\left[\frac{\sum_{\alpha_{K^*+1}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})p(\boldsymbol{\alpha}_{1:K''})}{\sum_{\alpha_{K^*+1}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})}\right]^2. \tag{17}$$

By definition of a correct q-vector,

$$p(\boldsymbol{\alpha}_{1:K''}) = p(\boldsymbol{\alpha}_{1:K^*}) \;\; \forall \;\; \alpha_{(K^*+1)},\ldots,\alpha_{K''}.$$

Thus, (17) is equal to

$$\sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_{K^*}=0}^{1} w(\boldsymbol{\alpha}_{1:K^*})\left[\frac{p(\boldsymbol{\alpha}_{1:K^*})\sum_{\alpha_{K^*+1}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})}{\sum_{\alpha_{K^*+1}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})}\right]^2$$
$$= \sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_{K^*}=0}^{1}\sum_{\alpha_{K^*+1}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})p^2(\boldsymbol{\alpha}_{1:K''})$$
$$= \sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})p^2(\boldsymbol{\alpha}_{1:K''})$$

which is the LHS of (16).

Case 2: When the provisional q-vector is *both* under- and overspecified, that is, $K' < K^* < K''$, $\boldsymbol{q} = \boldsymbol{q}_{K':K''}$ and $\varsigma^2 = \varsigma^2_{K':K''}$.
Case 2 will be proved by induction. By Lemma 1 and the result for Case 1, Case 2 of the theorem can be proved by showing that

$$\begin{aligned}&\sum_{\alpha_{K'}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{K':K''})p^2(\boldsymbol{\alpha}_{K':K''})\\ &\le \sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_{K^*}=0}^{1} w(\boldsymbol{\alpha}_{1:K^*})p^2(\boldsymbol{\alpha}_{1:K^*})\\ &= \sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})p^2(\boldsymbol{\alpha}_{1:K''})\end{aligned} \tag{18}$$

Step 1: Show that (18) is true for $K' = 1$.

The LHS of (18) is

$$\sum_{\alpha_{K'}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{K':K''})p(\boldsymbol{\alpha}_{K':K''}) = \sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})p(\boldsymbol{\alpha}_{1:K''}),$$

which is the RHS of (18).

Step 2: Assume that (18) is true for $K' = k$, that is,

$$\sum_{\alpha_k=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{k:K''})p^2(\boldsymbol{\alpha}_{k:K''}) \le \sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''})p^2(\boldsymbol{\alpha}_{1:K''}).$$

Step 3: Show that (18) is true for $K' = k + 1$, as in,

$$\sum_{\alpha_{k+1}=0}^{1}\cdots\sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{k+1:K''})p^2(\boldsymbol{\alpha}_{k+1:K''})$$
$$\le \sum_{\alpha_1=0}^{1}\cdots\sum_{\alpha_{K^*}=0}^{1} w(\boldsymbol{\alpha}_{1:K^*})p^2(\boldsymbol{\alpha}_{1:K^*})$$

$$= \sum_{\alpha_1=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''}) p^2(\boldsymbol{\alpha}_{1:K''}) \tag{19}$$

According to Lemma 2, the LHS of (19) is

$$\sum_{\alpha_{k+1}=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{k+1:K''}) p^2(\boldsymbol{\alpha}_{k+1:K''}) \leq \sum_{\alpha_k=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{k:K''}) p^2(\boldsymbol{\alpha}_{k:K''}). \tag{20}$$

According to the assumption in Step 2, (20) can further be expressed as

$$\sum_{\alpha_k=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{k:K''}) p^2(\boldsymbol{\alpha}_{k:K''}) \leq \sum_{\alpha_1=0}^{1} \cdots \sum_{\alpha_{K''}=0}^{1} w(\boldsymbol{\alpha}_{1:K''}) p^2(\boldsymbol{\alpha}_{1:K''}),$$

which is the RHS of (19). □

References

Barnes, T. (2010). Novel derivation and application of skill matrices: The q-matrix method. In *Handbook on educational data mining* (pp. 159–172). Boca Raton: CRC Press.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one-way classification. *Annals of Mathematical Statistics*, *25*, 290–302.

Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, *50*, 123–140.

Chiu, C.-Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement*, *37*, 598–618.

Chiu, C.-Y., & Douglas, J. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns. *Journal of Classication*, *30*, 225–250.

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*, 343–362.

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

de la Torre, J., & Douglas, J. (2004). Higher order latent trait models for cognitive diagnosis. *Psychometrika*, *63*, 333–353.

de la Torre, J., van der Ark, L. A., & Rossi, G. (in press). Analysis of clinical data from a cognitive diagnosis modeling framework. Measurement and Evaluation in Counseling and Development.

DeCarlo, L. T. (2012). Recognizing uncertainty in the Q-matrix via a Bayesian extension of the DINA model. *Applied Psychological Measurement*, *36*, 447–468.

DiBello, L. V., Roussos, L. A., & Stout, W. F. (2007). Review of cognitively diagnostic assessment and summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, Vol. 26, psychometrics* (pp. 979–1030). Amsterdam: Elsevier.

Doornik, J. A. (2007). *Object-oriented matrix programming using Ox* (3rd ed.). London: Timberlake Consultants Press.

Fu, J., & Li, Y. (2007). *An integrative review of cognitively diagnostic psychometric models*. Paper presented at the Annual Meeting of the National Council of Measurement in Education, Chicago, IL.

Haberman, S. J., & von Davier, M. (2007). Some notes on models for cognitively based skill diagnosis. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, vol. 26, psychometrics* (pp. 1031–1038). Amsterdam: Elsevier.

Haertel, E. H. (1984). An application of latent class models to assessment data. *Applied Psychological Measurement*, *8*, 333–346.

Hartz, S. M., & Roussos, L. A. (2008). The Fusion Model for skills diagnosis: Blending theory with practice. *Educational Testing Service, Research Report, RR-08-71*. Princeton, NJ: Educational Testing Service.

Henson, R. A., Templin, J. L., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*, 191–210.

Jaeger, J., Tatsuoka, C., & Berns, S. (2003). Innovation methods for extracting valid cognitive deficit profiles from NP test data in schizophrenia. *Schizophrenia Research*, *60*, 140–140.

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

Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement*, *36*, 548–564.

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

McCullagh, P., & Nelder, J. (1999). *Generalized linear models* (2nd ed.). Boca Raton, FL: Chapman and Hall.

Rupp, A. A., & Templin, J. L. (2008a). The effect of Q-matrix misspecification on parameter estimates and misclassification rates in the DINA model. *Educational and Psychological Measurement*, *68*, 78–96.

Rupp, A. A., & Templin, J. L. (2008b). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement*, *6*, 219–262.

Rupp, A. A., Templin, J. L., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York, NY: Guilford.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Mahwah, NJ: Erlbaum.

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

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). Hillsdale, NJ: Erlbaum.

Templin, J. L., & Henson, R. A. (2006a). *A Bayesian method for incorporating uncertainty into Q-matrix estimation in skills assessment*. Paper Presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.

Templin, J. L., & Henson, R. A. (2006b). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*, 287–305.

von Davier, M. (2005). A general diagnostic model applied to language testing data. *Educational Testing Service, Research Report, RR-05-16*.

Manuscript Received: 21 OCT 2013
Final Version Received: 7 APR 2015
Published Online Date: 6 MAY 2015