Statistical Consequences of Attribute Misspecification in the Rule Space Method

Seongah Im and James E. Corter

Educational and Psychological Measurement 2011 71: 712
DOI: 10.1177/0013164410384855

Educational and Psychological Measurement
71(4) 712–731

DOI: 10.1177/0013164410384855
http://epm.sagepub.com

Statistical Consequences of Attribute Misspecification in the Rule Space Method

Seongah Im[1] and James E. Corter[2]

Abstract

The present study investigates the statistical consequences of attribute misspecification in the rule space method for cognitively diagnostic measurement. The two types of attribute misspecifications examined in the present study are exclusion of an essential attribute (which affects problem-solving performance) and inclusion of a superfluous attribute (which does not). Results of a simulation study show that exclusion of an essential attribute tends to lead to underestimation of examinees' mastery probabilities for the remaining attributes, whereas inclusion of a superfluous attribute generally leads to overestimation of attribute mastery probabilities for the other attributes. In addition, order relations among attributes induced by superset/subset relationships affect the biases in the estimated attribute mastery probabilities in systematic ways. These results underscore the importance of correct attribute specification in cognitively diagnostic assessment and delineate some specific effects of using incorrect attribute sets.

Keywords

rule space method, cognitive diagnostic measurement method, attribute misspecification, Q-matrix

In recent years, there has been growing interest in using cognitively diagnostic measurement methods to better understand the cognitive processes required for test performance and enable assessing an examinee's performance as a profile of specific knowledge and skills. This shift in emphasis from outcomes to processes has posed

[1]University of Hawaii, Honolulu, HI, USA
[2]Columbia University, New York, NY, USA

Corresponding Author:
Seongah Im, Department of Educational Psychology, College of Education, Wist Hall 214, University of Hawaii, 1776 University Avenue, Honolulu, HI 96822, USA
Email: seongahi@hawaii.edu

a major challenge for psychometric theory and has begun to influence test construction and analysis procedures (Baker, 1993). Researchers have developed methodology that reflects this paradigm shift (e.g., de la Torre & Douglas, 2004; DiBello, Stout, & Roussos, 1995; Hartz, 2002; Leighton, Gierl, & Hunka, 2004; C. Tatsuoka, 2002; K. Tatsuoka, 1985, 2009; von Davier, 2005). The rule space method (RSM; K. Tatsuoka, 1985, 2009) is one of the original cognitive diagnostic approaches; its applicability and practicality have been demonstrated in various content areas such as architecture (Katz, Martinez, Sheehan, & Tatsuoka, 1998), English (Buck & Tatsuoka, 1998; Buck, Tatsuoka, & Kostin, 1997), mathematics (Birenbaum, Kelly, & Tatsuoka, 1993; Birenbaum, Tatsuoka, & Yamada, 2004; Dogan & Tatsuoka, 2007; K. Tatsuoka, Corter, & Tatsuoka, 2004; K. Tatsuoka et al., 2006), science (Yepes-Baraya & Allen, 2003), and statistics (Im & Yin, 2009).

The rule space method aims at explaining an individual examinee's problem-solving behavior in terms of specific knowledge, processes, and skills that together determine the individual's performance. The specific knowledge, skills, and processes needed for the solution of a set of test items are often termed *attributes*, and the assumptions made in this regard are represented in a *Q-matrix* as binary vectors representing all (1) or none (0) involvement in given test items (K. Tatsuoka, 1990). This process embodies the cognitive specification for test construction with the aim to provide maximum information about the specific attribute profiles of interests (de la Torre, 2008).

Ideally, cognitive diagnosis should proceed using the *true* Q-matrix representing complete and accurate specification of attribute–item involvement. But what if the attributes are incorrectly specified? This could happen for a number of reasons. For example, a diagnostic test in mathematics might be constructed to measure students' knowledge of exponents, and the Q-matrix might be constructed with this goal in mind. But some of the items might also tap skills in factoring or algebra, and these skills might be inadvertently omitted from the Q-matrix. Or, a skill might be presumed to be important in solving items by one particular method, but examinees may find an alternative method that dispenses with the need for that skill. For these and other reasons, the Q-matrix may not be the true Q-matrix, even when proposed by content experts. Therefore, the systematic investigation of possible consequences of attribute misspecification of a Q-matrix is important to establish if this concern is a serious threat to the accuracy and validity of cognitively diagnostic measurement methods.

Two previous studies have investigated the effects of Q-matrix misspecification using such methods. Baker (1993) investigated the sensitivity of the linear logistic test model (LLTM; Fischer, 1973) to Q-matrix misspecification. The two Q-matrices used for his study were composed of eight attributes and 21 items. The number of items was fixed across the misspecification conditions of six levels of error (defined by the percentage of misspecified elements) and four sample size conditions. He found that 5% to 10% misspecification in a Q-matrix seriously degrades attribute and item parameter estimations in the linear logistic test model. The effect of sample size was less substantial than the level of misspecification error. Rupp and Templin (2007) investigated the effects of Q-matrix misspecification on parameter estimates and classification consistency in the Deterministic-Inputs, Noisy ''And'' gate or DINA model

(Junker & Sijtsma, 2001), with controlled misspecification conditions in which certain attributes and items were deleted from the Q-matrix. The Q-matrix used was composed of four attributes and 15 items. They reported consequences at both item and examinee levels: The results showed overestimation of item slip parameters when attributes were deleted from the Q-matrix, overestimation of guessing item parameters when attributes were added to the Q-matrix, and high misclassification rates of examinees into attribute pattern vectors. Together these studies showed that a small degree of misspecification could substantively influence results in the LLTM and DINA models.

The present study investigates the influence of attribute misspecification in the RSM (K. Tatsuoka, 1985, 2009). Specifically, the consequences reported in this study focus on the results regarding estimation of examinees' characteristics, not on item characteristics, because results of the RSM do not include estimation of item characteristics. Examining examinees' attribute mastery parameter estimates and classification is important because the primary purpose of cognitive diagnosis is to diagnose examinees' latent skills and processes.

The effects of misspecification are addressed via a simulation study, as in the two previous studies cited above, because to study the effects of using a "misspecified" Q-matrix, the true Q-matrix used to generate the item responses must be known.

Two types of attribute misspecifications are examined in this study. The first type of misspecification studied is excluding an *essential* attribute that is needed for solving the test items. The second type of misspecification studied is including a *superfluous* attribute that is not necessary to solve the test items. These two types of misspecifications can easily arise from taking approaches to attribute definition that emphasize parsimony or completeness, respectively. The two types of misspecification errors may be illustrated by referring to the Q-matrices shown in Figure 1. First, let us assume that $\mathbf{Q_1}$ with A_1, A_2, A_3, and A_4 is the true Q-matrix. Then $\mathbf{Q_2}$ is a misspecified Q-matrix because essential attribute A_2 is excluded. $\mathbf{Q_3}$ and $\mathbf{Q_4}$ are also misspecified Q-matrices, obtained by omitting essential attributes A_3 and A_4, respectively. Conversely, if we assume that $\mathbf{Q_2}$, composed of three essential attributes A_1, A_3, and A_4, is the true Q-matrix, then $\mathbf{Q_1}$ is a misspecified Q-matrix because attribute A_2 is included superfluously.

The Rule Space Method

The first step of a rule space analysis is attribute specification. As discussed above, this step involves identifying specific attributes assumed to explain a student's performance on each test item and representing them in a Q-matrix, in which rows represent test items and columns represent attributes (K. Tatsuoka, 1990). Each entry q_{jk} in the Q-matrix is equal to 1 if the kth attribute is required to solve the jth item, and 0 otherwise, where $j = 1, \ldots, J$ and $k = 1, \ldots, K$. An example Q-matrix is shown in Table 1. The entries in the first row of the Q-matrix (corresponding to the item I_1) indicate that a student must have mastered A_1 and A_3 to get I_1 correct, whereas only A_3 is required to get I_2 correct.

$$\mathbf{Q}_1 = \begin{array}{cc} & \begin{array}{cccc} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \mathbf{A}_4 \end{array} \\ \begin{array}{c} \mathbf{I}_1 \\ \mathbf{I}_2 \\ \mathbf{I}_3 \\ \mathbf{I}_4 \\ \mathbf{I}_5 \\ \mathbf{I}_6 \end{array} & \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \qquad \mathbf{Q}_2 = \begin{array}{cc} & \begin{array}{ccc} \mathbf{A}_1 & \mathbf{A}_3 & \mathbf{A}_4 \end{array} \\ \begin{array}{c} \mathbf{I}_1 \\ \mathbf{I}_2 \\ \mathbf{I}_3 \\ \mathbf{I}_4 \\ \mathbf{I}_5 \\ \mathbf{I}_6 \end{array} & \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$\mathbf{Q}_3 = \begin{array}{cc} & \begin{array}{ccc} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_4 \end{array} \\ \begin{array}{c} \mathbf{I}_1 \\ \mathbf{I}_2 \\ \mathbf{I}_3 \\ \mathbf{I}_4 \\ \mathbf{I}_5 \\ \mathbf{I}_6 \end{array} & \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{array} \qquad \mathbf{Q}_4 = \begin{array}{cc} & \begin{array}{ccc} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 \end{array} \\ \begin{array}{c} \mathbf{I}_1 \\ \mathbf{I}_2 \\ \mathbf{I}_3 \\ \mathbf{I}_4 \\ \mathbf{I}_5 \\ \mathbf{I}_6 \end{array} & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array}$$

Figure 1. Example Q-matrices to illustrate attribute exclusion or attribute inclusion

Next, all possible attribute patterns, called *knowledge states*, are generated. For the example Q-matrix given in Table 1, the number of possible attribute patterns is $2^3 = 8$. Knowledge states in the RSM are represented in the form of a binary vector indicating which attributes have been mastered. For example, an attribute pattern (011) indicates a knowledge state of knowing A_2 and A_3 but not knowing A_1. The knowledge states correspond to certain item response patterns, termed as *ideal response patterns*. Formally, let $\mathbf{S_t}$ be the tth knowledge state ($t = 1, \ldots, T$) composed of K attributes and let $\mathbf{R_t}$ be the corresponding ideal response pattern composed of J items. The relationship between a knowledge state ($\mathbf{S_t} = S_{t1}, \ldots, S_{tK}$) and the corresponding ideal response pattern ($\mathbf{R_t} = R_{t1}, \ldots, R_{tJ}$) is given by a Boolean function as follows (K. Tatsuoka, 1991):

$$R_{tj} = \prod_{k=1}^{K} S_{tk}^{Q_{jk}}. \tag{1}$$

This equation of the production rule has a conjunctive nature, in which the examinee gets the item correct only if all involved attributes are mastered. Table 1 illustrates the possible knowledge states and ideal response patterns identified by the example Q-matrix. It is important to note that mapping of knowledge states to ideal response patterns is not a one-to-one correspondence; rather, given a particular set of items and a particular Q-matrix, two or more different knowledge states can result in the same ideal response pattern. For example, two different knowledge states, $\mathbf{S_1}$(000) and $\mathbf{S_2}$(100), correspond to the same ideal response pattern $\mathbf{R_1}$(00000) because no

Table 1. Generated Knowledge States and Ideal Response Patterns Implied by the Q-Matrix

Knowledge states, **S** (A_1 A_2 A_3)	Example Q-matrix	Ideal response patterns, **R** (I_1 I_2 I_3 I_4 I_5)
S_1(000)	$\begin{array}{c c} & \begin{array}{ccc} A_1 & A_2 & A_3 \end{array} \\ \begin{array}{c} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \end{array} & \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$	R_1(00000)
S_2(100)		R_1(00000)
S_3(010)		R_3(00001)
S_4(001)		R_4(01000)
S_5(110)		R_3(00001)
S_6(101)		R_6(11000)
S_7(011)		R_7(01101)
S_8(111)		R_8(11111)

items in the Q-matrix require knowing A_1 only. Knowing A_1 will only result in an item being performed correctly if A_3 is known as well. Similarly, $\mathbf{S_3}$(010) and $\mathbf{S_5}$(110) share the same ideal response pattern $\mathbf{R_3}$(00001).

It is important to note that attribute vectors in the Q-matrix can be compared, and ordered, using Boolean algebraic structures, because ordering of attributes can affect the identifiability of knowledge states. First, consider any attribute vector $\mathbf{A_m}$ to be a set whose elements are the items that require the subskill represented by the attribute $\mathbf{A_m}$. Then, for any two distinct attribute vectors $\mathbf{A_m}$ and $\mathbf{A_n}$ the set relationship $\mathbf{A_m} \subset \mathbf{A_n}$ induces an *order relationship*, <, between the two attributes, $\mathbf{A_m} < \mathbf{A_n}$. In this case, it follows that $\mathbf{A_m} + \mathbf{A_n} = \mathbf{A_n}$ and $\mathbf{A_m} \times \mathbf{A_n} = \mathbf{A_m}$. $\mathbf{A_n}$ is referred to as a *subset* of $\mathbf{A_m}$ and $\mathbf{A_m}$ a *superset* of $\mathbf{A_n}$. For example, if $\mathbf{A_1}$ is a subset of $\mathbf{A_3}$ and $\mathbf{A_3}$ a superset of $\mathbf{A_1}$, then the order relation between $\mathbf{A_1}$ and $\mathbf{A_3}$ is defined as $\mathbf{A_1} < \mathbf{A_3}$. Then $\mathbf{A_1} + \mathbf{A_3} = (10010)' + (11110)' = (11110)' = \mathbf{A_3}$ and $\mathbf{A_1} \times \mathbf{A_3} = (10010)' \times (11110)' = (10010)' = \mathbf{A_1}$. In general, in Boolean algebraic operations $1 + 1 = 1$, $1 + 0 = 0 + 1 = 1, 0 + 0 = 0, 1 \times 1 = 1, 1 \times 0 = 0 \times 1 = 0$, and $0 \times 0 = 0$.

The third step of a rule space analysis begins by mapping the observed responses $\mathbf{X}$ and ideal responses $\mathbf{R}$ onto a latent "rule space" composed of dimensions θ and ζ, where θ is the latent ability variable in item response theory and ζ is one of the item response theory–based caution indices, measuring atypicality of item responses (K. Tatsuoka, 1985). In other words, ζ indicates the tendency of a response pattern to answer difficult items correctly and easy items incorrectly. Formally, it takes a standardized form of the conditional covariance of two residual vectors, $\mathbf{P}(\theta) - \mathbf{x}$ and $\mathbf{P}(\theta) - \mathbf{T}(\theta)$, where $P_j(\theta)$ are the two-parameter logistic model probabilities for the J items, $\mathbf{x}$ is a binary response vector, and $T(\theta)$ is the mean of $P_j(\theta)$:

$$\zeta = \frac{\sum_j (P_j(\theta) - x_j)(P_j(\theta) - T(\theta))}{\sqrt{\sum_j P_j(\theta)(1 - P_j(\theta)(P_j(\theta) - T(\theta))^2}}. \tag{2}$$

The two axes of θ and ζ are proven to be orthogonal as long as local independence holds among the items (K. Tatsuoka, 1985). In some applications, it may be useful to incorporate more than one ζ dimension, if deviations of response patterns from a unidimensional ability continuum show several distinct patterns. The additional ζ measure the atypicality of item responses within subsets of the given test (K. Tatsuoka, 1996). The coordinates of ideal responses in the rule space, $\{(\theta_{Rt}, \zeta_{Rt})\ t = 1, \ldots, T\}$, are termed *centroids*, symbolizing the locations of the knowledge states. The projected points, $\{(\theta_{xi}, \zeta_{xi})\ i = 1, \ldots, N\}$, of examinees' observed responses **X** in the rule space are assumed to deviate from the ''true'' centroid because the examinees' observed responses reflect slips. Slips can be of two types: those because of lucky guesses without knowledge (up-slips) or because of careless mistakes with knowledge (down-slips). Thus, the observed examinee points form a cluster around the centroid.

The last step begins by examining the squared Mahalanobis distances between each examinee's coordinates (θ, ζ) and the centroids corresponding to ideal response patterns. The squared Mahalanobis distance is a chi-square variate with degrees of freedom equal to the number of dimensions of the rule space (K. Tatsuoka & Tatsuoka, 1987). If the distance between each examinee's location and some centroids are closer than a specified distance cutoff, the centroids are accepted as candidate knowledge states for that examinee. In addition, Bayes's decision rule for minimum error is also applied because the distance measure itself does not provide probabilities of misclassification. This procedure provides the posterior probabilities associated with being a member of the knowledge states for each examinee (K. Tatsuoka & Tatsuoka, 1987). The attribute mastery vector of each examinee is calculated in a probabilistic form using the binary attributes in the candidate knowledge states and the posterior probabilities corresponding to the knowledge states. Let a_{kt} be the attribute mastery (1 or 0) of the kth attribute in the tth knowledge state and w_t be the posterior probability corresponding to the tth knowledge state. The probability of mastery of an attribute k by an examinee i is calculated as follows (K. Tatsuoka, 2009):

$$P(a_k = 1|x_i) = \sum_t w_t \cdot a_{tk}. \tag{3}$$

This probability for examinee i is called the *attribute mastery probability* (AMP) of the examinee for the attribute k. When specific knowledge states for the examinee are identified and corresponding AMPs are estimated, then the cognitive diagnosis for the examinee is complete.

Changes in Knowledge States Because of Attribute Misspecification

In considering possible consequences of attribute misspecification, it is useful to trace through these possible effects in different stages of the RSM, to generate hypotheses about the effects of attribute exclusion or inclusion on the identification of knowledge states and classification of examinees.

Table 2. Knowledge States Generated by the Q-Matrices and Their Corresponding Ideal Response Patterns

Knowledge states generated by				
$\mathbf{Q_1}$	$\mathbf{Q_2}$(no A_2)	$\mathbf{Q_4}$(no A_4)	$\mathbf{Q_3}$(no A_3)	Corresponding ideal response patterns
S_1(1111)	T_1(111)	V_1(111)	U_1(111)	→ R_1(111111)
S_2(1011)		V_2(101)	U_2(101)	→ R_2(010111)
S_3(1110)	T_2(110)		U_3(110)	→ R_3(111000)
S_5(0111)	T_3(011)	V_3(011)	U_4(011)	→ R_4(001101)
S_4(1010)			U_5(100)	→ R_5(010000)
S_6(0110)	T_4(010)		U_6(010)	→ R_6(001000)
S_7(1001)	T_5(101)	V_4(100)		→ R_7(000011)
S_8(1000)	T_6(100)			→ R_{10}(000000)
S_9(0011)		V_5(001)	U_7(001)	→ R_8(000101)
S_{10}(0010)			U_8(000)	→ R_{10}(000000)
S_{11}(0001)	T_7(001)	V_6(000)		→ R_9(000001)
S_{12}(0000)	T_8(000)			→ R_{10}(000000)

Table 2 illustrates sets of knowledge states generated by $\mathbf{Q_1}$ to $\mathbf{Q_4}$ (see Figure 1) and the corresponding ideal response patterns identified through Equation (1). Let us use $\mathbf{Q_1}$ and $\mathbf{Q_2}$ for the illustration. $\mathbf{Q_1}$ generates 12 knowledge states for the attributes A_1 to A_4, $\mathbf{S_1}$ to $\mathbf{S_{12}}$, and 10 corresponding ideal response patterns, $\mathbf{R_1}$ to $\mathbf{R_{10}}$; $\mathbf{Q_2}$ generates 8 knowledge states for the attributes A_1, A_3, and A_4, $\mathbf{T_1}$ to $\mathbf{T_8}$, and 7 corresponding ideal response patterns, $\mathbf{R_1}$, $\mathbf{R_3}$, $\mathbf{R_4}$, $\mathbf{R_6}$, $\mathbf{R_7}$, $\mathbf{R_9}$, and $\mathbf{R_{10}}$. Note that the ideal response patterns generated by $\mathbf{Q_2}$ are a subset of the ideal response patterns generated by $\mathbf{Q_1}$.

Excluding an Essential Attribute

The direct consequence of exclusion of an essential attribute is the elimination of essential knowledge states and the corresponding ideal response patterns. To illustrate, let us assume that $\mathbf{Q_1}$ is the true Q-matrix and $\mathbf{Q_2}$ is a misspecified Q-matrix. As can be seen in Table 2, exclusion of A_2 results in erroneous elimination of four knowledge states and three ideal response patterns $\mathbf{R_2}$, $\mathbf{R_5}$, and $\mathbf{R_8}$. The general point is that attribute exclusion does not transform the ideal response patterns into different response patterns but simply leads to a decrease in the number of defined or identified ideal response patterns.

The effects of attribute misspecification on classification of examinees can be better understood with a graphical illustration of the rule space composed of (θ, ζ) pairs, where specific locations (centroids) correspond to specific knowledge states and/or ideal response patterns. Figure 2 displays 10 hypothetical centroids $\mathbf{C}(\theta_t, \zeta_t)$, where $t = 1, \ldots, 10$, corresponding to the 10 ideal response patterns identified by $\mathbf{Q_1}$ in

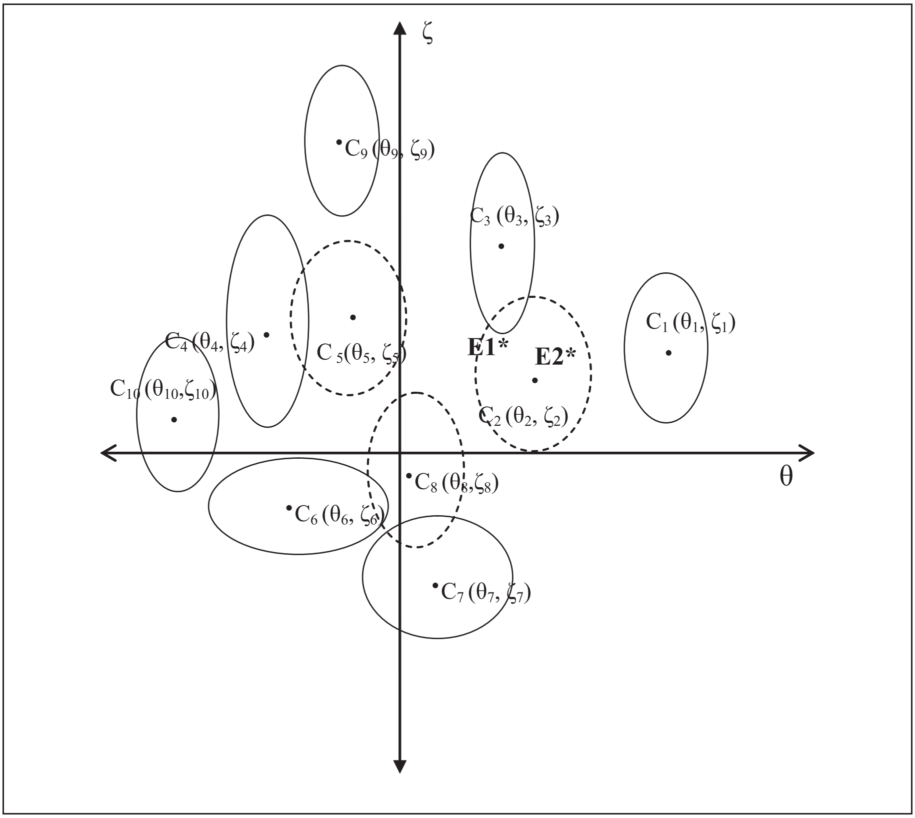

Figure 2. Hypothetical ellipses of 10 centroids projected onto the rule space

Table 2. The ellipses around the centroids signify the squared Mahalanobis distance cutoff for the classification of examinees. Note that C_2, C_5, and C_8 were eliminated when the attribute A_2 was excluded. The ellipses around the eliminated centroids, C_2, C_5, and C_8, are shown with dotted lines.

Let us assume that two hypothetical examinees E1 and E2 fall within the ellipse of C_2 when diagnosed using the true Q-matrix $\mathbf{Q_1}$. If attribute A_2 is excluded, several ellipses including C_2 will be eliminated, and the two examinees E1 and E2 will be reassigned into ellipses of the remaining centroids (if possible). In general, any examinees who are located within the eliminated knowledge state ellipses will be reclassified into the remaining knowledge states based on proximity, using the squared Mahalanobis distance from their locations in the rule space. With the exclusion of attribute A_2, examinee E1's closest ellipse becomes the one labeled C_3. But examinee E2 cannot be successfully reclassified given the specified distance cutoff because he or she is located outside the remaining ellipses. To sum up, attribute exclusion leads to elimination of some knowledge states. Accordingly, examinees who were positioned near the eliminated knowledge states might be either not classified or *underclassified*, meaning that they have fewer memberships in knowledge states than they are supposed to have. Another consequence of the misspecification, besides this lower successful classification rate, is that the AMPs of the examinees will be estimated based on the erroneous set of knowledge states, leading to higher error or possibly to bias.

Including a Superfluous Attribute

The inclusion of a superfluous attribute introduces some superfluous knowledge states. To illustrate, we assume $\mathbf{Q_2}$ to be the true Q-matrix and $\mathbf{Q_1}$ to be the misspecified one (see Table 2). Now ideal response patterns $\mathbf{R_2}$, $\mathbf{R_5}$, and $\mathbf{R_8}$ are considered to be superfluous because they are added because of erroneous inclusion of A_2. Thus, attribute inclusion merely leads to an increase in the number of possible ideal response patterns rather than transforming the ideal response patterns into different response patterns.

Hypotheses about the effects of attribute inclusion on examinees' classification and AMPs can be similarly inferred by taking the opposite approach to that used with attribute exclusion. If $\mathbf{Q_2}$ is the true Q-matrix, then $\mathbf{Q_1}$ represents the Q-matrix obtained by the inclusion of superfluous attribute A_2. The dotted ellipses in Figure 2 are now considered superfluous centroids. Then examinee E2 who initially does not belong to any ellipse now acquires a new membership in C_2. After inclusion of A_2, examinee E1 who had one membership in ellipse C_3 is now judged to have two memberships, in C_2 and C_3. Attribute inclusion results in superfluous creation of knowledge states that likely inject noise into the estimation process. Thus, examinees positioned around the superfluously added knowledge states would be *overclassified*, meaning that the examinees have more memberships than they are supposed to have. Consequently, the AMPs of the examinees are estimated based on a mix of valid and invalid memberships.

Effects of Order Relations Among Attributes

As previously mentioned, an order relation between attributes affects the identification of possible knowledge states. Within a single Q-matrix, attribute subset–superset relationships have the consequence that not all knowledge states can be identified. This point was made earlier, where it was pointed out that some distinct knowledge states may correspond to the same ideal response pattern. Specifically, knowledge states that are not generated in the RSM because of the order relation between the subset A_2 and superset A_3 in $\mathbf{Q_1}$ (see Figure 1 and Table 2) include (1101), (1100), (0101), and (0100). Thus, the number of possible knowledge states is reduced from 16 ($=2^4$) to 12 because of the order relation. Therefore, it is reasonable to hypothesize that order relationships among attributes may moderate the effects of attribute misspecification. For example, excluding the subset attribute A_2, or the superset attribute A_3, may lead to different consequences in terms of classification and estimation bias.

Summary

The previous section illustrated changes in the set of candidate knowledge states because of exclusion of an essential attribute and inclusion of a superfluous attribute. Attribute exclusion results in elimination of some knowledge states, which causes examinees being underclassified. In contrast, attribute inclusion results in additional knowledge states where examinees can be overclassified. The specific research questions addressed here are (a) how attribute misspecification affects the classification consistency between examinees' true mastery for each attribute and the estimated mastery and (b) how attribute exclusion and attribute inclusion affect the estimated AMPs of examinees. In addition to these two questions, we further examine (c) how the order relations of an excluded/included attribute with the other attributes affect the estimation of AMPs.

Method

Data Simulation

To maintain fidelity by using a realistic attribute structure, a Q-matrix constructed for an actual test is used. Table 3 shows a Q-matrix specifying attributes of a mixed-fraction subtraction test with 20 items constructed by Tatsuoka (1984), which has been also referred to by other studies of cognitive diagnostic measurement models (e.g., de la Torre & Douglas, 2004; C. Tatsuoka, 2002). The seven attributes were involved with the 20 items with varying degrees: A_6 is associated with only three items whereas A_7 is associated with all the 20 items. This Q-matrix is used as the base model to create Q-matrices with correct and incorrect attribute specifications.

A total of eight Q-matrices, $\mathbf{Q_0}$ to $\mathbf{Q_7}$, were used to simulate item responses. The Q-matrix in Table 1 is labeled $\mathbf{Q_0}$; it includes all seven attributes, A_1 to A_7. $\mathbf{Q_1}$ includes all attributes except A1, $\mathbf{Q_2}$ all attributes but A_2, and similarly for $\mathbf{Q_3}$ to $\mathbf{Q_7}$. Accordingly, a total of eight data sets were simulated, using $\mathbf{Q_0}$ to $\mathbf{Q_7}$ as the true knowledge structures. The number of simulated examinees per identified knowledge state was 100. This resulted in different sample sizes for each simulated data set, because the different ''true'' Q-matrices generated different number of knowledge states. Table 4 shows the design used in generating the eight simulated data sets, and the effective sample sizes.

To examine the effects of attribute exclusion, $\mathbf{Q_0}$ was used as the true Q-matrix, and the other Q-matrices, $\mathbf{Q_1}$ to $\mathbf{Q_7}$, were then considered as misspecified Q-matrices. For example, $\mathbf{Q_1}$ composed of A_2 to A_7 is a misspecified Q-matrix because it does not include A_1. A total of seven comparisons between $\mathbf{Q_0}$ and other misspecified Q-matrices (because of attribute exclusion) were performed. To examine the effects of including a superfluous attribute, the comparisons were reversed: $\mathbf{Q_1}$ to $\mathbf{Q_7}$ were considered the true Q-matrices (and used to generate simulated examinees), whereas $\mathbf{Q_0}$ was considered to be the misspecified attribute matrix.

Table 3. Q-Matrix for the Mixed Fraction Subtraction Test

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
I_1	0	0	0	1	0	1	1
I_2	0	0	0	1	0	0	1
I_3	0	0	0	1	0	0	1
I_4	0	1	1	0	1	0	1
I_5	0	1	1	1	0	0	1
I_6	0	0	0	0	0	0	1
I_7	1	1	1	0	0	0	1
I_8	0	0	0	0	0	0	1
I_9	1	1	0	0	0	0	1
I_{10}	0	1	1	0	1	1	1
I_{11}	0	1	1	0	1	0	1
I_{12}	0	0	1	0	0	0	1
I_{13}	0	1	0	1	1	0	1
I_{14}	0	1	0	0	0	0	1
I_{15}	1	1	1	0	0	0	1
I_{16}	0	1	0	0	0	0	1
I_{17}	0	1	0	0	1	0	1
I_{18}	0	1	1	0	1	1	1
I_{19}	1	1	1	0	1	0	1
I_{20}	0	1	1	0	1	0	1

Note. Adapted from Tatsuoka (1984).

Table 4. Simulation Design for Attribute Exclusion and Attribute Inclusion

True versus misspecified Q-matrix	Response set	Slip probability[a]	Simulated N	Classified N[b]	Percentage classified
Attribute exclusion					
$\mathbf{Q_0}$ vs. $\mathbf{Q_1}$	Set 0	(.00, .10)	3,300	2,994	90.75
$\mathbf{Q_0}$ vs. $\mathbf{Q_2}$	Set 0	(.00, .10)	3,300	3,299	99.97
$\mathbf{Q_0}$ vs. $\mathbf{Q_3}$	Set 0	(.00, .10)	3,300	3,158	95.67
$\mathbf{Q_0}$ vs. $\mathbf{Q_4}$	Set 0	(.00, .10)	3,300	3,059	92.70
$\mathbf{Q_0}$ vs. $\mathbf{Q_5}$	Set 0	(.00, .10)	3,300	3,059	92.70
$\mathbf{Q_0}$ vs. $\mathbf{Q_6}$	Set 0	(.00, .10)	3,300	3,266	98.97
$\mathbf{Q_0}$ vs. $\mathbf{Q_7}$	Set 0	(.00, .10)	3,300	3,263	98.88
Attribute inclusion					
$\mathbf{Q_1}$ vs. $\mathbf{Q_0}$	Set 1	(.00, .10)	2,000	2,000	100.00
$\mathbf{Q_2}$ vs. $\mathbf{Q_0}$	Set 2	(.00, .10)	2,700	2,700	100.00
$\mathbf{Q_3}$ vs. $\mathbf{Q_0}$	Set 3	(.00, .10)	1,800	1,800	100.00
$\mathbf{Q_4}$ vs. $\mathbf{Q_0}$	Set 4	(.00, .10)	2,100	2,100	100.00
$\mathbf{Q_5}$ vs. $\mathbf{Q_0}$	Set 5	(.00, .10)	2,100	2,100	100.00
$\mathbf{Q_6}$ vs. $\mathbf{Q_0}$	Set 6	(.00, .10)	2,100	2,100	100.00
$\mathbf{Q_7}$ vs. $\mathbf{Q_0}$	Set 7	(.00, .10)	3,200	3,200	100.00

a. Slip probability = probability {up-slip + down-slip}.
b. The number of classified examinees after the rule space analyses using misspecified Q-matrices, given the classification cutoff of 4.605, the critical value of χ^2 with two degrees of freedom, $p > .90$.

Evaluation

As a first step, rule space analyses were conducted on the eight simulated data sets using the BUGLIB program (C. Tatsuoka, Varadi, & Tatsuoka, 1992). Then the effects of attribute specification were measured in two ways. First, an evaluation was made of the consistency in classifying examinees' mastery for each attribute, based on the first knowledge state with the highest posterior probabilities. Specifically, to assess the classification consistency between examinees' true attribute mastery and the estimated mastery given attribute misspecification, percent agreement (p) and Cohen's kappa (K) are reported. Second, to assess errors in the estimated AMPs, the root mean square errors (RMSEs) showing the magnitude of errors and the mean bias showing the directionality of errors are reported along with the standard deviation.

Results

The consequences of attribute misspecification could be examined only for examinees successfully classified into any of the knowledge states, because those who were not assigned into any states could not be further analyzed in the RSM. The number and percentages of the simulated examinees who were successfully classified into any of the knowledge states are reported in Table 4. When a superfluous attribute is included, the successful classification rate is 100%, no matter which superfluous attribute was added. In contrast, when an essential attribute was excluded, the rate of the successfully classified examinees was lower. The percentages of successfully classified examinees were in general high, with rates varying from 90.75% to 99.97%, depending on which specific attribute was excluded.

Effects of Attribute Exclusion and Attribute Inclusion

Table 5 presents a summary of the classification consistencies between examinees' true attribute mastery and the estimated mastery given attribute misspecification and of the errors in the estimated AMPs under attribute misspecification. The classification consistencies across the different cases of attribute misspecification were generally high. The classification consistencies for attribute exclusion were lower than the values for attribute inclusion across all seven misspecification cases from A_1 to A_7. The magnitude of the errors is given by the RMSEs and the directionality of the errors is given by the means and ranges of the biases. The RMSEs of the estimated AMPs because of inclusion of a superfluous attribute are much smaller ($M = .038$) than those because of exclusion of an essential attribute ($M = .138$). The direction of the estimation bias differs for the two types of specification errors. When an essential attribute is excluded the directional bias of the errors for the remaining AMPs is generally negative ($M = -.031$), meaning that the estimated AMPs tend to be underestimated. When a superfluous attribute is included, the AMPs are generally overestimated for the other attributes, resulting in a positive bias ($M = .015$).

Table 5. Classification Consistency of Attribute Mastery in the First Knowledge States and Error in AMP When an Essential Attribute Is Excluded or When a Superfluous Attribute Is Included

	Classification		Error				
				Bias			
	p	K	RMSE	Mean	*SD*	Minimum	Maximum
Essential attribute excluded							
A_1	.949	.886	.135	−.018	.141	−.854	.621
A_2	.951	.813	.123	−.021	.124	−1.000	.821
A_3	.881	.670	.217	−.079	.211	−1.000	.718
A_4	.879	.738	.163	−.035	.159	−1.000	.757
A_5	.862	.734	.199	−.047	.211	−1.000	.812
A_6	.943	.871	.122	−.014	.131	−1.000	.757
A_7	1.000	1.000	.004	.000	.006	.000	.196
Superfluous attribute included							
A_1	.977	.950	.056	.013	.123	−.580	.854
A_2	.982	.893	.019	.011	.080	−.462	.810
A_3	.963	.856	.046	.035	.122	−.718	1.000
A_4	.985	.903	.025	.011	.094	−.757	.872
A_5	.955	.896	.055	.023	.160	−.812	1.000
A_6	.951	.893	.060	.014	.133	−.696	1.000
A_7	1.000	1.000	.002	−.000	.003	−.089	.000

Note. AMP = attribute mastery probability; RMSE = root mean square error.

Exclusion or inclusion of A_7 is the only exception to the general pattern, showing essentially zero bias, but a slight tendency toward the opposite pattern. Some analyses investigating why A_7 might be an exception to the general pattern are presented in the next section.

Effects of Attribute Order Relations in the Q-Matrix

As discussed above, an order relation between attribute vectors can affect the number of possible knowledge states and accordingly the number of ideal response patterns. This section focuses on the results of misspecification of attributes that have order relations with other attributes in the Q-matrix. Figure 3 presents all the order relations among the seven attributes. A_7 is a superset of all the other attributes, because all the 20 items are involved with A_7. A_2 is a superset of A_1 and A_5 and is a subset of A_7 at the same time.

Attribute exclusion. When an essential subset attribute was excluded, the RMSEs of superset attributes were higher than the RMSEs of subset attributes when a superset attribute was excluded (see Table 6). There was consistent directionality in biases when an excluded attribute is a superset or subset of some remaining attributes.

$$
\begin{array}{ccccc}
\mathbf{A_1} & \le & \mathbf{A_2} & \le & \mathbf{A_7} \\
\mathbf{A_3} & & & \le & \mathbf{A_7} \\
\mathbf{A_4} & & & \le & \mathbf{A_7} \\
\mathbf{A_5} & \le & \mathbf{A_2} & \le & \mathbf{A_7} \\
\mathbf{A_6} & & & \le & \mathbf{A_7}
\end{array}
$$

Figure 3. Order relations among the attributes

When a subset attribute is excluded, the biases of the estimated AMPs of superset attributes are nonpositive (upper-bounded at 0), indicating that all the examinees' AMPs in the superset attributes are underestimated or remain the same. In contrast, when an excluded attribute is a superset of some remaining attributes, the biases of the estimated AMPs of subset attributes are nonnegative, that is., lower-bounded at 0. Thus, all the examinees' estimated AMPs in the subset attributes are overestimated or remain the same after excluding a superset attribute.

Attribute inclusion. When a superfluous subset attribute was included the RMSEs of the remaining (superset) attributes were higher than the RMSEs of the remaining (subset) attributes when a superset attribute was included. When a subset attribute was included, the biases of the estimated AMPs in superset attributes are lower-bounded at 0. This indicates that all the estimated AMPs in the remaining (superset) attributes are overestimated or remain the same after superfluously including a subset attribute. In contrast, when an included attribute is a superset of some remaining attributes, the biases of the estimated AMPs of the remaining (subset) attributes are upper-bounded at 0, meaning that all the estimated AMPs in the subset attributes are underestimated or remain the same.

Summary and Discussion

The results show consistent bias from attribute misspecification and show bias in opposite directions from excluding an essential attribute and including a superfluous attribute. However, as a practical matter the amount of bias is not particularly large. The classification consistencies between the simulated examinees' true attribute mastery and the estimated mastery after attribute misspecification were high, RMSEs were not large, and mean biases were close to zero. One of the possible reasons would be that, as shown in the illustration, attribute exclusion/inclusion did not result in complete changes in knowledge states, but in elimination/addition of some knowledge states because of the characteristics of the Boolean function (Equation [3]).

Several specific conclusions seem warranted by the results. First, when an essential attribute was excluded, the classification consistencies of examinees' attribute mastery were lower than the consistencies when a superfluous attribute is included. This may be attributed to the fact that attribute exclusion results in reclassification of some examinees who belong to the excluded knowledge states; attribute inclusion

Table 6. Errors in AMPs When a Misspecified Attribute Is in Order Relation With the Remaining Attributes

		Attribute exclusion					Attribute inclusion				
			Bias					Bias			
Misspecified attribute	Remaining attribute	RMSE	Mean	*SD*	Minimum	Maximum	RMSE	Mean	*SD*	Minimum	Maximum
Subset	Superset										
A_1	A_2	.151	−.072	.133	−.676	.000	.145	.068	.128	.000	.676
A_1	A_7	.015	−.002	.229	−.147	.000	.018	.002	.018	.000	.147
A_2	A_7	.134	−.028	.015	−.825	.000	.105	.016	.104	.000	.810
A_3	A_7	.074	−.016	.131	−.656	.000	.082	.017	.081	.000	.656
A_4	A_7	.157	−.036	.073	−.872	.000	.120	.021	.118	.000	.872
A_5	A_2	.258	−.117	.153	−1.000	.000	.228	.089	.213	.000	1.000
A_5	A_7	.017	−.002	.017	−.168	.000	.020	.003	.048	.000	.168
A_6	A_7	.001	.000	.001	−.042	.000	.001	.000	.001	.000	.042
Superset	Subset										
A_2	A_1	.102	.038	.095	.000	.636	.064	−.015	.090	−.357	.000
A_7	A_2	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
A_7	A_3	.005	.000	.005	.000	.118	.004	.000	.004	−.089	.000
A_7	A_4	.012	.002	.012	.000	.196	.007	−.001	.007	−.081	.000
A_2	A_5	.004	.000	.004	.000	.194	.000	.000	.000	.000	.000
A_7	A_5	.091	.026	.087	.000	.882	.054	−.012	.081	−.462	.000
A_7	A_6	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
A_2	A_1	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

Note. AMP = attribute mastery probability; RMSE = root mean square error.

results in some examinees near the superfluous knowledge states in the rule space being overclassified because they now have membership in one or more additional knowledge states. Thus, inclusion of a superfluous attribute was less influential to the reclassified examinees. We also found that there were consistent asymmetric reclassifications because of both attribute exclusion and attribute inclusion: When an essential attribute is excluded, the rate of reclassified examinees moving from 1 (mastery) to 0 (nonmastery) in their first knowledge state was 5.4% and the rate of moving from 0 to 1 was 2.3% on average. In contrast, when a superfluous attribute is included, the rate of reclassified examinees moving from 1 to 0 was about 0.8% and the rate of moving from 0 to 1 was 1.9% on average. This shows that with attribute exclusion, there were more examinees who move from a mastery state to a nonmastery state, whereas attribute inclusion results in more examinees moving from a nonmastery state to a mastery state. This finding could be further investigated to establish its generalizability across various Q-matrix specification conditions.

Second, when an essential attribute was excluded, the mean biases of the estimated AMPs for the remaining attributes were negative, meaning that the AMPs were underestimated. The exception was the case of attribute A_7, which resulted in zero or slight overestimation. This exceptional behavior apparently occurred because A_7 is a ‘‘universal’’ attribute, that is, an attribute required by all items, meaning that it is a superset of any other attribute. When a superfluous attribute is included, the mean biases of the AMPs for the other attributes are positive, indicating that the AMPs of examinees are overestimated. Again, universal attribute A_7 was the exception, resulting in zero or slight underestimation. The underestimation of AMPs because of attribute exclusion and overestimation of AMPs because of attribute inclusion might be explained in relation to the results of classification. As Equation (3) makes clear, an examinee's AMP is an average of the attribute mastery values (1 or 0) defined by the relevant knowledge states, after weighting by the posterior probabilities of the knowledge states. Thus, the asymmetric reclassification changes, such that more examinees move from 1 to 0 because of attribute exclusion and more examinees move from 0 to 1 because of attribute inclusion. This directly leads to the negative bias for attribute exclusion and the positive bias for attribute inclusion, respectively.

Another possible explanation for the pattern of biases because of attribute exclusion and attribute inclusion can be constructed, based on general arguments about the nature of the modeling task in cognitive diagnosis using any conjunctive CDM, which assumes that all involved attributes must be mastered by the examinee to correctly answer the item (as in Equation [1]). Since the goal of any error-based modeling exercise is to match the predicted item response vectors to the observed vectors, excluding an essential attribute that is not mastered by some examinees will tend to raise the mean of the predicted response vector, unless ‘‘adjustments’’ (i.e., negative biases) are made to the AMPs of the remaining attributes. Similarly, including a superfluous attribute will tend to depress the mean predicted response vector (unless the superfluous attributes were assumed to be mastered by examinees), requiring positive ‘‘adjustments’’ (biases) to bring the mean level of the predicted item responses in line with the mean observed item response vectors.

A third general finding was that when an essential attribute is excluded, the RMSEs of the estimated AMPs were larger than the RMSEs when a superfluous attribute is included. One possible reason can be found in the larger percentages of reclassified examinees because of attribute exclusion rather than because of attribute inclusion, as no errors arise for the examinees who were not reclassified. The percentages of examinees reclassified because of attribute exclusion were 7.7 on average, whereas the percentages reclassified because of attribute inclusion were 2.7 on average.

To summarize, the effects of excluding an essential attribute are more detrimental to the classification of examinees' attribute mastery and estimation of examinees' AMPs, compared with the effects of including a superfluous attribute. These findings have an implication for attribute specification in test construction, namely, that content experts should not drop an important attribute in problem solving, for example, when it is considered as a basic attribute used for majority of the test items.

Effects of Order Relations Between Attributes

When an attribute is in an order relation with a misspecified attribute, the order relationships were associated with a systematic pattern in the directions of the biases of the estimated AMPs. The effects found were the following:

1. When an excluded essential attribute is the subset of some remaining attributes, all the examinees' AMPs on those attributes are underestimated or remain the same. In contrast, when an excluded attribute is the superset of some remaining attributes, all the examinees' AMPs on those attributes are overestimated or remain the same.
2. The effects of attribute inclusion are complementary to the effects of attribute exclusion. When an included superfluous attribute is the subset of some attributes, all the examinees' AMPs on those attributes are overestimated or remain the same. In contrast, when an included superfluous attribute is the superset of some attributes, all the examinees' AMPs on those attributes are underestimated or remain the same. The consistent patterns in overestimation/underestimation of AMPs were found across different attributes regardless of different degrees of involvement with the items. The effects of attribute inclusion were complementary.
3. RMSEs of the superset attributes because of misspecification of subset attributes were bigger than RMSEs of the subset attributes because of misspecification of the superset attributes. The sizes of the RMSEs were varied by different attributes.

Implications

Future research might investigate other types of attribute misspecification. An example is random misspecification (e.g., Baker, 1993) of elements in an attribute vector of a Q-matrix arising from random errors by coders, corresponding to low reliability in coding a specific attribute. That type of misspecification is different from the attribute exclusion/inclusion examined here, because misspecification of individual elements in the Q-matrix would instigate changes of locations of centroids in the rule space, instead of eliminating/adding some centroids without changing their locations in the rule space. The degree of changes of centroids' locations would be related to the number of misspecified elements, and examinees' attribute mastery would be changed depending on the changes of the locations of knowledge states in the rule space.

Here, the effects of misspecification error were mediated by order relations among attributes. Order relations among attributes can be created by the existence of prerequisite, necessary, and/or logically dependent skills and by the coding of multilevel intensity of attribute involvement. But this type of specification might also occur merely because of poor test construction that fails to create items with all possible combinations of attributes. For example, a test designer might include items that test essential attributes A_1 and A_2 together, and some items that test A_2 alone, but no items that test A_1 alone. Then attribute A_1 will be a subset of A_2, merely as an artifact of the particular design of the item set. In this case, order relationships among attributes are not logically determined; rather, they arise from flaws or gaps in test design. We have shown here that such haphazard test construction could bias the estimation of AMPs in predictable ways. A final issue worthy of future investigation is suggested by the fact that the sizes of RMSEs varied among the various attributes with different degrees of item involvement. It may be that attribute nesting is not the only factor that leads to estimation bias; other related factors such as overall level of item involvement might play a role as well. Future research might test this possibility as part of a broad program to find and mitigate factors that lead to bias in diagnostic measurement methods.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

This research was supported by a grant from the National Science Foundation Fund (REC. 0126064).

References

Baker, F. B. (1993). Sensitivity of the linear logistic test model to misspecification of the weight matrix. *Applied Psychological Measurement*, *17*, 201-210.

Birenbaum, M., Kelly, A. E., & Tatsuoka, K. (1993). Diagnosing knowledge states in algebra using the rule space model. *Journal for Research in Mathematics Education*, *24*, 442-459.

Birenbaum, M., Tatsuoka, C., & Yamada, Y. (2004). Diagnostic assessment in TIMMS-R: Between countries and within country comparisons of eight graders' mathematics performance. *Studies in Educational Evaluation*, *30*, 151-173.

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, *15*, 119-157.

Buck, G., Tatsuoka, K., & Kostin, I. (1997). The sub-skills of reading: Rule-space analysis of a multiple-choice test of second language reading comprehension. *Language Learning*, *47*, 423-466.

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., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333-353.

DiBello, L., Stout, W., & Roussos, L. (1995). Unified cognitive/psychometric diagnostic assessment: Likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-389). Hillsdale, NJ: Lawrence Erlbaum.

Dogan, E., & Tatsuoka, K. (2007). An international comparison using a diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R. *Educational Studies in Mathematics*, *68*, 263-272.

Fischer, G. (1973). The linear logistic test model as an instrument of educational research. *Acta Psychologica*, *37*, 359-374.

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

Im, S., & Yin, Y. (2009). Diagnosing skills of statistical hypothesis testing using the rule space method. *Studies in Educational Evaluation*, *35*, 193-199.

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.

Katz, I. R., Martinez, M. E., Sheehan, K. M., & Tatsuoka, K. (1998). Extending the rule space methodology to a semantically-rich domain: Diagnostic assessment in architecture. *Journal of Educational and Behavioral Statistics*, *24*, 254-278.

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.

Rupp, A., & Templin, J. (2007). The effects of Q-matrix misspecification on parameter estimates and misclassification rates in the DINA model. *Educational and Psychological Measurement*, *68*, 78-96.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *51*, 337-350.

Tatsuoka, C., Varadi, F., & Tatsuoka, K. (1992). BUGLIB (Computer program). Trenton, NJ: Tanar Software.

Tatsuoka, K. (1984). *Analysis of errors in fraction addition and subtraction problems*. Technical report NIE-G-81-0002. Computer-based education research laboratory, University of Illinois at Urbana-Champaign.

Tatsuoka, K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics*, *10*, 55-73.

Tatsuoka, 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: Lawrence Erlbaum.

Tatsuoka, K. (1991). *Boolean algebra applied to determination of universal set of knowledge states* (Research Report RR-91-44). Princeton, NJ: Educational Testing Service.

Tatsuoka, K. (1996). Use of generalized person–fit indexes, zetas for statistical pattern classification. *Applied Measurement in Education*, *9*, 65-75.

Tatsuoka, K. (2009). *Cognitive diagnostic assessment: An introduction to the rule space method*. New York, NY: Routledge, Taylor & Francis.

Tatsuoka, K., Corter, J. E., & Tatsuoka, C. (2004). Patterns of diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. *American Educational Research Journal*, *41*, 901-926.

Tatsuoka, K., Guerrero, A., Corter, J., Tatsuoka, C., Yamada, T., Xin, T., . . . Im, S. (2006). International comparison of mathematical thinking skills in the TIMSS-R. *Japanese Journal for Research on Testing*, *2*, 3-4.

Tatsuoka, K., & Tatsuoka, M. (1987). Bug distribution and statistical pattern classification. *Psychometrika*, *52*, 193-206.

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

Yepes-Baraya, M., & Allen, N. L. (2003, April). *An attribute-based study to obtain diagnostic assessment information on the attainment of cognitive dimensions relevant in science education*. Paper presented at the 2003 American Educational Research Association meeting, Chicago, IL.