An NCME Instructional Module on

Exploring the Logic of Tatsuoka's Rule-Space Model for Test Development and Analysis

Mark J. Gierl, *Centre for Research in Applied Measurement and Evaluation, University of Alberta*
Jacqueline P. Leighton, *Yale University*
Stephen M. Hunka, *Centre for Research in Applied Measurement and Evaluation, University of Alberta*

K. Tatsuoka's rule-space model is a statistical method for classifying examinees' test item responses into a set of attribute-mastery patterns associated with different cognitive skills. A fundamental assumption in the model resides in the idea that test items may be described by specific cognitive skills called attributes, *which can include distinct procedures, skills, or processes possessed by an examinee. The rule-space model functions by collecting and ordering information about the attributes required to solve test items and then statistically classifying examinees' test item responses into a set of attribute-mastery patterns, each one associated with a unique cognitive blueprint. The logic of Tatsuoka's rule-space model, as it applies to test development and analysis, is examined in this module. Controversies and unresolved issues are also presented and discussed.*

Mark J. Gierl is an Assistant Professor of Educational Psychology in the Centre for Research in Applied Measurement and Evaluation at the University of Alberta, 6-110 Education Centre North, Faculty of Education, Edmonton, Alberta, Canada, T6G 2G5. His specializations are educational measurement and evaluation, with an emphasis on assessment and cognition, differential item functioning, test translation and adaptation, and item response theory.

Jacqueline P. Leighton is a Social Sciences and Humanities Research Council of Canada Postdoctoral Fellow at Yale University, Department of Psychology, 2 Hillhouse Avenue, P.O. Box 208205, New Haven, CT, 06520-8205. Her specializations are deductive and inductive reasoning, critical thinking, and their assessment and evaluation.

Stephen M. Hunka is University Professor and professor emeritus of Educational Psychology in the Centre for Research in Applied Measurement and Evaluation at the University of Alberta, 6-110 Education Centre North, Faculty of Education, Edmonton, Alberta, Canada, T6G 2G5. His specializations are statistics and statistical computing.

Series Information

ITEMS is a series of units designed to facilitate instruction in educational measurement. These units are published by the National Council on Measurement in Education. This module may be photocopied without permission if reproduced in its entirety and used for instructional purposes. Dr. Michael Zieky has served as editor for this module. Information regarding the development of new ITEMS modules should be addressed to: Dr. Michael Zieky, Educational Testing Service, Mail Stop 16-C, Rosedale Rd., Princeton, NJ 08541.

More than a decade ago, Richard Snow and David Lohman published their seminal chapter entitled "Implications of Cognitive Psychology for Educational Measurement" in the third edition of *Educational Measurement* (1989). Cognitive psychology has made inroads into the field of psychometrics since their chapter appeared. There is a general acceptance in the psychometric community that the psychology of test performance must be understood in order to construct, score, and validly interpret results from tests. It appears that many researchers and practitioners agree with one of the main assumptions asserted by Snow and Lohman—namely, that cognitive psychology will become fundamental to psychometric research because most educational tests are based on cognitive problem-solving tasks. A survey of the psychometric and educational literature supports this statement as books, journal articles, and edited volumes are focusing on the implications of cognitive psychology for educational measurement (e.g., Frederiksen, Glaser, Lesgold, & Shafto, 1990; Frederiksen, Mislevy, & Bejar, 1993; Mislevy, 1996; Nichols, 1994; Nichols, Chipman, & Brennan, 1995; Ronning, Glover, Conoley, Witt, 1990; also see Embretson, 1985).

Potential Benefits of Applying Cognitive Research to Educational Measurement

The cognitive analyses of educational tests may contribute to psychometric theory in a number of ways. Snow and Lohman (1989) identified four possible contributions. First, cognitive psychology may provide a new way of thinking about and understanding scores on educational tests. Snow and Lohman (1989) reasoned:

As a substantive focus for cognitive psychology then, 'ability,' the latent trait θ in EPM (educational and psychometric measurement) models, is not considered univocal, except as a convenient summary of amount correct regardless of how obtained. Rather, a score reflects a complex combination of processing skills, strategies, and knowledge components, both procedural and declarative and both controlled and automatic, some of which are variant and some invariant across persons, or tasks, or stages of practice, in any given sample of persons or tasks. In other samples of persons or situations, different combinations and different variants and invariants might come into play. Cognitive psychology's contribution is to analyze these complexes. (p. 268)

Second, cognitive analyses of educational tests may improve our understanding of the constructs represented by tests. Perhaps constructs can be more carefully described using cognitive components such as problem representation, content knowledge, initial knowledge state, and strategy selection (VanLehn, 1989) or processes and representations (Siegler & Jenkins, 1989; Siegler & Shrager, 1984). Cognitive research has the potential to provide new evidence about the construct validity of educational tests.

Third, cognitive analyses of constructs across content areas may suggest alternative measurement methods and improvements to existing methods. This point seems most applicable in the current movement toward cognitively diagnostic assessments where one of the goals in testing is to make explicit the psychological assumptions used to create the test and assign test scores. These psychological assumptions are intended to describe the cognitive processes and knowledge structures used by examinees to solve items on the test (Frederiksen, Glaser, Lesgold, & Shafto, 1990; Nichols, 1994; Nichols, Chipman, & Brennan, 1995). However, the success of this movement depends, in part, on the accuracy of modeling examinees' processes and representations in order to produce valid and useful remedial information about test performance. Cognitive research can guide these attempts at modeling complex problem-solving behaviors.

Fourth, cognitive analyses may lead to more unified and predictive theories of aptitude, learning, instruction, and achievement in education if tests can be used to measure the cognitive components required to solve different academic tasks. Cognitively diagnostic tests, for example, could be used to evaluate and expand educational theory by assessing some of the assumptions in these theories.

In short, cognitive psychology has the potential to advance psychometric theory. Cognitive analyses will allow researchers to experiment with the internal characteristics of the test, evaluate the assumptions of existing psychometric models, build new psychometric models of test performance, and explain, substantively, the psychology by which tests are constructed, scored, and interpreted. With the advance of cognitive theory into psychometrics, educational testing may never be the same.

In an effort to represent the cognitive skills required to solve items on educational and psychological tests, Kikumi Tatsuoka and her associates developed a psychometric model called rule space (e.g., K. Tatsuoka, 1983, 1993, 1995; K. Tatsuoka & M. Tatsuoka, 1987). Although different cognitive diagnosis models have been developed (see, e.g., the review by Roussos, 1994; Nichols, Chipman, & Brennan, 1995; or van der Linden & Hambleton, 1997, Section 3), Tatsuoka's work on the rule-space model represents one of the first comprehensive and practical programs of research on psychometric models in the area of cognitive diagnosis. Moreover, Tatsuoka's approach accommodates stochastic variation and distributional theory. It is used to assess whether the cognitive skills or attributes required to solve items have been mastered by examinees, and it can be used to diagnose their misconceptions. Tatsuoka's rule-space model posits that exam questions can be described by specific cognitive skills. These skills are called *attributes*, and they can include different procedures, skills, or processes that an examinee must possess to solve a test item. Tatsuoka's model serves as a statistical method for classifying examinees' test item responses into a set of attribute-mastery patterns associated with different cognitive skills. The model provides a measure of cognitive proficiency. Tatsuoka (1995) provided this rationale for her approach:

If new measurement models are to measure complex and dynamic cognitive processes and to be linked to instruction, then understanding knowledge structures in highly specific detail provides a rational basis for proposing and evaluating potential improvements in the measurement of general proficiency. Without this understanding, improvement remains largely a trial-and-error process. (p. 328)

The purpose of this instructional module is to highlight key features in Tatsuoka's rule-space model[1] so the reader understands the logic behind this approach to cognitively diagnostic assessment. We also provide a tutorial on Tatsuoka's model developed with Mathematica 4.0 (Wolfram, 1996). Readers who would like to learn more about this approach can download the tutorial and, if they have Mathematica, the functions used to execute the computations required of the rule-space model with their own data. The functions provided in this Mathematica notebook are intended to provide users with some tools for further exploring the psychometric characteristics of the model.

Tatsuoka's Approach to Cognitive Diagnosis

An Overview

Broadly speaking, the rule-space model was developed to address two distinct problems first identified in engineering and the physical sciences and later applied to cognitive assessment—namely, the identification of feature variables and statistical pattern classification. The focus of this module is on the first problem, the identification of feature variables, as outlined by Tatsuoka and her associates. Tatsuoka calls the feature variables *attributes* in her approach. The second problem, statistical pattern classification, is beyond the scope of this paper. A discussion of statistical pattern classification using a two-dimensional Cartesian coordinate system called the *rule space*, characterized by the theta (i.e., Θ or ability) and zeta (i.e., ζ or response unusualness) axes, can be found in different sources including Tatsuoka (1984, 1995, 1996).

To begin, a general example is provided. In Figure 1a, three attributes make up the hierarchy. The hierarchy explicitly defines the *logical and/or psychological ordering* between the attributes required to solve a test item (i.e., the hierarchy specifies the dependent relations between the attributes). Attribute A1 may be considered hypothetical in the sense that it represents all the initial skills and competencies required of the examinee that are prerequisite to Attributes A2 and A3 or, alternatively, A1 may be considered a specific attribute. Furthermore, to say that Attribute A1 is prerequisite to Attributes A2 and A3 implies that an examinee is not expected to possess Attributes A2 and A3 unless Attribute A1 is present. Similarly, Attribute A2 is prerequisite to A3.

The ordering of the attributes may be derived from logical (e.g., addition is logically followed by multiplication) and/or psychological considerations (e.g., developmental sequences suggested by Piaget such as preoperational, concrete operational, and formal operational). In Figure 1b, three attributes again make up the hierarchy, and Attribute A1 is prerequisite to Attributes A2 and A3. However, this time Attribute A2 is not prerequisite to A3. Given the constraints implied by the hierarchical relationship among the attributes, the ideal item response patterns of the examinees to test items become a function of both the attributes possessed by the examinees and the attributes required to answer the items.

Identifying Attributes

Tatsuoka's rule-space model postulates that performance on test items depends on a set of specific skills or competencies called attributes, which the examinee must possess in order to answer the items correctly. The importance of correctly identifying the

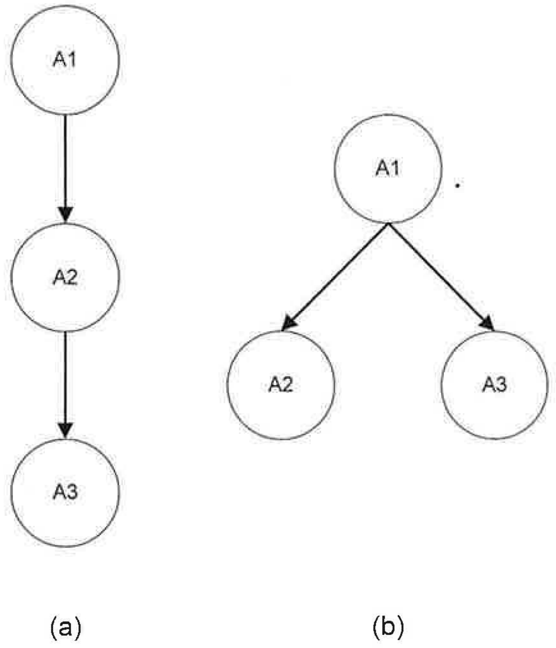


FIGURE 1. *Two simple relationships between the attributes A1, A2, and A3*

attributes and the hierarchy in Tatsuoka's approach cannot be overstated—the first step in making inferences with the rule-space model depends on identifying the cognitive skills required to solve items on a test and to order the attributes in a hierarchy. The attributes and the hierarchy serve as the most important input variables to the model because they provide the basis for interpreting the results in this approach to psychometric modeling. The attributes on a test can be identified in many ways. Ideally, however, the attributes are identified by specifying the cognitive skills required to solve a set or class of problems in a specific content area before the test is constructed. For example, the cognitive attributes required to solve mixed fractions may include addition, subtraction, multiplication, division of integers, conversion of mixed numbers into improper fractions, and finding a common denominator. After the attributes are identified, test items can be created to measure these attributes. To test the attributes required to solve mixed fractions, students could be presented with this problem, taken from Tatsuoka (1995):

$$2\frac{1}{2}+4\frac{2}{4}.$$

To solve the problem, an examinee might use the following four steps.

Step 1: Convert the mixed number into an improper fraction

$$\frac{5}{2}+\frac{18}{4};$$

Step 2: Find a common denominator using one of the fractions

$$\frac{5}{2}\times\frac{2}{2}=\frac{10}{4};$$

Step 3: Add the two fractions

$$\frac{10}{4}+\frac{18}{4}=\frac{28}{4};$$

Step 4: Reduce the fraction to the lowest common denominator

$$\frac{28}{4}=7.$$

More succinctly, the attributes used by an examinee to solve the fraction problem just presented include: addition of integers (Attribute 1 or A1), multiplication of integers (A2), conversion of mixed numbers into improper fractions (A3), and division of integers (A4). These four attributes can be ordered into a hierarchy based on their logical and/or psychological properties. The attribute hierarchy for the fraction problem is presented in Figure 2. Attributes A1 and A2 are prerequisites to Attribute A3 because the attribute *convert mixed numbers into improper fractions* requires that the examinee be able to add integers and multiply integers. Attribute A4 (division of integers) is only tenable, given the specific attribute hierarchy, if the examinee possesses Attributes A1, A2, and A3. Consequently, we can infer that an examinee who correctly solves this problem possesses Attributes A1 through A4 *under the assumption that the hierarchy is true*. Moreover, if a set of test items requires Attributes A1 through A4 and the examinee uses the same approach to solve the items, the expectation is that the examinee will get the items correct ignoring random error that might occur.

The validity of the measures produced by the rule-space model depends on the skills of the curriculum specialists and test developers to correctly identify the attributes required to solve problems in a specific content area, to specify the relationship among these attributes, and to create test items that accurately measure these attributes. Understandably, however, identifying the attributes and then specifying their logical structure in the hierarchy is a challenging task. A great deal of effort is required in the test development process to achieve this goal. To some extent this challenge raises the specter of Bloom's taxonomy (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956) if *objectives* is substituted for *attributes*, although Bloom's taxonomy does not specify the ordering of objectives as precisely as the rule-space model specifies the ordering of attributes. Hence, one advantage of Tatsuoka's approach is that it forces researchers and practitioners to go beyond identifying general, typically imprecise, cognitive skills

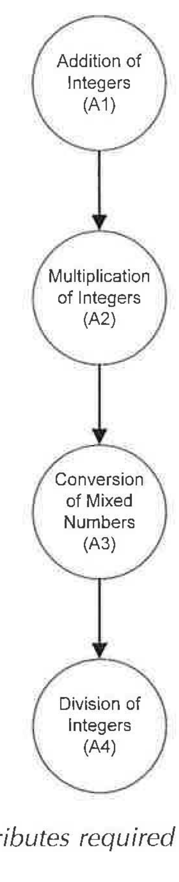


FIGURE 2. *The attributes required to solve a mixed fractions problem*

(the kind often seen in test blueprints) to carefully identifying and ordering the cognitive skills required to solve problems in a specific content area (Gierl, 1997a). This information, in turn, is used to evaluate students' cognitive skills and to provide specific information about their cognitive problem-solving strengths and weaknesses that may prove useful for instruction and remediation because of the high level of detail and specificity produced (see Birenbaum, Kelly, & Tatsuoka, 1993, for an example in mathematics).

"Hitting Paydirt:" Defining Relationships Among Attributes

In the rule-space model, the relationship among attributes is specified by a binary *adjacency matrix* (A) of order $k \times k$, where k is the number of attributes. Moreover, it is important to note that the adjacency matrix expresses only the direct relationship between attributes such that the ij element represents the presence or absence of a direct connection of one attribute to another. Consider, for example, the two attribute hierarchies illustrated in Figure 3, along with their respective adjacency matrices in Table 1. These two hierarchies will serve as the basis for the two examples in this module.

In Figure 3a, Attribute A1 is prerequisite to Attribute A2 and A4. This hierarchical relationship is expressed in the first row of the A_a matrix in Table 1 by the positions of a 1 in columns 2 and 4. The positions of 0 in row 1 indicate that Attribute A1 is neither directly connected to itself nor to Attribute A3. Figure 3a also shows that Attribute A2 is prerequisite to A3. The lone 1 in row 2 of the A_a matrix in Table 1 indicates that Attribute A2 is directly connected to A3. Lastly, rows 3 and 4 of the A_a matrix indicate that Attributes A3 and A4 are not direct prerequisites to any attributes because there are only 0s occupying these two rows. The adjacency matrix, A_b, in Table 1 is interpreted in the same fashion. Two additional points about the A matrix are worth noting. First, the A matrix is always square and of an order equal to the number of attributes. Second, the A matrix may be configured in upper triangular form with zeros in all diagonal positions, if the matrix reflects the sequence of attributes in the hierarchy. The upper triangular form implies that the sequence of attributes is unidirectional, for example, Attribute A4 cannot connect directly back to A1.

Although the A matrix expresses the direct relationship between attributes, it does not express the indirect relationships among attributes. To specify the direct and indirect relationships among attributes, Tatsuoka uses a *reachability matrix* (R) of order $k \times k$, where k is the number of attributes. A simple way of thinking about the reachability matrix is to consider its application in communications. For example, suppose you had telephone links between Edmonton and Toronto, Toronto and Chicago, and Chicago and New York. This kind of information is exactly that held by the adjacency matrix. The R matrix, which is calculated from the adjacency matrix, can be used to determine whether it is possible to reach New York from Edmonton. Of course, in this simple example, one can see intuitively that it is possible, but in instances where many routes exist, some of which may be inoperative, the solution is not always obvious. To obtain the R matrix from the A matrix, Boolean addition and multiplication operations are performed on the adjacency matrix (i.e., $R = (A + I)^n$; where n is the integer required to reach invariance, $n = 1, 2, \ldots m$, and I is the identity matrix), such that when the result becomes invariant the reachability matrix has been obtained. For example, if one were to try $n = 1$, 2, and 3 and to observe that $n = 2$ yielded the same result as $n = 3$, then the R matrix would have been defined. The R matrices derived from adjacency matrices, A_a and A_b, are shown in Table 2.

The R matrix is read in a similar fashion to the A matrix. For instance, the first row of the reachability matrix R_a indicates Attribute A1 is related (i.e., "reaches") to all other attributes because of the positions of a 1 in columns 1, 2, 3, and 4. Row 2 of this matrix indicates that A2 reaches A3 but not A4. Matrix R_b is

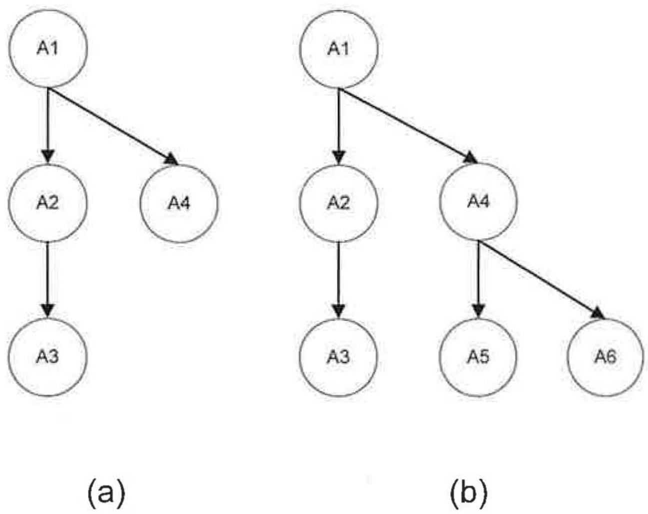


FIGURE 3. *Two different relationships between two sets of attributes*

Table 1
Adjacency Matrices, A_a and A_b, for Attribute Hierarchy Illustrated in Figure 3a and 3b, Respectively

A_a [a]	A_b [b]
0101	010100
0010	001000
0000	000000
0000	000011
	000000
	000000

Note. The adjacency matrix is of order $k \times k$, where k is the number of attributes.
[a] Matrix has four attributes corresponding to Figure 3a
[b] Matrix has six attributes corresponding to Figure 3b

Table 2
Reachability Matrices, R_a and R_b, for Adjacency Matrices, A_a and A_b, Respectively

R_a	R_b
1111	111111
0110	011000
0010	001000
0001	000111
	000010
	000001

Note. The reachability matrix is of order $k \times k$, where k is the number of attributes.

Table 3
Incidence Matrix Q_a Derived from Reachability Matrix R_a

101010101010101
011001100110011
000111100001111
000000011111111

Note. The incidence matrix is of order $k \times 2^k - 1$, where k is the number of attributes. This matrix is the unreduced Q matrix (i.e., it is not constrained by the logical characteristics of the hierarchy).

interpreted in the same manner. Note also that, unlike the A matrix, the R matrix has 1s occupying its diagonal positions, suggesting that each attribute can reach itself. The R matrix is ultimately used to select a subset of items, from a potential pool of items, which reflects the attribute hierarchy.

Potential Pool of Item Types

The potential pool of item types is considered to be those items representing all combinations of attributes when the attributes are considered independent of each other (i.e., when the corresponding reachability matrix is an identity matrix). The size of the potential pool is $2^k - 1$, where k is the number of attributes. Thus, for even a small k, say, 10, the pool of items is quite large (e.g., $2^{10} - 1 = 1023$ items). When the attributes are considered related in some way, the potential pool of items can be reduced. Tatsuoka refers to the potential pool of item types as the *incidence matrix* (Q) of order $k \times i$, where k is the number of attributes and i is the number of potential items. In the Q matrix, each item is described by the attributes required by the examinee to obtain a correct answer. The Q matrices for the attribute hierarchies depicted in Figure 3a ($k = 4$) and 3b ($k = 6$) are shown in Tables 3 and 4, respectively.

In Table 3, the incidence matrix, Q_a, is of order 4 by 15 (i.e., attributes by items), and each column of the matrix represents one item. For example, column 1 of this matrix represents Item 1, and it specifies that only Attribute A1 is required to correctly answer this item. Conversely, Item 15 (column 15) requires all four attributes for a correct response. Notice also that Item 8 requires only Attribute A4. If the constraints of the attribute hierarchy shown in Figure 3a are imposed, Item 8 does not conform or fit to the specified hierarchy because Attribute A4 is contingent on the existence of Attribute A1 and Item 8 does not have a 1 in row 1 to indicate that A1 is required. In Table 4, the matrix, Q_b, is of order 6×63, and it is interpreted in the same way as Q_a.

Reducing the Incidence Matrix

The incidence matrix Q may be reduced to form the *reduced Q matrix* (Q_r) by imposing the constraints of the attribute hierarchy as defined in the R matrix. The reduced Q matrix represents the items from the potential pool that fit the constraints of the specified attribute hierarchy. For instance, the number of items reflected in the incidence matrix Q_a shown in Table 3 can be reduced. Notice that column 1 of Q_a (i.e., Item 1) is represented in binary as 1000, indicating that only Attribute A1 is required to correctly answer this item, whereas Item 2 is represented as 0100, indicating that only Attribute A2 is required. Because A1 is prerequisite to A2 according to the hierarchy, and assuming that there are no slips or errors[2], Item 2 should be designated as requiring Attributes A1 and A2. As a result, Item 2 should be represented as 1100, indicating that both A1 and A2 are required for a correct answer. The binary representation 1100, however, is already used in Q_a to portray the attribute requirements of Item 3. Hence, Item 2 is taken to be equivalent to Item 3 in attribute requirements. The equivalency of Items 2 and 3 illustrates one reduction in the number of potential items represented in Q_a.

Another approach to reducing the potential set of items according to the attribute hierarchy is to note that if the attribute model is true, certain items do not conform to the specified hierarchy. For instance, constructing Item 2 according to Q_a would only require Attribute A2, but if Attribute A1 is prerequisite to A2, an item requiring only A2 should not be used.

The reduced Q matrix is formed by determining which columns of the incidence matrix Q are entailed in each column of the reachability matrix R, using Boolean inclusion. This procedure is more formally specified by Tatsuoka (1991). The reduced Q matrix, $Q_{r;a}$, derived from R_a, is shown in Table 5. As shown, only six items can be generated from the attribute hierarchy illustrated in Figure 3a.

The reduced Q matrix, $Q_{r;b}$, derived from R_b is shown in Table 6 and, as can be seen, from a potential pool of 63 items, only 15 items can be generated after imposing the constraints of the attribute hierarchy illustrated in Figure 3b.

Calculating the Ideal Item Response Patterns

Given a hierarchy of attributes postulated to exist for an examinee and required to answer a set of items, the ideal item-score pattern (i.e., vector) for the examinee can be calculated provided the examinee does not make any "slips." For example, consider the attribute hierarchy shown in Figure 3a. Imagine an examinee who has Attributes A1, A2, and A4, a pattern of attributes expressed by the vector 1101. Knowing the examinee's pattern of attributes, we can subsequently obtain the examinee's ideal item response pattern as 110110, which indicates that he or she is able to answer items 1, 2, 4, and 5 correctly. The ideal item response

Table 4
Incidence Matrix Q_b Derived from Reachability Matrix R_b

101010101010101010101010101010101010101010101010101010101010101
011001100110011001100110011001100110011001100110011001100110011
000111100001111000011110000111100001111000011110000111100001111
000000011111111000000001111111100000000111111110000000011111111
000000000000000111111111111111100000000000000001111111111111111
000000000000000000000000000000011111111111111111111111111111111

Note. The incidence matrix is of order $k \times 2^k - 1$, where k is the number of attributes. This matrix is the unreduced Q matrix (i.e., it is not constrained by the logical characteristics of the hierarchy).

Table 5

Reduced Incidence Matrix $Q_{r;a}$ Derived From Incidence Matrix Q_a

111111
011011
001001
000111

Note. The reduced incidence matrix is of order $k \times i$, where i is the number of possible items. This matrix is now *constrained* by the logical characteristics of the hierarchy.

pattern further indicates that the examinee is unable to answer Items 3 and 6, which are precisely those items that necessitate Attribute A3—the attribute the examinee lacks. It should be noted that when the Q_r matrix is transposed (i.e., the transposition of a matrix can be thought of as flipping the matrix by 180° over the main diagonal so that the rows and columns exchange positions), the resulting matrix, E, can be read as holding all tenable examinee attribute combinations consistent with the constraints of the A matrix. For example, assuming that all plausible attribute combinations associated with the hierarchy shown in Figure 3a are represented by six different examinees, the attributes of these six examinees would be described by the matrix E_a shown in Table 7.

In the matrix E_a, the rows represent six different examinees, and the columns represent the four attributes. If the model is correct, no other combination of attributes should be reflected by these examinees. Although we may observe many examinees with a similar attribute pattern, the specific nature of this distribution would depend on the sample of examinees (e.g., we would likely observe many examinees with the attribute pattern 1111 in an academically gifted sample).

Calculating the ideal item response vectors for the six examinees produces the six item response vectors shown in Table 8.

The total score on the test created according to this model for the six examinees would be 1, 2, 3, 2, 4, and 6, respectively. Notice, however, that by observing the total score we are not consistently able to uniquely determine which attributes an examinee possesses. For instance, a score of 2 can be obtained by an examinee who has Attributes A1 and A2 (i.e., 1100) or Attributes A1 and A4 (i.e., 1001). This is a limitation of any measurement procedure when a total score is based on the sum of the item scores. In the case of the attribute hierarchy shown in Figure 3b, the examinee attributes, ideal item response vectors, and total scores are shown in Table 9.

Keeping in mind the attribute hierarchy, row 1 of Table 9 should be interpreted as follows: An examinee who only has Attribute A1 (i.e., 100000) is expected to answer only the first item correctly, out of a 15 item test; that is, he or she should produce the ideal item response vector, 100000000000000. Alternatively, if Attribute A3 is not present but all others are (i.e., 110111), then the response vector is expected to be 110110110110110. Hence, the ideal item response vector an examinee produces is dependent on the attributes the examinee possesses. If an examinee lacks an attribute that is required by a subset of items for a correct response, then those items will be missed. Furthermore, like the previous example, notice that the examinee's total score does not indicate which attributes are present. For example, a score of 2 may be obtained by having attribute patterns 110000 or 100100. If the attribute hierarchy is true, the only scores that should be observed are 1, 2, 3, 4, 5, 6, 9, 10, and 15. Random responding can produce as many as 32768 different response vectors associated with total scores ranging from 0 to 15. When many other score patterns, outside of the ideal score patterns, are observed, this outcome suggests that (a) the attributes were not accurately identified, (b) the attribute hierarchy specified is inappropriate, (c) the items constructed or chosen do not fit the model, (d) the test was inappropriate for the student sample, and/or (e) random slips were made.

Table 6

Reduced Incidence Matrix $Q_{r;b}$ Derived From Incidence Matrix Q_b

111111111111111
011011011011011
001001001001001
000111111111111
000000111000111
000000000111111

Note. The reduced incidence matrix is of order $k \times i$, where i is the number of possible items. This matrix is now *constrained* by the logical characteristics of the hierarchy.

Table 7

Ideal Attribute Matrix, E_a, in a Hypothetical Pool of Six Examinees and Corresponding to Reduced Incidence Matrix $Q_{r;a}$

1000
1100
1110
1001
1101
1111

Note. Matrix E_a is the transposition of matrix $Q_{r;a}$.

Table 8

Ideal Item Response Vectors, Total Scores, and Examinee Attributes for the Hypothetical Pool of Six Examinees Originally Illustrated in Matrix E_a

Examinee	Ideal item response vectors	Total scores	Examinee attributes[a]
1	100000	1	1000
2	110000	2	1100
3	111000	3	1110
4	100100	2	1001
5	110110	4	1101
6	111111	6	1111

[a] This column illustrates the information provided by matrix E_a.

Table 9
Ideal Item Response Vectors, Total Scores, and Examinee Attributes for the Hypothetical Pool of 15 Examinees

Examinee	Ideal item response vectors	Total scores	Examinee attributes[a]
1	100000000000000	1	100000
2	110000000000000	2	110000
3	111000000000000	3	111000
4	100100000000000	2	100100
5	110110000000000	4	110100
6	111111000000000	6	111100
7	100100100000000	3	100110
8	110110110000000	6	110110
9	111111111000000	9	111110
10	100100000100000	3	100101
11	110110000110000	6	110101
12	111111000111000	9	111101
13	100100100100100	5	100111
14	110110110110110	10	110111
15	111111111111111	15	111111

[a] This column illustrates the E_b matrix.

Cognitively Diagnostic Assessments: Emerging Issues and Controversies

Kikumi Tatsuoka and her associates have conducted a great deal of research on the rule-space model. Some of this research is summarized in our module, especially as it relates to the logical aspects of the model. However, a considerable amount of work remains. In our concluding section, we identify three key issues, controversies and, subsequently, research opportunities, in the area of cognitively diagnostic assessment using Tatsuoka's approach.

Developing Cognitively Diagnostic Assessments: A 4-Step Process

Currently, there are no guidelines for developing a cognitively diagnostic assessment when the rule-space model is used to analyze the results. To this end, we believe the development of a cognitively diagnostic assessment based on the rule-space model is likely to produce the most meaningful results when the following sequence of steps is followed. First, the attributes must be identified and their hierarchical relationships specified. This step requires that test developers have a clear understanding of the cognitive processes measured within a specific group of examinees. Careful thought must be given to the definition of the attributes and their hierarchical relationships in order for the hierarchy to be represented as a tree structure (or sometimes called network structure; see Tatsuoka, 1995). Furthermore, the hierarchical relationships of the attributes must not be confused with the order of their invocation by examinees responding to test items. Although there is an interaction between the availability of attributes possessed by an examinee and the order of their invocation within each specific test item (i.e., the strategy of invoking the attributes), it is important not to confuse these two concepts because each represents a distinct level of cognitive analysis. Attributes are considered to represent building blocks to strategies and, consequently, need not always directly correspond with a strategy (this point is described in the next section). It is probably more fruitful, at least until more experience is gained with the rule-space model, to start with attributes required of examinees in the elementary grades rather than, for example, candidates in programs leading to professional certification. The clarity of the attribute hierarchy is probably more easily achieved when the attributes are clearly of a logical or psychological nature. We recommend that the attribute hierarchy have a single starting node to represent all the attributes prerequisite to the attributes contained within the remaining hierarchy. Such a starting node also serves the purpose of keeping the test developer aware of all the starting attributes that are *assumed* to be available to all examinees.

Second, from the attribute hierarchy generated during Step 1, the adjacency matrix is constructed from which, by logical operations, the reachability matrix and reduced incidence matrix can be obtained.

Third, construct the test items. During this process, it is probable that the prerequisite attributes represented by the starting node of the attribute hierarchy will be modified. For example, even though a test item may require the attribute of adding single-digit numbers, if the item requires the examinee to read a context for the problem, the prerequisites contained in the starting node will have to include a reading attribute in some form.

We note, parenthetically, that it is not recommended that the attribute hierarchy model be obtained from a set of items taken from an existing test by constructing a reduced incidence matrix through an analysis of the attributes required of each item. Test items constructed for the purpose of achievement testing are probably not ideal items for cognitively diagnostic assessments. Furthermore, the attributes associated with items selected from an existing test may not be adequately represented by an attribute hierarchy; that is, there are no assurances that one can work back from the reduced incidence matrix to the reachability matrix, and then work back to the adjacency matrix which is required to construct the tree structure representing the attribute hierarchy. For some attribute hierarchies, it is not possible to recover the appropriate adjacency matrix from the reachability matrix because the reachability matrix represents *both* direct and indirect connections. We know of no method for reliably disentangling the direct and indirect connections of the reachability matrix in order to obtain the appropriate adjacency matrix required to depict the attribute hierarchy (see sec. 5 in the Mathematica tutorial associated with this module under the heading "Ambiguity in Retrieving Attribute Hierarchy from the Reachability Matrix" for a detailed example using data from Tatsuoka, 1995). Also, even if an accurate attribute hierarchy can be obtained when starting from an existing set of test items, the reduced incidence matrix derived from the attribute hierarchy is likely to call for a set of test items having little similarity to those of the existing test items (i.e., the items will probe different combinations of attributes). This recommendation represents an important departure from Tatsuoka's approach because she starts, at times, with an existing set of test items before the attributes are identified.

Fourth, generate the ideal item response vectors once the reduced incidence matrix is formed. Assuming the attribute hierarchy is true and using the matrix of ideal item response vectors, a probability estimate can be made of a correct and incorrect answer for each item. A major unresolved issue is how best to determine whether the attribute hierarchy is true, or partially true. Although we know what the ideal item response vectors will be under the condition that the hierarchy is true, we do not know what evidence should be used from all the observed response vectors to warrant stating the hierarchy is appropriate or inappropriate in a given analysis.

Defining and Identifying Attributes

Attributes serve as the most important input variables to the rule-space model because they provide the basis for making inferences about examinees' cognitive skills. Yet despite their importance, attributes are inconsistently described in the rule-space literature. Various definitions for attributes have been proposed, various techniques for identifying attributes have been used, and various specialists have been consulted. As a result, it is difficult to characterize a test attribute. For some researchers, an attribute is general, and it can be viewed as the cognitive requirements of a task (see, for example, Tatsuoka, Birenbaum, Lewis, & Sheehan, 1992). For other researchers, an attribute is more specific, because it can be described as the combination of cognitively relevant subcomponents used to solve a problem (e.g., Tatsuoka, 1993), the specification of elementary cognitive skills needed for mastery of a domain (e.g., Sheehan, Tatsuoka, & Lewis, 1992), or the description of a set of procedures or operations that one can use in solving a problem in some well-defined procedural domain (e.g., M. Tatsuoka & K. Tatsuoka, 1989). Attributes are also characterized using a combination of general and specific descriptors. For example, Birenbaum, Kelly, and Tatsuoka (1993) stated

> An attribute of a task is a description of the processes, skills, or knowledge a student would be required to possess in order to successfully complete the target task. ... They may include, but are not limited to, a student's ability to perform some procedures. Attributes may also include a student's use of heuristics, or adoption of a strategy. (p. 442)

These examples demonstrate that little consensus exists on how to describe a test attribute.

There is also little consensus on how to identify the attributes from a test. Some researchers are guided by conceptual frames (i.e., models of problem solving are used to identify the attributes of a test). For example, Tatsuoka, Birenbaum, Lewis, and Sheehan (1992) used a mapping sentence, a procedure used in facet theory, to specify the content and process attributes from student protocol data on the Scholastic Assessment Test (SAT). Their mapping sentence contained 13 facets, and each facet contained elements (e.g., the facet-labeled processing contained seven elements, including reasoning, analytic thinking, and reading comprehension). The primary elements from the mapping sentence were selected as attributes for the test. Katz, Martinez, Sheehan, and Tatsuoka (1998) used a model of problem solving to evaluate a 22-item architecture test. The model contained three steps: Constructing the initial representation of an item, forming goals and performing actions based on those goals, and determining whether the goals have been attempted and satisfied. Item attributes were derived from the cognitive processes specified in each step of the model. Sheehan, Tatsuoka, and Lewis (1992) advocated the use of conceptual models more generally:

> Each item in the [examination] pool is classified [based on] a subset of skills required for successful completion. This classification must be performed by someone who is familiar both with the items and with the cognitive models proposed for solving the items. (p. 4)

Finally, Tatsuoka (1990) offered a conceptual frame that depended on the judgment of specialists. She stated, "The initial task analyses could be carried out by several experts or master teachers. If two or more experts use different methods to solve a set of problems, then they could get entirely different attribute by item matrices" (p. 461). In this case, several different incidence matrices are possible, and each could be tested with the rule-space model. In these examples, attributes are identified using different procedures and data sources (e.g., content specialist or student protocols), demonstrating that little consensus exists on how to identify the attributes on a test.

In short, attributes in the rule-space literature have a limited conceptual foundation. They have been identified using different models, techniques, and data. Consequently, the attributes tend to differ qualitatively from one study to the next, and there is no coherent knowledge base from which to understand the psychology of test performance despite their importance for interpreting the output produced by the rule-space model. Attributes are clearly related to what examinees think about when they solve items on a test, and they represent cognitive skills in some form. However, the conceptual foundation for attributes in the rule-space literature must be refined. To clarify what is meant by *attribute*, we propose the following new working definition (Gierl, 1997b; Leighton, Gierl, & Hunka, 1999):

> An attribute is a description of the procedural or declarative knowledge needed to perform a task in a specific domain. Although an attribute is not a strategy, attributes do provide the building blocks for strategies. Furthermore, the set of attributes organized into a strategy serves a momentary problem-solving role, but does not necessarily remain grouped as a strategy. Attributes are dynamic entities. They evolve with a student's increasing competency so that a set of attributes at time 1 may no longer function as useful descriptions of behavior at time 2. Finally, the time periods mentioned are developmentally and/or instructionally dependent, meaning that a student progresses from time 1 to time 2 in response to developmental and/or instructional factors. The attributes for a test can be identified by using different methods (e.g., expert opinion, task analysis, written responses from students). However, verbal think-aloud protocols should be included among the methods used to validate the attribute descriptions using both examinees and test items that are comparable to their target populations.

Given our definition, the identification of attributes should involve characterizing attributes and their interrelationships *before* test construction because, if test items are selected before the attribute hierarchy is specified, these items may not fit the chosen hierarchy. The rule-space model should be applied to domains where attributes can be clearly defined and identified. Finally, we recommend that the attribute hierarchy include a starting node, which includes all prerequisite attributes in the model.

Availability of Computer Software

Rule-space analyses are conducted with the computer program PMAIN (Varadi & Tatsuoka, 1989, 1992). However at the time of publication of this module, the program was for research purposes only, and not publicly available. This outcome limits the use of the rule-space model. In an attempt to overcome this limitation, we provide a tutorial based on Tatsuoka's model developed with Mathematica 4.0 (Wolfram, 1996). The functions presented in the Mathematica tutorial were developed from the concepts and procedures described in the research literature. Readers who would like to learn more about this approach can download the tutorial and view our results using the MathReader, a program that allows one to read a Mathematica notebook. If readers have the full version of Mathematica, they can access the functions used to execute the computations required of the rule-space model using their own data. The tutorial and the functions provided in this Mathematica notebook are intended to provide users with some tools for exploring the logical characteristics of the rule-space model. Comments and suggestions on the tutorial are welcome. The tutorial can be downloaded from http://www.education.ualberta.ca/educ/psych/crame/, or it can be obtained from the first or third author by request.

Summary

Cognitive psychology has the potential to advance psychometric theory because educational tests are based on cognitive problem-solving tasks. With the development of the rule-space model, Tatsuoka is one of the first psychometrians to embrace this assumption. This approach to cognitive assessment demonstrates how principles and practices in cognitive psychology and educational assessment can be combined in the spirit advocated by Sam Messick (1989):

> Almost any kind of information about a test can contribute to its construct validity, but the contribution becomes stronger if the degree of fit of the information with the theoretical rationale underlying score interpretation is explicitly evaluated. ... Possibly most illuminating of all are direct probes and modeling of the

processes underlying test responses, an approach becoming both more accessible and more powerful with continuing developments in cognitive psychology. (p. 17)

The rule-space model can be used to assess whether the cognitive skills or attributes required to solve items have been mastered. An attribute, as it is generally described, includes knowledge, procedures, skills, or processes an examinee must possess to solve a test item. The attributes of an examinee are not observable, and they must be inferred from their response patterns. Tatsuoka's rule-space model is a statistical method for classifying examinees' test item responses into a set of attribute-mastery patterns associated with different cognitive skills.

The logical aspects of Tatsuoka's model, under the assumption that the test developer's specifications of the attribute hierarchy are true, are used to establish a unique set of ideal item response vectors. These vectors, or a subset of them, are the only vectors that can be observed for a group of examinees. The total score associated with the response vector indicates an ability level for the examinee. To begin any cognitively diagnostic assessment using the rule-space model, the researcher must identify and specify the qualitative nature of the attributes. Their relationships are precisely defined in the adjacency matrix, which defines the direct relationships between the attributes. The reachability matrix is derived from the adjacency matrix, and it expresses both the direct and indirect relationships among the attributes. The potential pool of items to assess these attributes can then be generated as an incidence matrix, which describes each item by the attributes being assessed. The incidence matrix can be reduced or constrained by the relationships expressed in the reachability matrix so that the reduced incidence matrix contains the minimum number of items and attributes that each item can measure. Using the adjacency matrix, a plot of the attribute hierarchy can be made that specifies each unique-ordered combination of attributes in the cognitive model. Using the reduced incidence matrix, and its transpose, a tenable list of ideal item response vectors and the total scores can be calculated.

Despite the relatively well-defined logical aspects of this approach, much work remains because many issues and controversies are not resolved. For example, controversy exists about how to develop and analyze cognitively diagnostic assessments, how to conceptualize and describe attributes, and how to use the rule-space model in applied settings. Ultimately, the success or failure of applying cognitively rich psychometric models to understand student problem-solving will depend on the accuracy of the cognitive assumptions made.

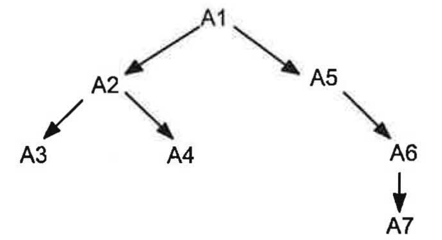
Self-Test

1. Children have been taught to add two proper fractions that do not have a common denominator. (For a proper fraction the numerator is less than or equal to the denominator.) The instructional procedure involved the following understandings and procedures:
   (a) $n/n = 1$, for any value of n not equal to zero or infinity
   (b) (n/n) times a proper fraction a/b equals (na/nb)
   (c) two fractions having a common denominator can be added by adding the numerators and maintaining the common denominator
   (d) simplification of a fraction can be made by obtaining factors for the numerator and denominator and then canceling

   For example, for the addition of $1/6$ and $1/3$, we could expect the student to carryout the following sequence of operations:
   1) $(3/3) \times (1/6) = 3/18$
   2) $(6/6) \times (1/3) = 6/18$
   3) $(3/18) + (6/18) = 9/18$
   4) $(9/18) = (3 \times 3)/(2 \times 3 \times 3) = 1/2$.

   The following attributes have been defined for these operations:
   1) prerequisite skills
   2) multiplication of two integers
   3) factor an integer
   4) multiply two fractions
   5) $n/n = 1$
   6) add two fractions having common denominators
   7) cancel common numbers in the numerator and denominator and replace with 1.

   The following hierarchy representing the conceptual complexity of the attributes is shown below.



1a. Create the Adjacency (A) matrix for this hierarchy.

Answer:

$$\begin{bmatrix} 0&1&0&0&1&0&0 \\ 0&0&1&1&0&0&0 \\ 0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&0 \end{bmatrix}$$

1b. Create the Reachability (R) matrix.

Answer:

$$\begin{bmatrix} 1&1&1&1&1&1&1 \\ 0&1&1&1&0&0&0 \\ 0&0&1&0&0&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&0&0&1&1&1 \\ 0&0&0&0&0&1&1 \\ 0&0&0&0&0&0&1 \end{bmatrix}$$

1c. Which matrix, the Adjacency or Reachability matrix, defines the direct connections among the attributes?

Answer: Adjacency matrix

1d. Which matrix, the Adjacency or Reachability matrix, defines the direct and indirect connections among the attributes?

Answer: Reachability matrix

2. The reduced incidence (Q_r) matrix of order (attributes by items) for the hierarchy of Question 1 is given below.

$$\begin{bmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 0&1&1&1&1&0&1&1&1&1&0&1&1&1&1&0&1&1&1&1 \\ 0&0&1&0&1&0&0&1&0&1&0&0&1&0&1&0&0&1&0&1 \\ 0&0&0&1&1&0&0&0&1&1&0&0&0&1&1&0&0&0&1&1 \\ 0&0&0&0&0&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 0&0&0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&1&1 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&1 \end{bmatrix}$$

2a. Why is there no single item probing simultaneously the Attributes A1, A3, and A4?

Answer: Attribute A2 (multiplication of two integers) is a prerequisite to Attributes A3 (factor an integer) and A4 (multiply two fractions). Factoring an integer (A3) and multiplying two factors cannot be done according to the model hierarchy without being able to multiply two integers (A2).

2b. How many times will Attribute A2 be probed by the examinee attempting the 20 questions?

Answer: Attribute A2 will be probed 16 times, i.e., the sum of the 2nd row.

2c. If an examinee possesses Attributes A1, A2, and A4 only (1101000), and the examinee's ideal item response vector is (11010000000000000000), i.e., Items 1, 2, and 4 are correct, and all others are wrong. How would we explain a binary response vector of (11110000000000000000) for this examinee?

Answer: An examinee's response vector that does not match one of the ideal item response vectors does not fit the model hierarchy. Assuming that our hierarchy is true, one or more slips have occurred. By examining the attributes probed by Items 1, 2, 3, and 4 as given by the reduced Q matrix, getting the first 4 items correct implies that Attributes A1, A2, A3, and A4 are available to the examinee. Thus, one possibility is that a slip has occurred for Question 3 of the form $0 \rightarrow 1$.

2d. With reference to question 2c, in reality we never know with certainty what attributes an examinee possesses and the best we can do is to make some inferences based upon the observed and ideal item response vectors. Suppose that we observe (11110000000000000000) and wish to make some inferences about the attributes possessed by the examinee. Is it reasonable to assume the examinee possesses all 7 attributes and that the observed binary response vector came from the ideal item response vector (11111111111111111111) with 16 slips of the form $1 \rightarrow 0$?

Answer: It is highly unlikely that 16 slips of the form $1 \rightarrow 0$ were made. An examinee possessing all 7 attributes would be expected to get all 20 items correct indicating a high level of competence, thus it would be unlikely that 16 slips would appear.

2e. How can all the combinations of examinee attributes, as represented by the hierarchy, be obtained from the reduced Q matrix?

Answer: The columns of the reduced Q matrix provide all the combination of examinee attributes. There are 20 such combinations.

3. For the following Adjacency (A) matrix, create the attribute hierarchy.

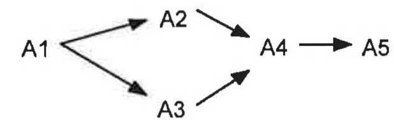
$$\begin{bmatrix} 0&1&0&0&0&0&0 \\ 0&0&1&0&0&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&0&0&1&0&0 \\ 0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&0 \end{bmatrix}$$

Answer:

A1 → A2 → A3 → A4 → A5 → A6 → A7

4. Consider the following two hierarchies:

a) A1 → A2 → A3 → A4 → A5

b) 

Which hierarchy, (a) or (b), represents the following:

4a. In order to possess attribute A5, attributes A1 AND A2 AND A3 AND A4 must be present?

Answer: hierarchy (a)

Note: Consider an electrical switch at each attribute. Any one switch in the open position would prevent reaching A5. In effect we have a series of AND gates.

4b. In order to possess attribute A5, attributes A1 AND (A2 OR A3) AND A4 must be present?

Answer: hierarchy (b)

5. What is an attribute?

Answer: Currently, there is little consensus among researchers on how to define or identify attributes. However, it is generally believed that attributes represent the declarative and procedural knowledge needed to perform a cognitive problem-solving task in a specific domain.

6. If the hierarchy of attributes given in question 1 is associated with a new set of attributes, say for an elementary topic in science, would all the calculations remain the same, i.e., would we obtain the same Adjacency, Reachability, and reduced Q matrices?

Answer: Yes, all the calculations would be the same. Thus, once the structure of the hierarchy is set, the results are applicable without modification to any subject matter.

Note: One must keep in mind that although all the ideal response vectors are identical regardless of the topic of the examination, the observed response vectors could differ substantially as a function of the topic and characteristics of the examinees.

Notes

We thank two anonymous reviewers for their constructive comments on an earlier version of the manuscript. Their suggestions enhanced the clarity of the text and the controversy surrounding some of the issues outlined in our module.

[1] One reviewer of this module argued that rule space is not a psychometric model. Instead, the reviewer saw it as an analytic approach or method stating that the Q or incidence matrix does not include a probabilistic item response function or any other stochastic model of response as would be expected in a psychometric model (for a specific example of this type of model, see DiBello, Stout, & Roussos, 1995). The reviewer also acknowledged that the Q matrix could be thought of as a simple cognitive model that specifies which cognitive attributes are required by each item. It is our position that the A or adjacency matrix, which is based on the attribute hierarchy, is an explicit cognitive model (more so than the Q matrix) designed to account for examinee performance. As a result, we consider Tatsuoka's approach to be a model of cognitive performance. In addition, we maintain the convention for this module that is found repeatedly in the literature, which is to describe Tatsuoka's approach as the *rule-space model.*

[2] Because many factors, such as fatigue or test-wiseness, can mediate an examinee's response to an item, ideal item response patterns are subject to error called *slips*. A slip is considered to be a response to an item that is inconsistent with the attributes probed by the item and present in the examinee. Thus, when the examinee makes slips or errors in responding, the result is an observed response pattern that is inconsistent with the ideal item response pattern for that examinee.

References

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

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *The taxonomy of educational objectives: Handbook I: Cognitive domain.* New York: David McKay.

DiBello, L. V., Stout, W. F., & Roussos, L. A. (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: Erlbaum.

Embretson, S. E. (Ed.). (1985). *Test design: Developments in psychology and psychometrics.* Orlando, FL: Academic Press.

Frederiksen, N., Glaser, R. L., Lesgold, A. M., & Shafto, M. G. (1990). *Diagnostic monitoring of skills and knowledge acquisition.* Hillsdale, NJ: Erlbaum.

Frederiksen, N., Mislevy, R. J., Bejar, I. I. (1993). *Test theory for a new generation of tests.* Hillsdale, NJ: Erlbaum.

Gierl, M. J. (1997a). Comparing the cognitive representations of test developers and students on a mathematics achievement test using Bloom's taxonomy. *Journal of Educational Research, 91*, 26–32.

Gierl, M. J. (1997b). An investigation of the cognitive foundation underlying the rule-space model (Doctoral dissertation, University of Illinois at Urbana-Champaign, 1996). *Dissertation Abstracts International, 57*(08), 5351-B (University Microfilms No. AAC 97-02524).

Katz, I. R., Martinez, M. E., Sheehan, K. M., & Tatsuoka, K. K. (1998). Extending the rule space model to a semantically-rich domain: Diagnostic assessment in architecture. *Journal of Educational and Behavioral Statistics, 23*, 254–278.

Leighton, J., Gierl, M. J., & Hunka, S. (1999, April). Attributes in Tatsuoka's attribute-hierarchy model. Poster presented at the annual meeting of the National Council on Measurement in Education, Montréal, Quebec, Canada.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education, Macmillian.

Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement, 33*, 379–416.

Nichols, P. (1994). A framework of developing cognitively diagnostic assessments. *Review of Educational Research, 64*, 575–603.

Nichols, P. D., Chipman, S. F., Brennan, R. L. (1995). *Cognitively diagnostic assessment.* Hillsdale, NJ: Erlbaum.

Ronning, R., Glover, J., Conoley, J. C., & Witt, J. (1990). *The influence of cognitive psychology on testing and measurement: The Buros-Nebraska Symposium on measurement and testing* (Vol 3.). Hillsdale, NJ: Erlbaum.

Roussos, L. (1994). *Summary and review of cognitive diagnosis models.* Unpublished manuscript, University of Illinois, Urbana-Champaign, The Statistical Laboratory for Educational and Psychological Measurement.

Siegler, R. J. & Jenkins, E. (1989). *How children discover new strategies.* Hillsdale, NJ: Erlbaum.

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills.* Hillsdale, NJ: Erlbaum.

Sheehan, K., Tatsuoka, K. K., Lewis, C. (1992). *A diagnostic classification model for problem solving skills* (Tech. Rep. No. RR-92-39-ONR). Princeton, NJ: Educational Testing Service.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 263–331). New York: American Council on Education, Macmillian.

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

Tatsuoka, K. K. (1984). Caution indices based on item response theory. *Psychometrika, 49*, 95–110.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Fredrickson, R. L. Glaser, A. M. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skills and knowledge acquisition* (pp. 453–488). Hillsdale, NJ: Erlbaum.

Tatsuoka, K. K. (1991) *Boolean algebra applied to determination of universal set of knowledge states* (Tech. Rep. No. RR-91-44-ONR). Princeton, NJ: Educational Testing Service.

Tatsuoka, K. K. (1993). Item construction and psychometric models appropriate for constructed responses. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 107–133). Hillsdale, NJ: Erlbaum.

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327–359). Hillsdale, NJ: Erlbaum.

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

Tatsuoka, K. K., Birenbaum, M., Lewis, C., & Sheehan, K. (1992). *Proficiency scaling based on attribute characteristic curves* (Tech. Rep. No. RR-92-14-ONR). Princeton, NJ: Educational Testing Service.

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

Tatsuoka, M. M., & Tatsuoka, K. K. (1989). Rule space. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (pp. 217–220). New York: Wiley.

van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory.* New York: Springer.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 527–579). Cambridge, MA: MIT Press.

Wolfram, S. (1996). *The mathematica book* (3rd ed.). Cambridge: Cambridge University Press.

Varadi, F., & Tatsuoka, K. K. (1989). BUGLIB [Computer program]. Trenton, NJ: Tanar Software.

Varadi, F., & Tatsuoka, K. K. (1992). BUGLIB Modified version [Computer program]. Trenton, NJ: Tanar Software.

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION *(Required by 39 U.S.C. 3685).* 1. Publication Title: Educational Measurement: Issues and Practice. 2. Publication No.: 6808-90. 3. Filing Date: September 25, 2000. 4. Issue Frequency: Quarterly. 5. No. of Issues Published Annually: Four. 6. Annual Subscription Price: $10/$30/$35. 7. Complete Mailing Address of Known Office of Publication: 1230 17th St., NW, Washington, DC 20036-3078. Contact Person: Trish Thomas. Telephone: 202-223-9485. 8. Complete Mailing Address of Headquarters of General Business Office of Publisher: 1230 17th St., NW, Washington, DC 20036-3078. 9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor. Publisher: National Council on Measurement in Education, 1230 17th St., NW, Washington, DC 20036-3078. Editor: Jeffrey K. Smith, School of Education, Rutgers University, New Brunswick, NJ 08903. Managing Editor: National Council on Measurement in Education, 1230 17th St., NW, Washington, DC 20036-3078. 10. Owner: National Council on Measurement in Education, 1230 17th St., NW, Washington, DC 20036-3078. 11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1% or More of Total Amount of Bonds, Mortgages, or Other Securities: None. 12. For completion of nonprofit organizations authorized to mail at special rates. The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes has not changed during preceding 12 months. 13. Publication Name: Educational Measurement: Issues and Practice. 14. Issue Date for Circulation Data Below: July 2000. 15. Extent and Nature of Circulation. a. Total No. Copies: Average No. Copies Each Issue During Preceding 12 Months 2,750; No. Copies of Single Issue Published Nearest to Filing Date 2,736. b. Paid and/or Requested Circulation. (1) Paid/Requested Outside-County Mail Subscriptions Stated on Form 3541: Average No. Copies Each Issue During Preceding 12 Months 2,360; No. Copies Single Issue Published Nearest to Filing Date 2,482. (2) Paid In-County Subscriptions: Average No. Copies Each Issue During Preceding 12 Months 0; No. Copies of Single Issue Published Nearest to Filing Date 0. (3) Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Non-USPS Paid Distribution: Average No. Copies Each Issue During Preceding 12 Months 0; No. Copies of Single Issue Published Nearest to Filing Date 0. (4) Other Classes Mailed Through the USPS: Average No. Copies Each Issue During Preceding 12 Months 0; No. Copies Single Issue Published nearest to Filing Date 0. c. Total Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months 2,360; No. Copies of Single Issue Published Nearest to Filing Date 2,482. d. Free Distribution by Mail. (1) Outside-County as Stated on Form 3541: Average No. Copies Each Issue During Preceding 12 Months 0; No. Copies Single Issue Published Nearest to Filing Date 0. (2) In-County as Stated on Form 3541: Average No. Copies Each Issue During Preceding 12 Months 0; No. Copies Single Issue Published Nearest to Filing Date 0. (3) Other Classes Mailed Through the USPS: Average No. Copies Each Issue During Preceding 12 Months 23; No. Copies Single Issue Published Nearest to Filing Date 41. e. Free Distribution Outside the Mail: Average No. Copies Each Issue During Preceding 12 Months 0; No. Copies of Single Issue Published Nearest to Filing Date 0. f. Total Free Distribution: Average No. Copies Each Issue During Preceding 12 Months 23; No. Copies of Single Issue Published Nearest to Filing Date 41. g. Total Distribution: Average No. Copies Each Issue During Preceding 12 Months 2,383. No. Copies of Single Issue Published Nearest to Filing Date 2,523. h. Copies Not Distributed. Average No. Copies Each Issue During Preceding 12 Months 367; No. Copies of Single Issue Published Nearest to Filing Date 213. i. Total: Average No. Copies Each Issue During Preceding 12 Months 2,750; No. Copies of Single Issue Published Nearest to Filing Date 2,736. j. Percent Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months 99%; No. Copies of Single Issue Published Nearest to Filing Date 98.3%. 16. This Statement of Ownership will be printed in the Fall 2000 issue of this publication. I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions *(including fines and imprisonment)* and/or civil sanctions *(including multiple damages and civil penalties).* Trish Thomas, Director of Publications, September 25, 2000.