An alternative examination of Chinese Taipei mathematics achievement: Application of the rule-space method to TIMSS 1999 data

Yi-Hsin Chen
University of South Florida, Tampa, Florida, United States of America

Joanna S. Gorin and Marilyn S. Thompson
Arizona State University, Tempe, Arizona, United States of America

Kikumi K. Tatsuoka
Columbia University, New York, United States of America

Rule-space methodology employing a well-designed cognitive model validated for the Chinese Taipei TIMSS 1999 sample of Grade 8 students was used to produce a diagnostic description of these students' cognitive abilities and skills relative to the TIMSS assessment items. The study also looked at the students' distribution across identified knowledge states. Using a mastery probability cut-off criterion of 0.85 for the 23 previously established attributes, the researchers determined that the Chinese Taipei students, as a group, had not mastered five attributes, including number sense, approximation and estimation, recognizing patterns, logical reasoning, and quantitative and logical reading. The descriptions of cognitive attributes suggest that Chinese Taipei students have somewhat weak high-level mathematical thinking skills compared to their other mastery attributes. Twelve clustered knowledge states were identified from all individual knowledge states and were categorized into four performance levels. Some suggestions based on the cognitive diagnostic information provided by this study are linked to the Chinese Taipei educational context and instructional practices.

INTRODUCTION

It is reasonable to assume that the primary goal of standardized educational assessment is to obtain information on student learning so as to improve it (Linn, 1989). Clear definitions of what constitutes student learning are therefore critical. One means by which educators define learning is through the use of behavioral objectives that specify the types of behavior an individual is expected to perform. Because the majority of behaviors in classrooms center on problem-solving or other mental functions, objectives are often phrased in terms of cognitive processes. Research that analyzes test results explicitly in terms of cognitive processes therefore is timely.

Although test scores are generally presented as "the number correct" or as a scale score, they are nonetheless useful measures of learning to the extent that they measure the cognitive processes represented by the tasks on the test. Unfortunately, the traditional psychometric models that underlie most educational tests (notably those based on classical test theory or item response theory) do not reflect the *cognitive* information embedded in test scores (Snow & Lohman, 1989). As Embretson (1993) has observed, "although the item response theory models that are typically applied have many advantages over earlier testing methods, they have little connection with the concerns of cognitive theory about the processes, strategies, and knowledge structures that underlie item solving" (p. 125). Thus, one of the greatest limitations of traditional psychometric models is that they provide only external estimates of student performance on tests, such as total scores or theta estimates, rather than internal information regarding what knowledge or skills students actually possess as evidenced by correct answers. The restriction exists primarily because traditional psychometric models are limited on cognitive grounds (Herman, 1991; Snow & Lohman, 1989). The deficiency of cognitive information available from traditional ability estimates in turn limits the utility of the tests as a means of diagnostic feedback to instructors in classrooms. As a result, achievement tests tend to be used for selection, placement, and certification purposes, rather than for diagnosing student strengths and weaknesses.

Having recognized the potential benefits of cognitive analysis, a number of researchers have combined cognitive information with modern psychometric models in order to fulfill various educational assessment purposes. Examinees' responses are still important within each of these approaches, but they are not dominant. The models, which collectively, as Stout (2002) maintains, can be conceptualized as skills-based (i.e., cognitive-based), include the following: Fischer's (1973) logistic latent-trait model (LLTM); Embretson's (1984, 1985) series of multidimensional, continuous-trait, non-compensatory logistic IRT models; Tatsuoka's (1983, 1990, 1995) rule-space method; and Mislevy's (1994) Bayes Net Approach. Although these methods differ in their major goals and estimation approaches, they all have one feature in common—the addition of cognitive information (Embretson & Gorin, 2001; Embretson & Reise, 2000; Gorin, 2002). In this paper, we take a closer look at one of these approaches—the rule-space method (RSM)—in relation to data from TIMSS 1999 (also known as TIMSS-Repeat or TIMSS-R).

The Rule-Space Method

Developed by K. K. Tatsuoka (1983, 1990, 1995), the rule-space method (RSM) is a mathematically probabilistic approach to analyzing test data that involves incorporating cognitive skills-based information into psychometric models. In addition to taking account of the traditional ability estimate provided for each examinee, the RSM diagnoses the cognitive attributes of individuals and groups of students according to their patterns of responses on a test. Attributes include knowledge, skills, and the processing abilities required to successfully solve a test problem (Birenbaum, Kelly, & Tatsuoka, 1993). The approach taken when describing them is very similar to that for behavioral objectives. RSM thus provides a diagnostic profile for an individual or a group that is based on their mastery or non-mastery of the cognitive attributes underlying the items of a test. Each particular pattern of mastery or non-mastery of attributes relative to a test is known within RSM terminology as the individual's or group's "knowledge state" (Tatsuoka & Tatsuoka, 1987), and each examinee is assigned, based on his or her responses to the test items, to one of several predetermined knowledge states. The RSM model, therefore, is predicated on the assumption that examinees' performance on a task or a test can be directly tied to the knowledge, skills, and processing abilities required for successful task completion that each examinee personally has at hand (Everson, Guerrero, & Yamada, 2003).

RSM comprises two main phases that together include four steps (Tatsuoka, 1995; Tatsuoka & Boodoo, 2000):

- *Phase 1, determination of latent knowledge states:* The two steps in this phase involve (i) identifying unobservable cognitive attributes in a domain of interest and creating the incidence matrix (also known as the Q-matrix), and (ii) determining ideal latent knowledge states (the ideal item-response patterns or classification groups).
- *Phase 2, classification of examinees' knowledge states:* The steps here require (iii) mapping the observed response patterns into the ideal item-response patterns, and (iv) classifying each examinee's responses into the most appropriate knowledge state.

The Q-matrix of the first step depicts the relationships between required cognitive attributes and test items. The list of cognitive attributes and the Q-matrix are thus seen as a representation of the cognitive model used in the rule-space analyses, and it is this incidence matrix that provides the reference point during application of Boolean Descriptive Functions (BDF) in order to determine all possible latent knowledge states (Tatsuoka, 1991). The *predetermined* latent knowledge state, also called an attribute mastery pattern, is expressed as an attribute pattern consisting of a list of the mastered/non-mastered attributes that an examinee uses (Tatsuoka & Boodoo, 2000). Because the latent knowledge states are represented by binary patterns of cognitive attributes that cannot be observed directly, the BDF plays a linking role by setting out these unobservable knowledge states as observable ideal item-response patterns. These patterns, within the context of RSM, are regarded as classification groups.

After identifying the ideal item-response patterns, analysts then compare the examinees' item-response patterns against the *ideal* item-response patterns in order to classify the examinees into one of the predetermined knowledge states. From here, Mahalanobis distances between examinees' response patterns and the ideal response patterns are calculated, and a Bayesian decision rule is used to determine which of several ideal response patterns best matches each examinee's response patterns. Once an individual examinee has been classified into a knowledge state, information regarding his or her mastered/non-mastered attributes is available as useful learning feedback for both examinee and instructor.

The Cognitive Attribute Model Prepared for the TIMSS 1999 Mathematics Test

Corter and Tatsuoka (2002) generated a set of the cognitive knowledge and processing skills underlying performance on the TIMSS 1999 mathematics test. They used three sources of data to create this set: the content and performance frameworks used for the TIMSS 1995 and the TIMSS 1999 mathematics tests (Gonzalez & Miles, 2001); a proposed set of attributes assumed to explain performance on SAT and GRE mathematics items (Tatsuoka, 1995; Tatsuoka & Boodoo, 2000); and written protocols from domain experts. The researchers also validated these attributes by interviewing several students and high school mathematics teachers and by conducting statistical analyses of the TIMSS 1999 sample for the United States (Corter & Tatsuoka, 2002).

To develop the Q-matrix (i.e., to establish the relationships between attributes and test items), the two researchers and one doctoral student independently coded each of the 162 test items according to its identified attributes. They then discussed any differences in coding with one another and reviewed the student protocols in order to reach consensus on the coding for each item. After obtaining an acceptable Q-matrix, the researchers made a series of improvements to it by using statistical approaches such as multiple regression and correlations among attributes until they created what they deemed to be a valid and reliable incidence matrix. In addition, they prepared a manual featuring the coding for the TIMSS 1999 test-item attributes (Tatsuoka, Corter, & Guerrero, 2004). The cognitive attributes and the Q-matrix that eventually emerged from this work comprise the cognitive attribute model for the TIMSS 1999 mathematics test, and this model was the one we applied in the current study.

It is important to note that Tatsuoka and her colleagues (Tatsuoka, Guerrero, Corter, Yamada, & Tatsuoka, 2003) validated the set of cognitive attributes and the Q-matrix through reference to data from 20 of the countries that participated in TIMSS 1999. (These countries included the United States but not Chinese Taipei.) The researchers concluded that the cognitive model adequately represented eighth-graders' mathematics performance on the TIMSS 1999 tests across the 20 countries (Tatsuoka et al., 2003). They also found that the top-scoring countries in the TIMSS 1999 mathematics test (see also Mullis et al., 2000) differed in terms of how they "produced" good performance. For instance, Japanese students tended to learn mathematical thinking skills earlier than they learned content knowledge, while students from Hong Kong acquired solid content knowledge first and mathematical thinking skills later (Tatsuoka et al., 2003). We used the list of cognitive attributes and the Q-matrix validated for the Chinese Taipei sample (Chen, Gorin, Thompson, & Tatsuoka, 2008) to conduct research that might provide us with a better understanding of what a high-performing TIMSS country, such as Chinese Taipei, the third-ranking country in the study, does to educate its students to this degree of success.

Research Purpose

Our overall purpose in this present study was to apply RSM to the TIMSS 1999 mathematics achievement data for Chinese Taipei eighth-graders. We examined the knowledge states most evident among these students in order to provide better descriptive accounts of how well they performed on the TIMSS 1999 items in terms of cognitive attributes and knowledge states. To state our goal in different terms, we wanted to identify which cognitive attributes these students had mastered or not mastered. Another aim was to explore how the students "distributed" across these knowledge states. Finally, we wanted to determine if and how this cognitively diagnostic information linked to Chinese Taipei's educational context and instructional practices.

METHOD

Participants

A total of 5,772 Chinese Taipei students nested within 150 schools participated in TIMSS 1999. The sample for the current study was based on students who completed TIMSS Booklets 1, 3, 5, and 7. Our eventual sample comprised 2,874 students out of the 5,772 students tested.

Instrument

Although the TIMSS mathematics test involved eight question-and-answer booklets, the data that we used for the current study came only from those Chinese Taipei eighth-graders who completed Booklets 1, 3, 5, and 7. We selected these booklets according to the criterion that each attribute to be analyzed in the study had to be included in *at least* three items (see Corter & Tatsuoka, 2002). Note that the mathematics tests used in TIMSS 1999 involved five content categories: (i) fractions and number sense (38% of the total number of items); (ii) measurement (15%); (iii) data representation, analysis, and probability (13%); (iv) geometry (13%); and (v) algebra (22%). Item types involved multiple-choice (77%), short-answer (13%), and extended-response formats (10%) (Gonzalez & Miles, 2001).

Cognitive Attributes and Incidence Matrix

Originally, Corter and Tatsuoka (2002) classified the 27 cognitive attributes that represent the underlying performance of the TIMSS 1999 mathematics test into the three categories shown in Table 1: content knowledge (C1 to C6), cognitive processes (P1 to P10), and skill/item-type (S1 to S11). Also on the basis of previous research (Tatsuoka, Corter, & Tatsuoka, 2004), we decided to delete Attributes C6, S1, S9,

Table 1: Knowledge, skill, and process attributes underlying performance on the TIMSS 1999 mathematics test

CONTENT ATTRIBUTES	
C1	Basic concepts, properties, and operations in whole numbers and integers
C2	Basic concepts, properties, and operations in fractions and decimals
C3	Basic concepts, properties, and operations in elementary algebra
C4	Basic concepts and properties of two-dimensional geometry
C5	Data, probability, and basic statistics
C6	Using tools to measure (or estimate) length, time, angle, and temperature
PROCESS ATTRIBUTES	
P1	Translate/formulate equations and expressions to solve a problem
P2	Computational applications of knowledge in arithmetic and geometry
P3	Judgmental applications of knowledge in arithmetic and geometry
P4	Applying rules in algebra
P5	Logical reasoning such as case reasoning, deductive thinking skills, if–then, necessary and sufficient, generalization skills
P6	Problem search, analytical thinking, problem restructuring, and inductive thinking
P7	Generating, visualizing, and reading figures and graphs
P8	Applying and evaluating mathematical correctness
P9	Management of data and procedures
P10	Quantitative and logical reading
SKILL (ITEM-TYPE) ATTRIBUTES	
S1	Unit conversion
S2	Applying number properties and relationships; number sense/number line
S3	Using figures, tables, charts, and graphs
S4	Approximation/estimation
S5	Evaluate/verify/check options
S6	Patterns and relationships (able to apply inductive thinking skills)
S7	Using proportional reasoning
S8	Solving novel or unfamiliar problems
S9	Comparing two or more entities
S10	Open-ended item, in which an answer is not given
S11	Using words to communicate questions

Note: Adapted from Corter and Tatsuoka (2002, p. 18).

and P8 from the current study because these did not appear in a sufficient number of items in Booklets 1, 3, 5, and 7. We accordingly used in the current study a total of 23 initial cognitive attributes of knowledge and skills (5 content attributes, 9 process attributes, and 9 skill/item-type attributes) as an initial list of attributes for the Chinese Taipei data. Table 2 gives the frequencies of attributes required for the items in the four booklets.

Table 2: Frequencies of attributes required in items for each booklet of the TIMSS 1999 mathematics test

Attribute		Booklet 1	Booklet 3	Booklet 5	Booklet 7	Total items
C1	Whole numbers and integers	14	6	10	10	40
C2	Fractions and decimals	15	19	19	17	70
C3	Elementary algebra	10	9	4	4	27
C4	Two-dimensional geometry	13	14	10	6	43
C5	Data and basic statistics	10	6	10	7	33
S2	Number sense	8	6	5	4	23
S3	Figures, tables, and graphs	19	17	22	15	73
S4	Approximation and estimation	4	6	5	5	20
S5	Evaluate and verify options	17	11	15	17	60
S6	Recognize patterns	6	1	3	4	14
S7	Proportional reasoning	11	12	14	12	49
S8	Novel/unfamiliar problems	10	12	10	8	40
S10	Open-ended items	13	10	12	9	44
S11	Word problems	25	18	18	18	79
P1	Translate	15	16	13	13	57
P2	Computation application	17	17	19	19	72
P3	Judgmental application	9	10	7	8	34
P4	Rule application in algebra	7	9	4	4	24
P5	Logical reasoning	17	13	11	7	48
P6	Solution search	9	10	5	8	32
P7	Visual figures and graphs	13	10	13	8	44
P9	Data management	20	14	13	8	55
P10	Quantitative reading	11	6	9	10	36

After completing their first step of the work, Tatsuoka and her colleagues (Tatsuoka, Corter, & Guerrero, 2004) used the identified attribute list to construct an incidence matrix (a Q-matrix). This specified the exact relationship between the items and the cognitive processing attributes. Unlike the design framework of the TIMSS mathematics test, where each item is categorized into only one content category or one performance category, the Q-matrix that Tatsuoka and her colleagues used did not restrict selection to a specific number of attributes associated with each item (Tatsuoka, Corter, & Tatsuoka, 2004). Thus, an item could involve any number of possible attributes. For instance, one item could involve two content attributes, three process attributes, and one item-type attribute. Figure 1 provides an example of an item and its associated attributes (see also Tatsuoka, Corter, & Guerrero, 2004).

Figure 1: Example of item attribute coding on a TIMSS mathematics item

Problem: A teacher and a doctor each have 45 books. If 4/5 of the teacher's books and 2/3 of the doctor's books are novels, how many more novels does the teacher have than the doctor?

A. 2 B. 3 C. 6 D. 30 E. 36

Attributes involved in: C2, P1, P5, P9, S11

The content is fractions	⟶ C2
Translate the expression into arithmetic	⟶ P1
To understand the comparison expressed by "how many more"	⟶ P5
Two goals problem	⟶ P9
This is a verbally expressed problem	⟶ S11

ANALYSIS

This process involved *rule-space analyses and knowledge state analyses*. We used a special computer software program called BUGSHELL, programmed by Tatsuoka, Varadi, and Tatsuoka (1992), for the rule-space analysis, and the three-dimensional Cartesian coordinate space to formulate the classification space, which consisted of the IRT ability (θ), ζ, and generalized ζ. The variable ζ measured the extent of the unusualness of item-response patterns in the whole test (Tatsuoka, 1984), while the generalized ζ focused the extent of the unusualness on the particular subset of interest (Tatsuoka, 1996).

For the purpose of classification, we also set in advance several relevant parameters for rule-space analyses. We began by setting the acceptable Mahalanobis distance and the difference of θ values between an examinee's item-response pattern and the ideal item-response pattern to 4.5 and 1.5, respectively. Next, we set the number of slips (i.e., the number of differing responses between the observed and the ideal item patterns in the test) to comprise no more than one third of the total items (Corter & Tatsuoka, 2002). Finally, we performed a separate rule-space analysis for each booklet.

Our second main analysis, the knowledge states analysis, initially involved a large number of knowledge states because of the large samples used in the current study. In order to establish more meaningful and interpretable knowledge states for diagnostic information, similar knowledge states were grouped together into a smaller number of such states. We first merged the attribute mastery probability vectors from the four rule-space analyses into a single dataset. According to Tatsuoka, Xin, Corter, and Tatsuoka (2004), combining examinee attribute mastery probabilities derived from different booklets into a single dataset is justified by the common set of cognitive attributes underlying each booklet.

Our next step was to conduct a *K*-means cluster analysis that included applying several criteria in order to obtain an appropriate number of clusters. These criteria were as follows:

1. The number of classified students for each cluster should be greater than 1% of the sample size;
2. The average distance between each classified student's knowledge state and the hypothesized cluster centroid of a knowledge state should be smaller than 2.0, which means that each student's knowledge state in his or her own cluster is closely located; and
3. Observed *F* statistics from univariate ANOVAs conducted for each separate attribute should be high (Tatsuoka et al., 2003).

Because we selected the clusters to maximize the differences among cases in the K-mean cluster analysis, we applied only the *F* values for our descriptive purposes.

After obtaining the hypothesized clusters, we computed the probability estimate for each attribute in a particular cluster by averaging all mastery probabilities on one attribute across the examinees classified into that cluster. Each probability vector represented the centroid of a hypothesized cluster. Using the numeric cut-off of 0.85, we then transformed the probability vector for each cluster into the binary indicator vector, in which 0s represented non-mastery and 1s indicated mastery. (This binary vector is equivalent to the attribute mastery pattern for the clustered knowledge state.) We also calculated a single estimate of mastery probability for each cluster by averaging all of the attribute mastery probabilities in a particular clustered knowledge state.

RESULTS

We begin this section of the paper by providing examples of the attribute mastery probabilities that we obtained for several students. We then present descriptive statistics of the attribute mastery probabilities for the entire sample of Chinese Taipei eighth-graders whose data were used in the current study. The second part of this results section presents findings relating to the clustered knowledge states.

Attribute Mastery Probabilities

Individual Diagnostic Information

The most fundamental and important output from rule-space analyses is the individual diagnostic information relating to students' (both individual and whole sample or population) mastery probabilities of content and process attributes. Figure 2 provides three examples of students' item-response patterns and the corresponding attribute mastery probabilities derived from the rule-space analyses.

Figure 2: Example of students' item-response patterns and the corresponding attribute mastery probabilities

Student A

Booklet taken = 3; Total scores = 38; Percentage correct responses = 90.48;
Item-response pattern (42 items)
111111111111111111111111111011111110110011
Attribute Mastery Probability (23 attributes)

1.00	1.00	1.00	1.00	0.78				
1.00	1.00	0.65	1.00	0.75	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Student B

Booklet taken = 5; Total scores = 32; Percentage correct responses = 76.19;
Item-response pattern (42 items)
111111011111010011111111110011110010101111
Attribute Mastery Probability (23 attributes)

1.00	1.00	1.00	1.00	1.00				
0.60	1.00	0.64	1.00	1.00	1.00	0.79	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.21

Student C

Booklet taken = 3; Total scores = 38; Percentage correct responses = 90.48;
Item-response pattern (42 items)
111111011111111111111111111001111110111111
Attribute Mastery Probability (23 attributes)

0.82	1.00	1.00	1.00	0.80				
0.80	1.00	0.58	1.00	0.44	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

As shown in Figure 2, Student A, who answered around 90% of the test items correctly, had 20 perfect mastery attributes and only three attributes whose mastery probabilities were other than 1. These three attributes were data and basic statistics (C5), approximation and estimation (S4), and recognize patterns (S6). Thus, Student A had lower probabilities (below 0.80) of having mastered these three attributes.

Student B answered 76% of the items correctly and had 19 perfect mastery attributes and 4 attributes with lower mastery probabilities. These four attributes were number sense (S2), approximation and estimation (S4), novel/unfamiliar problems (S8), and quantitative reading (P10). Student B therefore had lower mastery probabilities for these four attributes, with a particularly low mastery probability for quantitative reading (P10; P_{P10} = 0.21), indicating he or she had a low likelihood of correctly answering items requiring this attribute.

On comparing Student C with Student A in Figure 2, we can see that both students obtained the same total score (38) in Booklet 3, which translated into their correctly answering about 90% of the items. However, the attribute mastery profile of Student C differed from that of Student A because of differences in the item-response patterns. Student C had five imperfect mastery attributes. Of these, the item-response patterns for data and basic statistics (C5), approximation and estimation (S4), and recognize patterns (S6) were imperfect, as was the case for Student A. Additionally, Student C had less than ideal response patterns for whole numbers and integers (C1) and number sense (S2). Furthermore, both students had somewhat different patterns of probability values on the imperfect mastery attributes.

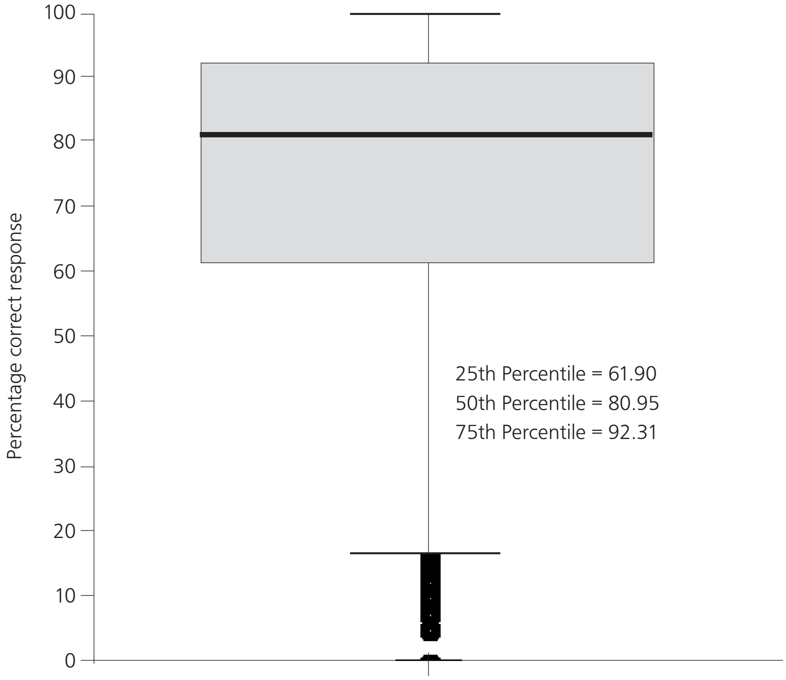
Test Percentage Score Distribution for the Chinese Taipei Sample

Before summarizing the attribute mastery probabilities, we consider it useful to briefly mention the performance of the Chinese Taipei students on the TIMSS 1999 mathematics test. The box plot in Figure 3 shows that 25% of the students correctly answered 92% or more of the items, 50% of the students correctly answered 81% or more of the items, and 75% of the students correctly answered 62% or more of the items. In general, the distribution shows that Chinese Taipei students performed very well on the TIMSS 1999 mathematics items.

Descriptive Statistics of Attribute Mastery Probabilities

Table 3 presents the means and standard deviations of the mastery probabilities on the 23 attributes across the entire sample. As shown in the table, all attributes other than recognize patterns (S6) had mastery probabilities greater than 0.80. Fifteen of the 23 attributes had mean mastery probabilities above 0.90. This finding reflects the excellent performance of the Chinese Taipei students on the TIMSS 1999 mathematics tests, which is consistent with the fact that Chinese Taipei was the third highest performing country (according to scaled scores) of the 38 countries that participated in the study (Gonzalez & Miles, 2001). While it is useful to know how well the Chinese Taipei students fared, the information we obtained relative to the attribute mastery probabilities provides a more detailed examination of student learning than does the information provided by the scale scores.

Figure 3: Box plot of percentage of correct responses on the TIMSS 1999 mathematics test for Chinese Taipei students



Among the five content attributes, Chinese Taipei students had the lowest mean probability on elementary algebra (C3) (P_{C3} = 0.88). They had the highest mean probabilities on whole numbers and integers (C1) (P_{C1} = 0.97) and fractions and decimals (C2) (P_{C2} = 0.98). In the skill/item-type attributes, the students performed very well on figures, tables, and graphs (S3), evaluate and verify options (S5), and word problems (S11), with mean probabilities of 0.99. For figures, tables, and graphs (S3) and word problems (S11), students had minimum mastery probabilities greater than 0 (P_{S3} = 0.20 and P_{S11} = 0.17, respectively). These findings mean that students who had the lowest overall performance still had approximately a 20% chance of mastering each of the two attributes. In addition, for recognize patterns (S6), two attributes had lower mean probabilities of 0.82. They were number sense (S2) and approximation and estimation (S4). The Chinese Taipei students performed particularly well on three process attributes: translate (P1) (P_{P1} = 0.98), computation application (P2) (P_{P2} = 0.98), and judgmental application (P3) (P_{P3} = 0.97). Lower achievement attributes were logical reasoning (P5) (P_{P5} = 0.84) and quantitative reading (P10) (P_{P10} = 0.83).

Using a mastery probability cut-off criterion of 0.85, we found that the Chinese Taipei students collectively mastered 18 of the 23 attributes and failed to master five. These non-mastered attributes included three skill attributes—number sense (S2), approximation and estimation (S4), and recognize patterns (S6)—as well as two process attributes—logical reasoning (P5) and quantitative reading (P10). The most

Table 3: Descriptive statistics of attribute probabilities for total sample (*N*=2,863)

Attribute		Mean[a]	*SD*	Minimum	Maximum
C1	Whole numbers and integers	0.97	0.11	0.00	1.00
C2	Fractions and decimals	0.98	0.11	0.00	1.00
C3	Elementary algebra	0.88	0.24	0.00	1.00
C4	Two-dimensional geometry	0.93	0.18	0.00	1.00
C5	Data and basic statistics	0.95	0.13	0.00	1.00
S2	Number sense	**0.82**	0.22	0.00	1.00
S3	Figures, tables, and graphs	0.99	0.05	0.20	1.00
S4	Approximation and estimation	**0.82**	0.19	0.00	1.00
S5	Evaluate and verify options	0.99	0.06	0.00	1.00
S6	Recognize patterns	**0.65**	0.29	0.00	1.00
S7	Proportional reasoning	0.97	0.12	0.00	1.00
S8	Novel/unfamiliar problems	0.92	0.17	0.00	1.00
S10	Open-ended items	0.87	0.26	0.00	1.00
S11	Word problems	0.99	0.06	0.17	1.00
P1	Translate	0.98	0.11	0.00	1.00
P2	Computation application	0.98	0.10	0.00	1.00
P3	Judgmental application	0.97	0.10	0.00	1.00
P4	Rule application in algebra	0.88	0.25	0.00	1.00
P5	Logical reasoning	**0.84**	0.29	0.00	1.00
P6	Solution search	0.92	0.18	0.00	1.00
P7	Visual figures and graphs	0.94	0.19	0.00	1.00
P9	Data management	0.94	0.17	0.00	1.00
P10	Quantitative reading	**0.83**	0.26	0.00	1.00

Note: [a] Mean mastery probabilities fall below cut-off of 0.85 for the attributes in bold type.

difficult attribute with the lowest probability of mastery for the students was recognize patterns (S6). In an earlier analysis, Tatsuoka and her colleagues (Tatsuoka, Corter, & Tatsuoka, 2004) found that recognize patterns was the most difficult attribute for students to master across the entire 20-country sample examined.

Although these numeric results are important, it is their interpretations in terms of students' skills that are of primary interest. According to Tatsuoka, Corter, and Guerrero's (2004) descriptions of cognitive attributes, recognize patterns (S6) is a skill that involves recognizing numeric, geometric, and/or algebraic patterns and finding a rule for generating those patterns. Within the realm of mathematics learning, this attribute can be viewed as an inductive thinking skill. Number sense (S2) is a skill that involves converting two or three different units into a comparable unit. In other words, the skill involves applying number properties and relationships. Approximation and estimation (S4) is a skill that requires the ability to estimate and approximate decimals or fractions in numbers, as well as in relation to areas or volumes in geometrical shapes. Logical reasoning (P5) includes case reasoning, deductive thinking skills, if–then, necessary and sufficient, and generalization skills. Quantitative and logical reading (P10) involves the ability to read and comprehend sentences containing mathematically quantitative terminology, such as "at least," "comparisons," "must be," and "increasing and decreasing," as well as logical quantifiers like "for every," "for any," and "for a given" (Tatsuoka, Corter, & Guerrero, 2004). These descriptions, and the results of the rule-space analyses with the Chinese Taipei sample, suggest that Chinese Taipei students are, in terms of the various mastery attributes, weakest in high-level mathematical thinking skills.

Clustered Knowledge States

The Cluster Analysis

This analysis specified solutions for 8 to 12 clusters. Table 4 presents the numbers of mastery attributes of the clustered knowledge states for each cluster solution. Because the goal of clustering is to explore educationally interpretable groups of students' attribute mastery probabilities and hierarchical relationships among these groups, we selected the 12-cluster solution as the final solution for the *K*-mean cluster analysis in this study. We considered other solutions, including 8- to 11-cluster solutions, but decided that these were not optimal. Our rationale for endorsing the 12-cluster solution follows.

First, some solutions did not yield the clustered knowledge state representing students who mastered all 23 attributes, such as the 9- and 10-cluster solutions (refer to Table 4). However, 325 students from the Chinese Taipei sample did master all attributes. In contrast, some solutions, such as the 8-cluster, did not yield the knowledge state reflecting students who mastered only a few attributes. Table 4 shows that the lowest numbers of clustered knowledge states in the eight-cluster solution were seven and eight attributes (except for 0 attributes). However, the clustered knowledge state characterized by four mastery attributes was evident in the 9-, 11-, and 12-cluster solutions. As for the 11-cluster solution, we were unable to derive interpretable hierarchical relationships among the clustered knowledge states. Essentially, interpretations for the 8- to 11-cluster solutions were problematic, while the 12-cluster solution yielded the clustered knowledge states that not only reflected students' attribute performances at different achievement levels, but also supported interpretable hierarchical relationships among them. Further statistical evidence is provided below to support our claim that the 12-cluster solution allowed apportionment of attribute masteries to separate distinct clusters.

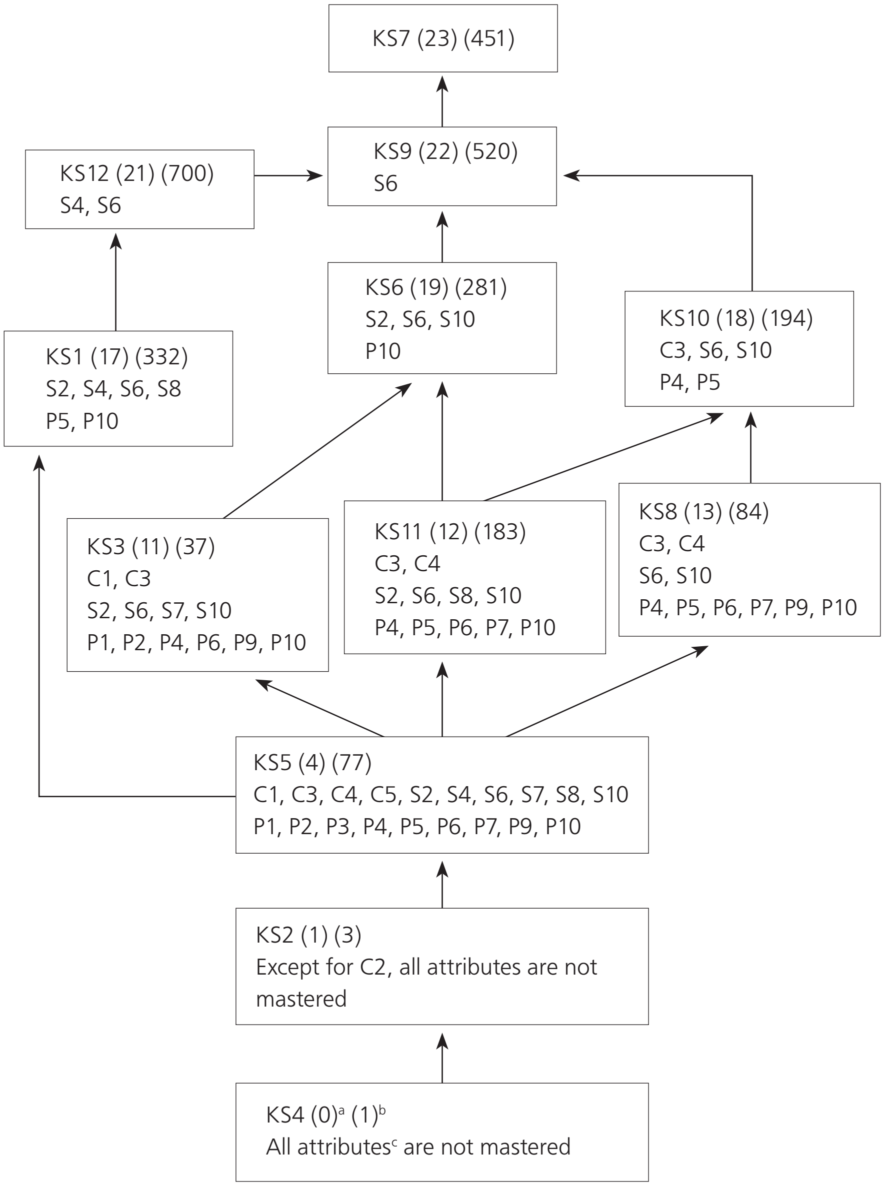
Table 4: Number of mastered attributes associated with each of the clustered knowledge states for different cluster solutions

Knowledge state	Solution				
	8-cluster	*9-cluster*	*10-cluster*	*11-cluster*	*12-cluster*
KS 1	18	0	8	19	17
KS 2	18	18	12	16	1
KS 3	13	21	0	10	11
KS 4	23	9	17	10	0
KS 5	8	14	18	4	4
KS 6	21	4	7	23	19
KS 7	7	18	21	18	23
KS 8	0	18	18	13	13
KS 9		10	9	0	22
KS 10			22	19	18
KS 11				21	12
KS 12					21

We labeled the 12 clusters KS1 to KS12. The average distance between the students and their classified cluster centroids was 0.51 and ranged from 0 to 1.62. The distance values from cluster centroids were also quite small, thereby indicating that, with the 12-cluster solution, the classified students were representative of their clusters. In addition, the numbers of students in 10 of the clusters for the 12-cluster solution were substantial, with each cluster containing more than 1% of the total sample (> 30 students). Exceptions were the clusters for KS2 and KS4 (three students and one student, respectively). KS2 and KS4 represented students with low mastery probabilities on all or almost all of the attributes. Few students were classified into KS2 and KS4 because of the high mathematics achievement overall of the Chinese Taipei students. These two are cognitively and educationally interpretable knowledge states in terms of the hierarchical relationships shown in Figure 4. These reasons led us to apply, in this study, the 12-cluster solution that included KS2 and KS4.

We also used univariate analysis of variance (ANOVA) on each attribute in order to understand how important each was in separating out the groups in the 12-clustered solution. The results on each attribute for the 12-cluster solution appear in Table 5. However, these results, especially the *F* values, serve only as guidelines for descriptive purposes because the *K*-mean cluster analysis maximized the differences among the cases in the different clusters. The results in Table 5 show that the *F* values were far from 1.0 and quite large; thus, by using the 12-cluster solution, we were able to show that every attribute was useful for separating clusters. Accordingly, we agreed that the 12-cluster solution was a cognitively interpretable solution in this study.

Figure 4: A hierarchically ordered network among the clustered knowledge states



Notes:

[a] The first number in parenthesis represents the number of mastery attributes.

[b] The second number in parenthesis represents sample size.

[c] Attributes presented in the rectangles were not mastered.

Table 5: One-way ANOVA results evaluating differences across the 12 knowledge states for each attribute

		Attribute	Cluster	Error
		Mean square (*df* = 11)	Mean square (*df* = 2851)	*F*
C1	Whole numbers and integers	1.162	0.007	175.786
C2	Fractions and decimals	0.391	0.011	35.502
C3	Elementary algebra	9.714	0.018	532.022
C4	Two-dimensional geometry	3.652	0.020	181.231
C5	Data and basic statistics	0.842	0.014	59.642
S2	Number sense	4.073	0.033	123.826
S3	Figures, tables, and graphs	0.120	0.002	58.944
S4	Approximation and estimation	3.751	0.022	167.813
S5	Evaluate and verify options	0.199	0.002	86.830
S6	Recognize patterns	15.578	0.026	607.143
S7	Proportional reasoning	1.537	0.010	157.827
S8	Novel/unfamiliar problems	1.835	0.021	88.816
S10	Open-ended items	9.438	0.033	289.037
S11	Word problems	0.136	0.003	40.905
P1	Translate	1.287	0.006	202.054
P2	Computation application	0.773	0.008	100.606
P3	Judgmental application	0.360	0.009	38.004
P4	Rule application in algebra	11.263	0.019	577.865
P5	Logical reasoning	11.367	0.040	284.532
P6	Solution search	3.193	0.019	171.370
P7	Visual figures and graphs	5.468	0.014	387.972
P9	Data management	1.645	0.023	73.051
P10	Quantitative reading	9.347	0.030	313.577

Note: $p < 0.001$ for all attributes.

Description for the Clustered Knowledge States

Table 6 presents the means (centroids) of the mastery probabilities of the 23 attributes for each of the 12 clustered knowledge states across the students classified into each group. The mean probabilities across the 23 attributes and the sample sizes for each of the 12 clustered knowledge states appear at the bottom of Table 6. By applying the cut-off of 0.85, we were able to transform the vectors of the mean mastery probabilities for the 12 clustered knowledge states into binary indictor vectors, thereby forming attribute mastery patterns for each of the 12 clustered knowledge states. These are shown in Table 7.

The attribute mastery patterns corresponding to the knowledge states presented in Table 7 facilitate interpretation of the cluster-analysis results in Table 6. Here, we computed the total numbers of mastery attributes for the clustered knowledge states and then sorted the knowledge states by the descending order of the total number from the left-most to the right-most columns. Table 7 therefore also presents the mean mastery probabilities and sample sizes adopted from Table 6 for the 12-cluster solution of knowledge states.

As is evident in Table 7, knowledge states KS7, KS9, and KS12 had the highest mean mastery probabilities with 0.98, 0.96, and 0.96, respectively. These three knowledge states not only corresponded to the binary indicator vectors for mastery attributes but also had the highest numbers of mastery attributes. KS7 was represented by all 23 mastery attributes. KS9 consisted of 22 mastery attributes and one non-mastery attribute, namely, recognize patterns (S6). KS12 comprised 21 mastery attributes and two non-mastery attributes—recognize patterns (S6) and approximation and estimation (S4)—which were the most difficult attributes for the students to master. Because these three knowledge states had the highest mastery probabilities and included the largest numbers of mastery attributes, it was possible to categorize them into the highest level of performance knowledge states for the Chinese Taipei sample. The students classified into these three knowledge states failed to master only two (at most) of the 23 attributes. Approximately 58% of the students (i.e., 1,671 students out of the total sample of 2,863) "classified" into this highest level of performance knowledge state.

KS1, KS6, and KS10 had mean mastery probabilities of 0.91, 0.88, and 0.87, respectively. These three can therefore be thought of as the second highest level of performance knowledge state for the Chinese Taipei students because their mean mastery probabilities of knowledge states were still greater than 0.85, which was a criterion for individual attribute mastery. Approximately 28% of the students (i.e., 807 out of 2,863) fell within this second highest level of performance knowledge state. In sum, 86% of Chinese Taipei students fell within the top two levels of knowledge states.

KS3, KS8, and KS11 had mean mastery probabilities of 0.70, 0.76, and 0.79, respectively. The corresponding numbers of mastery attributes were 11, 13, and 12. Interestingly, more process attributes than other attributes were not mastered in these three knowledge states. Given that KS3, KS8, and KS11 had around half of all 23 attributes mastered, we categorized these three knowledge states into the middle level of performance. Approximately 11% of the students (304 out of 2,863) classified into these mid-level performance knowledge states.

Finally, KS5, KS2, and KS4 had the lowest mean mastery probabilities of 0.68, 0.46, and 0.17, respectively. KS5 had four attributes mastered, while KS2 had only one. KS4 had no attributes mastered. The mastery attributes in KS5 comprised fractions and decimals (C2), figures, tables, and graphs (S3), evaluate and verify options (S5), and word problems (S11). These mastery attributes in KS5 were the easiest attributes

Table 6: Centroids of the clustered knowledge states

Attribute		Knowledge states											
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
C1	Whole numbers and integers	1	0.27	0.77	0.20	0.67	0.94	1	0.89	0.99	0.97	0.97	1
C2	Fractions and decimals	0.98	1	0.86	0.00	0.89	0.98	1	0.97	1	0.97	0.89	1
C3	Elementary algebra	0.96	0.07	0.21	0.00	0.51	0.94	0.99	0.72	0.99	0.54	0.47	0.98
C4	Two-dimensional geometry	0.93	0.20	0.93	0.00	0.66	0.97	0.97	0.46	1	0.91	0.71	0.97
C5	Data and basic statistics	0.99	0.20	0.95	0.20	0.72	0.88	0.99	0.94	0.95	0.93	0.98	0.96
S2	Number sense	0.60	0.60	0.76	0.00	0.58	0.74	0.97	0.90	0.91	0.89	0.65	0.85
S3	Figures, tables, and graphs	1	0.67	1	0.40	0.91	0.99	1	0.98	1	0.99	0.98	1
S4	Approximation and estimation	0.77	0.73	0.89	0.00	0.74	0.87	0.98	0.87	0.90	0.87	0.86	0.64
S5	Evaluate and verify options	1	0.53	0.96	0.20	0.89	0.99	1	0.98	1	0.99	1	1
S6	Recognize patterns	0.83	0.00	0.18	0.00	0.19	0.24	0.98	0.33	0.53	0.72	0.39	0.78
S7	Proportional reasoning	0.97	0.73	0.37	0.40	0.80	0.98	0.98	0.97	0.99	0.94	0.99	0.99
S8	Novel/unfamiliar problems	0.73	0.73	0.85	0.60	0.81	0.94	0.96	0.96	0.98	0.92	0.81	0.98
S10	Open-ended items	0.97	0.07	0.47	0.00	0.46	0.63	0.99	0.49	0.97	0.81	0.49	1
S11	Word problems	1	0.67	0.95	0.40	0.90	0.99	1	0.99	1	0.98	0.98	1
P1	Translate	0.99	0.40	0.48	0.20	0.79	0.99	0.99	1	1	0.95	0.98	1
P2	Computation application	0.99	0.80	0.84	0.00	0.70	0.99	0.99	0.98	1	0.96	0.97	1
P3	Judgmental application	0.99	0.67	0.86	0.20	0.83	0.98	0.99	0.96	0.99	0.95	0.92	0.98
P4	Rule application in algebra	0.97	0.13	0.18	0.00	0.41	0.96	0.97	0.59	0.97	0.42	0.59	1
P5	Logical reasoning	0.69	0.47	0.86	0.60	0.68	0.90	0.96	0.26	0.95	0.79	0.27	0.98
P6	Solution search	0.97	0.53	0.30	0.20	0.59	0.88	0.99	0.78	0.95	0.88	0.79	0.98
P7	Visual figures and graphs	0.98	0.67	0.93	0.00	0.72	0.93	0.99	0.19	1	0.94	0.83	1
P9	Data management	0.94	0.20	0.78	0.00	0.59	0.95	0.96	0.77	0.98	0.90	0.91	0.99
P10	Quantitative reading	0.58	0.33	0.76	0.20	0.59	0.50	0.98	0.43	0.95	0.87	0.77	0.97
Mean		0.91	0.46	0.70	0.17	0.68	0.88	0.98	0.76	0.96	0.87	0.79	0.96
Sample size		332	3	37	1	77	281	451	84	520	194	183	700

for the entire Chinese Taipei sample. It was therefore possible to categorize KS5, KS2, and KS4 into the low level of performance because students mastered so few of their attributes. Only 3% (81 out of 2,863) of the students fell within this level of performance. The mean mastery probability across 23 attributes for each of the middle- and low-level performing knowledge states was below the 0.85 criterion. Thus, only

Table 7: Order attribute mastery patterns for the clustered knowledge states

Attribute		Knowledge states											
		7	*9*	*12*	*6*	*10*	*1*	*8*	*11*	*3*	*5*	*2*	*4*
C1	Whole numbers and integers	1	1	1	1	1	1	1	1	0	0	0	0
C2	Fractions and decimals	1	1	1	1	1	1	1	1	1	1	1	0
C3	Elementary algebra	1	1	1	1	0	1	0	0	0	0	0	0
C4	Two-dimensional geometry	1	1	1	1	1	1	0	0	1	0	0	0
C5	Data and basic statistics	1	1	1	1	1	1	1	1	1	0	0	0
S2	Number sense	1	1	1	0	1	0	1	0	0	0	0	0
S3	Figures, tables, and graphs	1	1	1	1	1	1	1	1	1	1	0	0
S4	Approximation and estimation	1	1	0	1	1	0	1	1	1	0	0	0
S5	Evaluate and verify options	1	1	1	1	1	1	1	1	1	1	0	0
S6	Recognize patterns	1	0	0	0	0	0	0	0	0	0	0	0
S7	Proportional reasoning	1	1	1	1	1	1	1	1	0	0	0	0
S8	Novel/unfamiliar problems	1	1	1	1	1	0	1	0	1	0	0	0
S10	Open-ended items	1	1	1	0	0	1	0	0	0	0	0	0
S11	Word problems	1	1	1	1	1	1	1	1	1	1	0	0
P1	Translate	1	1	1	1	1	1	1	1	0	0	0	0
P2	Computation application	1	1	1	1	1	1	1	1	0	0	0	0
P3	Judgmental application	1	1	1	1	1	1	1	1	1	0	0	0
P4	Rule application in algebra	1	1	1	1	0	1	0	0	0	0	0	0
P5	Logical reasoning	1	1	1	1	0	0	0	0	1	0	0	0
P6	Solution search	1	1	1	1	1	1	0	0	0	0	0	0
P7	Visual figures and graphs	1	1	1	1	1	1	0	0	1	0	0	0
P9	Data management	1	1	1	1	1	1	0	1	0	0	0	0
P10	Quantitative reading	1	1	1	0	1	0	0	0	0	0	0	0
Number of mastery attributes		23	22	21	19	18	17	13	12	11	4	1	0
Mean of mastery probability		0.98	0.96	0.96	0.88	0.87	0.91	0.76	0.79	0.70	0.68	0.46	0.17
Sample size		451	520	700	281	194	332	84	183	37	77	3	1

Note: The cut-off point for mastery probability was set at 0.85.

14% of the Chinese Taipei students fell into these two lowest-level knowledge states.

A Hierarchically-ordered Network

Our analysis also involved applying the principles of inclusion relations and graph theory to establish the hierarchical relationships and the network existing among the knowledge states. As seen in Table 7, knowledge state 7 (KS7) had a hierarchical relationship with knowledge state 9 (KS9) because each component in the binary mastery vector of KS7 was larger than or equal to the relative component in the mastery vector of KS9. Figure 4 presents a hierarchically ordered network that depicts the various relationships among these 12 clustered knowledge states. The knowledge states with few mastered attributes, such as KS2 and KS4, were located in the lower portion of the network. In contrast, the knowledge states with more mastered attributes, such as KS7 and KS9, were located in the upper portion of the network.

DISCUSSION AND CONCLUSIONS

Implications of the Results

The empirical study presented in this paper involved a substantive examination of the mathematics achievement of Chinese Taipei Grade 8 students who participated in TIMSS 1999 (TIMSS-R). We applied rule-space methodology (RSM) to produce a diagnostic description of these students' cognitive abilities and skills as related to the TIMSS items. RSM employed within the auspices of a well-designed cognitive model successfully provided diagnostic information for the Chinese Taipei sample, and in so doing confirmed, as have other studies, that RSM is a viable alternative to traditional psychometric analysis of test scores.

In the TIMSS 1999 survey, Chinese Taipei ranked third among the 38 participating countries, based on the first plausible values on the mathematics test. Thus, in terms of an overall estimate of student ability across countries, Chinese Taipei students proved to be among the most able. The reproduction of these results with RSM supports the usefulness of this method for providing diagnostic information that augments overall score estimates with descriptions of cognitive attributes for a particular population of interest. Note, however, that RSM cannot provide diagnostic information relating to attribute mastery probabilities for students who have not been successfully classified into one of the predetermined knowledge states. If methods such as rule-space were implemented for score reporting, then handling of score reports for students not classified into a diagnostic group would need to be considered carefully. The best way to avoid this situation is to develop a complete and accurate Q-matrix that accounts for the skills underlying test performance. In situations where there are many unclassified students, the components of the proposed cognitive models (e.g., the list of attributes and the Q-matrix) would need to be revisited. By checking the item-response patterns of unclassified students, researchers conducting the analysis should be able to find new attributes from which they can create a new cognitive model for the test.

In general, the Chinese Taipei students performed well on all cognitive attributes other than recognize patterns, and their excellent performance was evident in the distributions of the 12 clustered knowledge states. The students also showed some weaknesses on thinking skills (inductive and deductive reasoning) and algebra content, compared with the other mastery attributes. These findings suggest that the teaching practices associated with mathematics in classrooms in Chinese Taipei middle schools may be having an effect. As in other Asian countries, such as Japan, South Korea, Singapore, and Hong Kong (SAR), Chinese Taipei students in middle and high schools study within a highly academic and peer-competitive learning environment. The foremost task of middle and high school teachers in Chinese Taipei is to help students acquire high scores on entrance examinations so that students can gain admission from high schools into colleges and universities. Hence, much instruction is examination-oriented. Teachers in these educational environments tend to emphasize memorization and repeated practice as important elements in students' learning, including mathematics learning (Leung, 2001; Liu, 1986).

Mathematics teachers in Chinese Taipei have sufficient and diverse mathematics knowledge because they have to pass the rigorous college entrance examination and take mathematics-related courses amounting to 80 credits in their colleges (Saul, 2000). As such, they are generally knowledgeable about the topics they teach and are able to provide students with diverse ways to solve mathematical questions. However, the lecture approach is still a prevalent pedagogy in mathematics classrooms in Chinese Taipei. Under this approach, students seldom have opportunity to ask questions. Their major task is to listen carefully to what the mathematics teachers teach, and to show their motivation by concentrating on mathematics learning through practice and repetition (Saul, 2000). Mathematics teachers also frequently assign homework to help students apply mathematics concepts and become more familiar with content knowledge. After class, most students go to enrichment programs or have private tutors to review what they have learned in schools and to teach them effective test-taking strategies. At home, students work hard on homework and repeatedly practice using supplemental materials. Given these instructional techniques, it is not surprising that Chinese Taipei students perform well not only on TIMSS but also on other studies of international mathematics achievement.

However, an overemphasis on mathematics examinations can result in the pursuit of correct answers rather than in a true understanding of mathematics. Students may use the right solutions to solve the test items, but they may or may not know why the solutions are appropriate for the questions. The current mathematics learning processes appear to lead to Chinese Taipei students not receiving instruction that encourages individual thinking skills. Instead, they are encouraged to spend time becoming familiar with test item solutions as taught by their mathematics teachers rather than developing their own solutions. These teaching practices may, in part, explain why Chinese Taipei students are somewhat weaker in mathematical thinking and reasoning skills relative to their performance on the other mastery cognitive attributes.

The hypothesized relationship between instructional style and skills mastery can be further highlighted by considering earlier research from other countries. For example, RSM results for Japan, another country that performed extremely well on TIMSS 1999, show that the Japanese students achieved this result because they had excellent high-level thinking skills. In particular, Japanese students performed well on logical reasoning (P5), solution search (P6), judgmental application (P3), data management (P9), and recognize patterns (S6) (Tatsuoka, Corter, & Tatsuoka, 2004). Tatsuoka and her colleagues hypothesized that this result may be a product of the mathematics teaching practices evident in Japanese classrooms. Their review of Kawanaka and

Stigler's (1999) study regarding TIMSS video data of classroom practices showed that Japanese teachers encourage students to develop divergent solutions to problems instead of simply lecturing them on mathematical content knowledge and explaining how to solve problems in classrooms. In other words, teachers in Japan emphasize developing students' mathematical thinking skills rather than having students simply solve problems. It seems, then, that instructional strategies that develop particular cognitive skills advantage students' mathematics achievement. Accordingly, Chinese Taipei mathematics teachers may want to consider providing students with more time to develop their thinking regarding mathematical questions and content.

Singapore provides another example in support of this claim. Students in this country also gained high mean scores on the TIMSS 1999 mathematics test. Moreover, Singapore has an educational context similar to that of Chinese Taipei. Birenbaum, Tatsuoka, and Xin (2005) conducted a study to compare the mathematics performance of students in the United States, Singapore, and Israel. They explored various aspects of the educational context of Singapore and concluded that Singapore has a strong examination culture. Parents support their children's education, particularly in terms of encouraging them to prepare for examinations. Singapore's teachers also focus on helping students prepare for the frequent examinations, and they set time aside for instruction directed at increasing students' test-taking abilities. The teachers also tend to emphasize rote memorization. Children spend a lot of time after class studying mathematics. As Tatsuoka, Corter, & Tatsuoka (2004) found, Singapore's high scores on mathematics resulted from students' excellence in computational and reading skills rather than from their high-level thinking skills. This finding aligns strongly with the findings of the current study. This consistent relationship between the type of instruction and skill mastery is useful and promising because it provides evidence of how instructional strategies affect students' learning in terms of the type of cognitive skills those strategies encourage.

In summary, within Chinese Taipei, strong emphases on memorizing lessons and repeated practice appear to lead to overall high achievement in mathematics. However, this approach may limit development of specific high-level thinking skills among students. Although Chinese Taipei students achieve excellence with respect to total mathematics scores as well as many cognitive attributes related to mathematics, the Chinese Taipei government could consider ways to improve relative weaknesses in the mathematics curriculum, in pedagogical approaches, and in mathematics teacher education. Models such as those employed by the Singaporean educational authorities that promote, despite students' existing high levels of achievement, the need for students to continue acquiring the competencies underpinning a successful economy (Birenbaum et al., 2005) could provide a template for modifications to Chinese Taipei mathematics instruction.

SUGGESTIONS AND RECOMMENDATIONS

The findings from this study bring forth several suggestions for further research relating to cognitive diagnostic assessment. One suggestion relates directly to the rule-space method in terms of the cut-off point of probability for attribute mastery. Different cut-off values affect the outputs of rule-space analysis, such as the hierarchical relationships among the knowledge states. In order to make these hierarchical relations most reliable, it is very important to determine an optimal cut-off score. How this might be done within the context of rule-space analyses is a crucial and interesting topic for future research.

Alternative methods of analyzing the information gained from RSM may improve our understanding of student proficiencies and skills. By using the new attribute probability dataset, researchers should be able to conduct statistical analyses that make rule-space results more interpretable and meaningful. For example, they could apply factor analysis to explore attributes with common underlying factors. The results of such an analysis should allow precise definition of the strengths and weaknesses of student performance on mathematics tests as well as comparisons of groups of students. These analyses could also be improved through use of multilevel models and/or hierarchical linear models. These complex models could allow for analysis of cognitive attributes alongside consideration of educational context variables, such as teaching strategies, teacher characteristics, and school context variables (see, for example Xin, Xu, & Tatsuoka, 2004, in this regard). Additional studies utilizing a multilevel approach may provide useful information for researchers and practitioners interested in the contextual and environmental factors that affect students' mathematics achievement.

Overall, RSM can be used to provide diagnostic information relative to cognitive skills and thereby enrich interpretations of test scores by and for students, teachers, and administrators. Ultimately, continued development of RSM and statistical methods for analyzing the resulting information should help improve the interpretability of assessment results and lead to more widespread application of RSM in the future.

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.

Birenbaum, M., Tatsuoka, C., & Xin, T. (2005). Large-scale diagnostic assessment: Comparison of eighth graders' mathematics performance in the United States, Singapore and Israel. *Assessment in Education, 12*, 167–181.

Chen, Y.-H., Gorin, J. S., Thompson, M. S., & Tatsuoka, K. K. (2008). Cross-cultural validity of the TIMSS-1999 mathematics test: Verification of a cognitive model. *The International Journal of Testing*, *8*, 251–271.

Corter, J. E., & Tatsuoka, K. K. (2002). *Cognitive and measurement foundations of diagnostic assessments in mathematics* (College Board technical report). New York: Teachers College, Columbia University.

Embretson, S. E. (1984). A general latent trait model for response processes. *Psychometrika, 49*, 175–186.

Embretson, S. E. (Ed.). (1985). *Test design: Developments in psychology and psychometrics* (pp. 195–218). Orlando, FL: Academic Press.

Embretson, S. E. (1993). Psychometric models for learning and cognitive processes. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 125–150). Hillsdale, NJ: Lawrence Erlbaum Associates.

Embretson, S. E., & Gorin, J. S. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement, 38*, 343–368.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates.

Everson, H. T., Guerrero, A., & Yamada, T. (2003). *Understanding group differences in mathematical knowledge states: A rule-space analysis of the black-white correlates of mathematical reasoning*. Paper presented at the annual meeting of the National Council of Measurement in Education (NCME), Chicago, Illinois.

Fischer, G. H. (1973). Linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*, 359–374.

Gonzalez, E. J., & Miles, J. A. (2001). *TIMSS 1999 user guide for the international database: IEA's repeat of the Third International Mathematics and Science Study at the eighth grade.* Chestnut Hill, MA: Boston College.

Gorin, J. S. (2002). *Cognitive and psychometric modeling of text-based reading-comprehension GRE-V items*. Unpublished doctoral dissertation, University of Kansas.

Herman, J. J. (1991). Research in cognition and learning: Implications for achievement testing practice. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 154–165). Englewood Cliffs, NJ: Prentice Hall.

Kawanaka, T., & Stigler, J. W. (1999). Teachers' use of questions in eighth-grade mathematics classrooms in Germany, Japan, and the United States. *Mathematical Thinking and Learning, 1*, 255–278.

Leung, F. K. S. (2001). In search of an East Asian identity in mathematics education. *Educational Studies in Mathematics, 47*, 35–51.

Linn, R. L. (1989). *Has item response theory increased the validity of achievement test scores?* (CSE technical report 302). Los Angeles, CA: UCLA Center for Research on Evaluation, Standards, and Student Testing.

Liu, I. M. (1986). Chinese cognition. In M. N. Bond (Ed.), *The psychology of the Chinese people* (pp. 73–105). Hong Kong: Oxford University Press.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika, 60*, 439–483.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden R. A., O'Connor, K. M., Chrostowski, S. J., & Smith, T. A. (2000). *TIMSS 1999 international mathematics report: Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.

Saul, M. (2000). Reifying the research: Mathematics education in Taiwan. *Notices of the AMS, 47*, 360–363.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (pp. 263–331). New York: Macmillan.

Stout, W. (2002). Psychometrics: From practice to theory and back. *Psychometrika, 67*, 485–518.

Tatsuoka, C., Varadi, F., & Tatsuoka, K. K. (1992). *BUGSHELL* (computer software). Ewing, NJ: Tanar Software.

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. Grederiksen, R. Glazer, A. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453–488). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tatsuoka, K. K. (1991). *Boolean algebra applied to determination of the universal set of knowledge states* (technical report ONR-91-1). Princeton, NJ: Educational Testing Service.

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: Lawrence Erlbaum Associates.

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

Tatsuoka, K. K., & Boodoo, G. (2000). Subgroup differences on the GRE quantitative test based on the underlying cognitive processes and knowledge. In A. E. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 821–857). Hillsdale, NJ: Erlbaum.

Tatsuoka, K. K., Corter, J. E., & Guerrero, A. (2004). *Coding manual for identifying involvement of content, skill, and process subskills for the TIMSS-R 8th grade and 12th grade general mathematics test items* (technical report). New York: Department of Human Development, Teachers College, Columbia University.

Tatsuoka, K. 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. K., Guerrero, A., Corter, J. E., Yamada, T., & Tatsuoka, C. (2003). *International comparisons of mathematical thinking skills in the TIMSS-R*. Paper presented at the annual meeting of the National Council of Measurement in Education (NCME), Chicago, Illinois.

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

Tatsuoka, K. K., Xin, T., Corter, J. E., & Tatsuoka, C. (2004). *Patterns of attribute characteristic curves and their relationships to teachers' background variables*. Paper presented at the annual meeting of the National Council of Measurement in Education (NCME), San Diego, California.

Xin, T., Xu, Z., & Tatsuoka, K. K. (2004). Linkage between teacher quality, student achievement, and cognitive skills: A rule-space model. *Studies in Educational Evaluation, 30*, 205–223.