Abstract

The traditional way for research concerning group differences on mathematics performance is highly dependent on raw scores (total scores) or standardized test scores of mathematical topics. Therefore, traditional comparisons of group means merely identify group differences, but provide no substantive description of the nature of the differences. Rule-space methodology (RSM), as an alternative to total or subscale scores based on sums of item responses, allows for group comparisons made at the cognitive attribute level, which is made up of cognitive skills, knowledge, and strategies that an individual can employ to solve a problem. In the current study we aimed to apply RSM to explore group differences of Taiwanese mathematics performance on TIMSS-1999 from a micro-skills perspective.

The results indicated that Recognize pattern (S6) and Rule application in algebra (P4) had the largest difference in mastery probabilities between highest- and lowest-achieving students. Gender differences of Taiwanese students in terms of mathematical skills are quite minimal. Students in urban schools had significantly higher raw scores than students in rural schools. And consistent findings in terms of cognitive attributes indicated that urban school students had higher attribute mastery probabilities than rural school students for 22 out of 23 attributes.

Keywords: rule-space methodology, group comparisons, mathematics assessment
Subgroup Comparisons of Taiwanese Mathematics Performance:

From a Perspective of Cognitive Attributes

The traditional way for research concerning group differences on mathematics performance is highly dependent on raw scores (total scores) or standardized test scores of mathematical topics. This could be called a single-score-based paradigm (Stout, 2002). For instance, in the Trends in International Mathematics and Science Study in 1999 (TIMSS-R), there were four forms of achievement scores calculated and reported in their database, including raw scores, standardized raw scores, national Rasch scores, and plausible values. Even international benchmarks of achievement were established along with plausible values (Gonzalez & Miles, 2001). Most of so-called diagnostic scores are the subscores of cognitive domains, such as the scores for four cognitive domains provided in the TIMSS-2003 study, including knowing factors and procedures, using concepts, solving routine problems, and reasoning (Mullis, Martin, & Foy, 2005). Each item in the test of TIMSS-2003 belongs to only one cognitive domain and sum of item scores that belong to the same cognitive domain is the subscore of a particular cognitive domain. Based on a single-score approach along with these various types of single scores, numerous studies of student mathematics performance among different groups, such as gender, rural and urban, and even cross-countries, were conducted.

As an alternative to total or subscale scores based on sums of item responses, rule-space methodology (RSM; Tatsuoka, 1995) allows for group comparisons made at the cognitive attribute level, which is made up of cognitive skills, knowledge, and strategies that an individual can employ to solve a problem. RSM has potential for comparing and contrasting subgroup differences (Everson, Guerrero, & Yamada, 2003; Tatsuoka & Boodoo, 2000). Traditional comparisons of subgroup means or item difficulty parameters merely identify group differences,
but provide no substantive description of the nature of the differences. Knowledge states obtained by RSM allow comparisons to be made across groups in terms of the types of skills mastered by percentages of students who have mastered or not mastered certain skills. Such insight into differences in subgroup performance has direct implications for instructional design and policy targeted at improving learning of underperforming students.

**Rule-Space Methodology**

Rule-space methodology is a cognitive-psychometric approach that incorporate cognitive information into psychometric models. Hence, RSM can provides an individual diagnostic profile regarding the current knowledge state of mastery or non-mastery of attributes that underlie performance on the test. RSM is comprised of two main phases. The first is *determination of latent knowledge states*. Two steps are required in the first phase: 1) identify unobservable cognitive attributes in a domain of interest and create the incidence matrix (Q-matrix); 2) determine latent knowledge states (classification groups or the ideal item-response patterns). The second phase is *classification of examinees’ knowledge states*. In the second phase, two additional steps are performed: 3) to map the observed response patterns into the ideal item-response patterns; and 4) to classify an examinee’s responses into one of the closest knowledge states (Tatsuoka, 1995; Tatsuoka & Boodoo, 2000).

**Research Question**

We aimed to apply rule-space methodology to explore group differences of Taiwanese mathematics performance on TIMSS-1999 from a micro-skills perspective. Group comparisons we focused on were among students of different performance levels, between male and female students, and between rural and urban students. The more specific research questions in the current study were presented as follows:
1) Are there group differences in terms of cognitive attributes among Taiwanese students at different performance levels on the TIMSS-1999 items?

2. Are there group differences between male and female as well as between rural and urban Taiwanese students’ performances on the TIMSS-1999 items in terms of cognitive attributes?

**Methods**

**Participants**

There were a total of 5772 Taiwanese students nested within 150 schools participating in the TIMSS-1999 study. Schools were sampled by four implicit strata: north (2424 students, 42%), middle (1512 students, 26%), south (1662 students, 29%), and east (174 students, 3%) of Taiwan (Chang, Hung, & Lo, 2004). In each classroom, eight different booklets of TIMSS-R mathematics tests were assigned randomly to students. Only Booklets 1, 3, 5, and 7 were used for the current study, including 2874 students distributed in 150 classrooms (Corter & Tatsuoka, 2003). In these 150 schools, there were 4 schools (2.7%) located in a geographically isolated area, 15 schools (10%) in village or rural areas, 58 schools (38.7%) in suburban areas, and 73 schools (48.7%) in urban settings.

**Data Management**

For the purpose of comparisons across subgroups, these data were divided into nine different performance subgroups, male and female subgroups as well as rural and urban subgroups. The different performance subgroups were based on the first plausible values, which are scores that can be compared across booklets with a mean of 500 and a standard deviation of 100. The nine different performance groups were categorized as follows: 50-150, 150 to 250, 250 to 350, 350-450, 450 to 550, 550 to 650, 650 to 750, 750 to 850, and 850 to 950. These
performance groups were labeled 100 (3 students, 0.1%), 200 (8 students, 0.3%), 300 (44 students, 1.5%), 400 (243 students, 8.5%), 500 (600 students, 21.0%), 600 (1167 students, 40.8%), 700 (710 students, 24.8%), 800 (84 students, 2.9%), and 900 (4 students, 0.1%). The sample was 50.6% male (1453 students) and 49.4% female (1421 students). As for rural and urban subgroups, schools located in a geographically isolated area and in village or rural areas were categorized into rural schools; while schools located in suburban areas and in urban settings were categorized into urban schools. Since rural schools were sampled only in north and middle of Taiwan, the sample used to make comparisons of rural and urban schools was derived from these two locations. There were 345 students (17.3%) in rural schools and 1649 students (82.7%) in urban schools.

Instrument

There were five content categories included in the TIMSS-R mathematics tests: a) fractions and number sense (38%); b) measurement (15%); c) data representation, analysis, probability (13%); d) geometry (13%); and e) algebra (22%). Item types involved multiple-choice (77%), short answer (13%), and extended response formats (10%) (Gonzalez & Miles, 2001). There were eight different test booklets were designed from a 162-item pool in the TIMSS-R study. Each student was requested to answer only one out of eight booklets and only Booklets 1, 3, 5, and 7 were used for the current study. These booklets were selected based on the criterion that each attribute to be analyzed in the study had to be included in at least three items (Corter & Tatsuoka, 2003).

Analysis

Two questions were as follows: a) Are there group differences in terms of cognitive attributes among students at different performance levels on the TIMSS-1999 items? a) Are there
group differences between male and female as well as between rural and urban students’ performances on the TIMSS-1999 items in terms of cognitive attributes? In order to compare different groups, the single dataset including all examinees’ vectors of attribute mastery probabilities was separated into the desired groups.

Comparisons of Different Performance Levels. To compare students in different performance levels, the dataset was stratified into nine groups based on the first plausible values. The patterns across different performance groups on three types of cognitive attributes can be examined. The graph for each attribute with different performance levels on the horizontal axis and mastery probability on the vertical axis is termed the Attribute Characteristic Curve (ACC). Attribute characteristic curves describe a relationship between attribute mastery probability and performance levels, like an item characteristic curve (ICC) in IRT models (Tatsuoka, Xin, Corter, & Tatsuoka, 2004). In the typical case, the probability will be high for examinees with high performance levels and low for examinees with low performance levels. By looking at these attribute curves, group differences in terms of cognitive attribute mastery among students at different performance levels on the TIMSS-1999 items were uncovered.

Comparisons of Different Groups. Group comparisons were made between male and female students as well as between rural and urban students. For the gender comparisons, the dataset was separated into male and female student groups. For the comparisons of rural and urban students, the dataset was divided into rural and urban student groups. The following comparisons were conducted for these two kinds of groups. First, a mastery vector of each attribute probability was computed for male and female groups as well as rural and urban groups. The graph was drawn with attributes on the horizontal axis and mastery probability on the vertical axis. The patterns of different groups were plotted on separate lines and then compared.
Second, attribute characteristic curves (ACC) for male and female examinees as well as for rural and urban examinees were compared. Each ACC graph was shown to compare whether or not different group students have different patterns along different performance levels. Third, the percentage of students in each clustered knowledge state was computed for compared groups. The first to third highest percentages of clustered knowledge states were highlighted to compare performance differences in terms of attributes mastery/non-mastery.

Results and Conclusions

The results in the current study indicated that Recognize pattern (S6) and Rule application in algebra (P4) had the largest difference in mastery probabilities (.70 and .67). Other attributes that had larger difference probabilities included Open-ended items (S10), Elementary algebra (C3), and Geometry (C4). These attributes may suggest that Taiwanese students between low and high achievement groups have larger learning differences in abstract thinking skills (S6 and S10), high-level content areas of Algebra and Geometry (C3 and C4), and skills of applying sophisticated algebra (P4). Similar results were found in Corter and Tatsuoka’s (2003) study for the American sample. They found that these same five attributes had large mastery differences for American students at various levels of achievement, in addition to content area of Fraction (C2).

Results regarding gender comparisons in this study showed that only 2 out of 23 attributes, Evaluate and verify options (S5) and Logical reasoning (P5), were significantly different between males and females at the .05 level. Female students had higher mean mastery probability than male students on Evaluate and verify options; whereas male students had higher probability on Logical reasoning. The finding indicates that gender differences of Taiwanese students in terms of mathematical skills are quite minimal. Male and female middle school
students in Taiwan show comparatively equivalent potential in learning mathematical skills. Concerning group comparisons between urban and rural school students in the current study, students in urban schools had significantly higher raw scores than students in rural schools. Consistent findings in terms of cognitive attributes indicated that urban school students had higher attribute mastery probabilities than rural school students for 22 out of 23 attributes, nine of which were statistically significant at an alpha level of .01. These nine attributes included high-level mathematics contents, such as Algebra (C3) and Geometry (C4), and high-level abstract thinking and reasoning skills, such as Proportional reasoning (S7), Logical reasoning (P5), Solution search (P6), and Open-ended items (S10). This finding highlights that students in urban schools perform better than those in rural schools on high-level mathematics contents and abstract thinking skills.

Educational Significance

What significant impact the results of this study can have is that examination of performance-group, gender, and region differences along with substantive description of the nature of the differences might help identify equity issues in the Taiwanese education system much more appropriately. Given that RSM has thus far been limited in use, this study is considered both as an opportunity to verify the feasibility of the method for group comparisons in terms of diagnostic information, as well as a chance to answer interesting questions about group differences of Taiwan’s mathematics performance.

The complete analyses have been conducted and the more detailed literature review, results, conclusions, and suggestions will be presented in the paper.
Reference


