Verification of Cognitive Attributes Required to Solve the TIMSS-1999 Mathematics Items for Taiwanese Students

Yi-Hsin Chen, Joanna Gorin, and Marilyn Thompson

Arizona State University

Kikumi Tatsuoka

Columbia University

Presented at the Annual Conference of American Educational Research Association

April 2006, San Francisco, CA

Please address correspondence concerning this manuscript to:

Yi-Hsin Chen

Division of Psychology in Education

Box 870611

Arizona State University

Tempe, Arizona 85287-0611

Abstract

Educational assessment is a process of collecting evidence and interpreting it to provide instructors with information regarding students' learning. However, the current design and scoring of most standardized educational tests are insufficient to serve this purpose. The limitation exists primarily due to the lack of cognitive information incorporated into traditional psychometric models. To date, some psychometricians have applied cognitive psychology principals to psychometric models of educational assessment data. Of these models, one of the most extensively researched and empirically supported is Tatsuoka's rule-space methodology (RSM). RSM can be used to validate a proposed cognitive model.

A list of cognitive attributes, which describe what knowledge, strategies, and processing skills the TIMSS-R mathematics test measures, was developed with RSM by using the U.S. sample. Research using these cognitive attributes of the TIMSS items appears to adequately describe student performance for several other countries. In our current study, we intended to explore how a list of cognitive attributes expresses student performance in Taiwan.

Three analyses, including calculation of classification rates, multiple regression analyses, and comparisons of attribute mastery probabilities across four booklets, were conducted. Successful classification rates ranged from 99.3% to 99.9% for four booklets, and R^2 and adjusted R^2 for each booklet were estimated using multiple regression analyses and ranged from .943 to .979. All ranges of mean attribute probabilities across booklets were less than .25, except for the Recognize patterns attribute (S6). Generally speaking, a list of cognitive attributes and the incidence matrix used as a proposed cognitive model in the current study represent the performance of Taiwanese eighth graders on the TIMSS-1999 mathematics tests very well.

Verification of Cognitive Attributes Required to Solve the TIMSS-1999 Mathematics Items for Taiwanese Students

Educational assessment is a process of collecting evidence and interpreting it to provide instructors with information regarding students' learning (Glaser, 1962). Instructors can use this information to identify what knowledge students already have prior to instruction, diagnose learning errors or misconceptions during instruction, and detect learning effects and outcomes after instruction. Thus, assessment ought to play a critical role in the instructional process. However, the current design and scoring of most standardized educational tests are insufficient to serve this purpose. More often, test scores are used to scale students in terms of relative ability to one another, or in some circumstances, a criterion level of mastery.

Since the use of tests is a popular way to assess students' learning, considerable debate has been made as to whether or not the tests measure what they are intended to measure. This is regarded as an issue of construct validity (Cronbach & Meehl, 1955). Given that the majority of standardized tests based on traditional psychometric models (eg., CTST and IRT) report students' learning as a single estimate, evidence of construct validity of test scores typically consists of correlations between test scores and other measures. Little information is gathered that is more directly concerned with the theoretical mechanisms underlying successful test performance. The limitation exists primarily due to the lack of cognitive information incorporated into traditional psychometric models (Herman, 1991; Snow & Lohman, 1989).

In order to overcome the limitations of traditional test scores, pyschometricians have explored the use of a new family of statistical models that incorporate more substantive theory into the data modeling procedure. These models, often called *cognitively diagnostic models*, are thought to be particularly useful for examining the cognitive structure of student responses for several purposes. First, a more complete understanding of student responses will increase our confidence in the valid use and interpretation of test scores. Second, detailed descriptions of student processes underlying test scores can help to supplement the feedback given to students and teachers that could be useful for improving student performance and learning.

To date, some psychometricians have applied cognitive psychology principals to psychometric models of educational assessment data, such as Embretson's cognitive design system (CDS; 1994, 1998), Mislevy's evidence centered design (ECD; 1994), Bennet and Bejar's integrated generative approach (1998), and Tatsuoka's rule space methodology (RSM; 1985, 1995). The commonality among these approaches is the auxiliary cognitive information added in the use of statistical models (Embretson & Reise, 2000). Of these models, one of the most extensively researched and empirically supported is Tatsuoka's rule-space methodology. RSM can be used to validate a proposed cognitive model, which is made up of cognitive skills, knowledge, and strategies that an individual can employ to solve a problem. If the attributes in a model can accurately and reliably describe student performance, then evidence is provided supporting the model and its components. Researchers can then examine the extent to which the validated model matches the intended measurement purpose of the test.

One area to which RSM has been applied is mathematics achievement performance in TIMSS-1999. Trends in International Mathematics and Science Study (TIMSS), which was developed by the International Association for the Evaluation of Educational Achievement (IEA), was designed to measure trends in students' mathematics and science achievement. This study is a long-term project for which data were collected in 1995, 1999, and 2003. The TIMSS-1999, also known as TIMSS-Repeated or TIMSS-R, was a repeat of the international achievement study conducted in 1995.

A list of cognitive attributes, which describe what knowledge, strategies, and processing skills the TIMSS-R mathematics test measures, was developed with RSM by using the U.S. sample (Coter & Tatsuoka, 2002). Research using these cognitive attributes of the TIMSS items appears to adequately describe student performance for several other countries (Birenbaum, Tatsuoka, & Yamada, 2004; Tatsuoka, Corter, & Guerrero, 2004). Taiwan was for the first time included in this project in 1999 and had the third ranking place in terms of performance on the TIMSS-R mathematics test. Thus, it is of interest to explore how a list of cognitive attributes expresses student performance in the high performing countries, such as Taiwan.

Based on the aforementioned reasons, the purpose of this study aims to provide evidence for construct validity of the TIMSS-1999 mathematics test for Taiwanese eighth graders. This will involve verification of a set of cognitive attributes including component-processes, strategies, and knowledge structures students must possess to correctly solve items of the TIMSS-1999 mathematics test. That is, the current study will validate the nature of the TIMSS-1999 mathematics test for the Taiwanese student population. Specifically, we investigate whether a set of previously identified cognitive attributes represents the performance of Taiwanese eighth graders on the TIMSS-1999 mathematics tests.

Rule-Space Methodology

Rule Space Methodology (RSM) is a cognitive-psychometric model of test data that provides estimates of student proficiency in terms of individuals' mastered and non-mastered skills. RSM results can be a powerful tool for substantive examinations of construct validity. Through the parameterization of item statistics in terms of *cognitive attributes* underlying item solutions, RSM can be used to verify the relationship between item responses and hypothesized construct definition. In other words, RSM allows for a formal hypothesis test of a cognitive model of test items in terms of cognitive skills, knowledge, and strategies that an individual can employ to solve a problem. We briefly review the four steps in rule-space methodology: identification, determination, mapping and classification.

First, the identification step identifies the attributes for the test items of interest. Item attributes may include knowledge, strategies, and processing skills that are required to answer the items correctly (Birenbaum, Kelly, & Tatsuoka, 1993). In this study, cognitive attributes for the TIMSS-1999 mathematics items include three categories: content attributes, process attributes, and skill/item-type attributes (Corter & Tatsuoka, 2002). Once a set of cognitive attributes is identified, an incidence matrix (Q-matrix; Tatsuoka, 1983) is constructed based on these cognitive attributes. The incidence matrix represents the relationships between items and attributes by assigning 1s to those attributes that are related to solving a particular item and 0s to those that are not related (Birenbaum & Tatsuoka, 1993; Tatsuoka, 1995).

Second, the goal of the determination step is to determine ideal item-response patterns. The rule of Boolean Descriptive Function (BDF; Tatsuoka, 1991) is applied to connect the latent attribute mastery pattern (so-called latent knowledge state) to the observable ideal item-response patterns in this step. The assumption behind Boolean Descriptive Function is that an item can be answered correctly if and only if the attributes involved in this item have been mastered. Based on the attributes involved in an item (indicated by Q-matrix) and applying BDF, each latent attribute mastery pattern can correspond to the observable item-response pattern. These logically interpretable item-response patterns are called ideal item-response patterns.

However, students' observed item-response patterns are often different from the ideal item-response patterns due to using a rule inconsistently. So the third step, mapping, is conducted to map these two patterns onto a classification space called rule space. A Cartesian Coordinate System is utilized to formulate a two-dimensional classification space, which consists of θ (IRT ability parameter) along the X-axis and ζ (the unusualness of item response pattern) along the Y-axis. The classification space could be a multiple-dimensional space by adding the generalized ζs (the unusualness of item response pattern for the subset of items) (Tatsuoka, 1996). Generalized ζs can help form better classification groups. In this step, the estimates of θ, ζ, or generalized ζs were obtained.

Fourth, the classification step aims to classify an examinee based on the examinee's item-response pattern into one or more of the predetermined latent knowledge states. Mahalanobis distances (D^2) between the examinees' item-response patterns and the ideal item-response pattern were calculated in order to decide which classification group examinees belong to. The Bayesian Decision rule for minimum error was utilized to give each examinee the attribute mastery profile with the highest posterior probability.

Methods

Participants

The data for the current study were previously collected in the 1999 administration of TIMSS and only the Taiwanese data were utilized. Based on a two-stage sample design used by TIMSS-R, schools were sampled in the first stage and each intact classroom within the sampled schools was selected in the second stage (Foy & Joncas, 2000). There were a total of 5772 students nested within 150 schools in the Taiwan sample. 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.

Instrument

The TIMSS-1999 mathematics tests contained a pool of 162 items, including five content categories: 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).

A 162-item pool was designed and used to compose eight different test booklets in the TIMSS-R study. Each student was requested to answer only one out of eight booklets, each taking 90 minutes to complete. Although there were eight booklets in TIMSS-R mathematics tests, 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, 2002).

Analysis

Item parameter estimations, rule-space analyses, multiple regression analyses, classification rates, and comparisons of attribute probabilities of four booklets were used to validate the attributes and the incidence matrix.

Estimating item parameters. The examinee ability parameters (θ), the item difficulty parameters (b), and the item discrimination parameters (a) were estimated with BILOG-MG (Zimowski et al., 1996). BILOG-MG analyses proceeded separately for each booklet.

Conducting rule-space analysis. Special computer software called BUGSHELL, programmed by Tatsuoka, Varadi, and Tatsuoka (1992), was utilized for the rule-space analysis. Three-dimensional Cartesian coordinate space was used to formulate the classification space, which consists of the IRT ability (θ), ζ, and generalized ζ. Several relevant parameters for rule space analyses were set in advance. First, the acceptable Mahalanobis distance and the difference of θ values between an examinee's and ideal item-response patterns were set to 4.5 and 1.5, respectively. Second, the number of slips, which is the number of the different responses between the observed and ideal item patterns in the test, was not more than one-third of the total items (Corter & Tatsuoka, 2002). Finally, a mastery probability of each attribute was computed to form an attribute vector for each examinee by performing the four separate rule-space analyses for four booklets.

Computing classification rate. Classification rate in the RSM study was an alternative approach used to validate the attributes and the incidence matrix. Classification rate is the proportion of examinees who are classified successfully into at least one of the predetermined knowledge groups. If the classification rate is low, this suggests that many examinees' item response patterns are inconsistent with the predetermined latent knowledge states. A high classification rate occurs when the extent of reflection of the identified attributes and the incidence matrix to performance on the TIMSS-1999 mathematics test is high.

Multiple regression analyses. Several multiple regression analyses were conducted for each booklet. An examinee ability parameter, such as a raw score, the first plausible value on the TIMSS-R scale, or an IRT ability estimate, was regressed on each student's attribute mastery probability. The *R*-squared and adjusted *R*-squared indices were examined to determine how well the cognitive attributes and the incidence matrix accounted for the item parameter. If low *R*-squareds are obtained, the attributes and the incidence matrix are revisited. Attributes with low predictive contribution are either refined or deleted in the cognitive model.

Comparing descriptive statistics of each attribute across four booklets. The means and standard deviations of the mastery probabilities for all attributes across examinees were computed for each booklet. The consistency of the estimated attribute mastery probabilities was checked across four booklets. If any one attribute has significantly different probabilities across the four booklets, it reflects a problem with respect to the theoretical definition of attributes and/or attribute coding (Corter & Tatsuoka, 2002).

Results

In order to verify the cognitive model of the TIMSS-R math test using the Taiwanese sample, including the attributes and the incidence matrix, three analyses were conducted: 1) calculation of classification rates, 2) multiple regression analyses, and 3) comparisons of attribute mastery probabilities across four booklets. Detailed results are presented below.

Classification Rates

Classification rates for four rule space analyses are shown at the bottom of Tables 1 to 4. Successful classification rates ranged from 99.3% to 99.9%, which were extremely high, for the four booklets. The average classification rate for all students was 99.6%, with only 11 out of 2874 students not assigned to at least one of the predetermined knowledge states. In other words, almost all item response patterns of Taiwanese students on the TIMSS-R mathematics test for Booklets 1, 3, 5, and 7 could be accounted for by the predetermined knowledge states derived from the proposed attributes and the incidence matrix.

Multiple Regression Analyses

To further check the adequacy of the proposed cognitive model using the Taiwanese sample, multiple regression analyses were performed with total scores regressed on students' attribute mastery probabilities that were derived from rule space analyses. Four individual regression analyses were conducted for Booklets 1, 3, 5, and 7. Results from four separate analyses are represented in Tables 1 to 4.

In these multiple regression models, R^2 and adjusted R^2 for each Booklet were checked first. Extremely high R^2 and adjusted R^2 were obtained for Booklets 1, 3, 5, and 7, ranging from .943 to .979. For the entire sample, R^2 and adjusted R^2 were also quite high, .925 and .924, respectively. This means that over 92% of the variance in total scores on the TIMSS-R mathematics items can be accounted for by the attribute mastery probabilities obtained from RSM with the proposed cognitive model.

Similar multiple regression analyses were performed on the first plausible values, which were used in international comparisons for the TIMSS achievement study. Although lower than for the total scores, the R^2 and adjusted R^2 for Booklets 1, 3, 5, and 7 were still very high, ranging from .853 to .814. For the entire sample, R^2 and adjusted R^2 were .805 and .804, respectively. That is, attribute mastery probabilities also predicted the first plausible values well.

In order to assess the relationships of individual attributes and the total scores, the regression coefficients and correlation coefficients of each attribute with total scores for Booklets 1, 3, 5, and 7 were examined (refer to Tables 1 to 4). The results indicate that Approximation and Estimation (S4) had a low correlation coefficient (.07) with total scores in Booklet 3 and negative correlation coefficients (-.15 and –.06) in Booklets 1 and 5. Elementary Algebra (C3) was excluded in the regression analysis for Booklet 5 due to 0 tolerance, which means that Elementary Algebra (C3) independently contributes no information independent of other attributes in the model. Except for these two attributes, correlations between attributes and the total scores were positive and strong in value.

Consistency of Attribute Probabilities across Four Booklets

Given that the same list of attributes was measured across the four rule space analyses for Booklets 1, 3, 5, and 7, the consistency of attribute mastery probabilities across four booklets was checked to provide further evidence of the adequacy of the cognitive model. The means and standard deviations of each individual attribute for four booklets are presented in Table 5. The results in Table 5 showed that the largest difference in range of mean attribute probabilities across booklets was .27 for Recognize pattern (S6), with the lowest mean probability of .47 occurring in Booklet 3. The ranges for Number sense (S2) and Logical reasoning (P5) were .23 and .20, respectively. Except for these three attributes, the ranges of mean attribute probabilities across booklets were less than .20. Thirteen out of 23 attributes had probability difference ranges less than .10.

Table 6 displays frequencies of attributes required for items in the four booklets and ranges of mean probability differences adopted from Table 5. Recognize patterns (S6) had the highest probability difference range. And compared with the other attributes, Recognize patterns (S6) was required in the fewest total items (14 items) across the four booklets and was included in only one item for Booklet 3, where it had the lowest mastery probability among attributes of .47 (see Table 6).

Discussion

The proposed cognitive model in the current study consists of a list of cognitive attributes that are required to answer the TIMSS-R test items correctly and an incidence matrix that describes the relationship of the cognitive attributes with items on the TIMSS-R test. Three different kinds of evidence were used to verify the proposed cognitive model, including classification rates, multiple regression analyses, and comparisons of attribute mastery probabilities across four booklets.

Classification rates in the current study were extremely high, and few students were not assigned to at least one or more of the predetermined knowledge states. The results of the high classification rates in the current study can be considered as positive evidence that the cognitive model used here, including the list of attributes and the incidence matrix, explains the Taiwanese students performance on the TIMSS-R mathematics test very validly. That is, the latent knowledge states derived from the proposed cognitive model used in the current study do reflect the actual examinee performance, represented by examinee item response patterns, on the TIMSS-1999 mathematics test.

The four individual multiple regression analyses were conducted to check the adequacy of the proposed cognitive model. R^2 and adjusted R^2 for each booklet were checked to examinee what extent of the variances in total scores and in plausible values can be accounted for by the attribute mastery probabilities. Regression analyses on both total scores and plausible values yielded extremely high R^2 and adjusted R^2 for the four booklets as well as for the entire sample. Similar regression analyses with the first plausible values also obtained high R^2 and adjusted R^2. Overall, these findings suggest that the proposed cognitive model for the TIMSS-R mathematics test explained the Taiwanese students achievement performances well.

Furthermore, the regression coefficients and correlation coefficients of each attribute with total scores were examined in order to assess the relationships of individual attributes and the total scores. The results show that almost all attribute probabilities among the four booklets have positive and strong correlations with the total score. The examination of regression and correlation coefficients for each attribute suggests that the list of attributes and the incidence matrix used in the current study were stable in predicting the students' performance on the TIMSS-R mathematics test.

Regarding consistency of attribute probabilities across four booklets, only Attribute S6 (Recognize pattern) has a probability difference range greater than .25 and approximately half of the 23 attributes had probability difference ranges less than .10. These results show that the list of attributes and the coding of relations between items and attributes were quite stable across four booklets for Taiwanese students. A further finding is that Recognize pattern (S6), an attribute which has the largest range, is required in the fewest items and involved in only one item in Booklet 3 with the most extreme mastery probability. This may indicate that unstable attributes showing inconsistency of mean mastery probabilities among booklets are involved in relatively few items, so it might be problematic not to involve enough items for each measured attribute.

Conclusion

Construct validation of standardized educational tests is not a one-step process. Iterations of construct validation using diverse populations are expected. Thus, one of the significant implications of this study is the generalizability of the construct measured by TIMSS items for the Taiwanese sample (generalizability validity). Generally speaking, a list of cognitive attributes and the incidence matrix used as a proposed cognitive model in the current study represent the performance of Taiwanese eighth graders on the TIMSS-1999 mathematics tests very well.

In the future research, two attributes in the list should be checked again. One is Approximation and estimations (S4), an attribute which has low or even possibly negative correlation with the total scores. The other is Recognize patterns (S6), an attribute which has unstable mean mastery probabilities among four booklets. Perhaps creating more items requiring for each measured attribute will yield more stable attribute probability estimates.

References

Bennett, R. E., & Bejar, I. I. (1998). Validity and automated scoring: It's not only the scoring. *Educational Measurement: Issues and Practice, 17,* 9-16.

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, K. K. (1993). Applied an IRT-based cognitive diagnostic model to diagnose students' knowledge states in multiplication and division with exponents. *Applied Measurement in Education*, *6,* 255-268.

Carroll, J. B., & Maxwell, S. (1979). Individual differences in ability. *Annual Review of Psychology, 30,* 603-640.

Corter, J. E., & Tatsuoka, K.K. (2002). *Cognitive and Measurement Foundations of Diagnostic Assessments in Mathematics.* Technical Report, College Board.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, *93*, 179-197.

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*. Hillsdale, NJ:Lawrence Erlbaum Associates.

Embretson S. E. (1994). Application of cognitive design systems to test development. In C. R. Reynolds (Ed.), *Cognitive assessment: A multidisciplinary perspective*). New York: Plenum Press.

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.

Gierl, M. J. (1996). *An investigation of the cognitive foundation underlying the rule-space model.* Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign, IL.

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: TIMSS International Study Center, Boston College.

Herman, J. J. (1991). Research in cognition and learning: Implication for achievement testing practice. In M. C. Wittrock, & E. L. Baker (Eds.), *Testing and Cognition.* Englewood Cliffs, N.J. : Prentice Hall

Mislevy, R. J. (1994). Evidence and inference in education al assessment. *Psychometrika, 59,* 439-483.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), Educational measurement. New York: American Council on Educational, Macmillian.

Tatsuoka, C., Varandi, 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. (1985). A probabilistic model for diagnostic misconceptions in the pattern classification approach. *Journal of Educational Statistics, 10,* 55-73.

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.* 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., 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.

Wittrock, M. C. (1986). Education and recent research on attention and knowledge acquisition. In S. L. Friedman., K. A. Klivingston., & R. W. Peterson (Eds.), *Brain, cognition, and education.* New York: Academic Press.

Wittrock, M. C. (1992). Knowledge acquisition and comprehension. In M. C. Alkin (Ed.), *Encyclopedia of educational research.* New York: Macmillan Publishing Compny.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). *BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items*. Chicago: Scientific Software.

Table 1

Regression Coefficients of Booklet 1 for Total Score Model

Attribute	*B*	SE_B	β	r	$r_{partial}$	r_{part}
C1: Whole numbers and integers	-1.22	1.21	-.01	.39	-.04	-.01
C2: Fractions and decimals	-1.07	.94	-.01	.42	-.04	-.01
C3: Elementary algebra	-3.49**	.83	-.08	.72	-.16	-.03
C4: Two-dimensional geometry	3.77**	.71	.05	.57	.20	.04
C5: Data and basic statistics	.58	1.15	.01	.33	.02	.00
S2: Number sense	4.79**	.57	.08	.54	.31	.06
S3: Figures, tables & graphs	-7.53*	2.73	-.03	.33	-.10	-.02
S4: Approximation & estimation	1.22**	.40	.02	-.15	.11	.02
S5: Evaluate and verify options	.81	1.28	.01	.28	.02	.01
S6: Recognize patterns	6.52**	.49	.14	.67	.45	.10
S7: Proportional reasoning	9.25**	.89	.09	.34	.37	.07
S8: Novel/unfamiliar problems	7.55**	.69	.10	.40	.38	.08
S10: Open-ended items	4.68**	.58	.09	.73	.29	.06
S11: Word problems	2.57	1.69	.02	.38	.06	.01
P1: Translate	1.71	1.12	.02	.43	.06	.01
P2: Computation application	7.57**	.63	.11	.42	.42	.09
P3: Judgmental application	3.97**	1.01	.04	.40	.15	.03
P4: Rule application in algebra	11.88**	.67	.32	.72	.56	.13
P5: Logical reasoning	11.22**	.38	.29	.68	.74	.21
P6: Solution search	3.44**	.64	.06	.68	.20	.04
P7: Visual figures and graphs	8.08**	.65	.12	.52	.42	.09
P9: Data management	9.16**	.65	.13	.58	.47	.10
P10: Quantitative reading	7.19**	.55	.11	.50	.44	.09

R^2 = .965, adjusted R^2 = .964, F(23, 697) = 833.445, p<.01

Classification Rate with D^2 < 4.5 = 99.3%

* p < .05 ** p < .01

Table 2

Regression Coefficients of Booklet 3 for Total Score Model

Attribute	*B*	SE_B	β	r	$r_{partial}$	r_{part}
C1: Whole numbers and integers	1.93*	.93	.03	.48	.08	.02
C2: Fractions and decimals	-2.49	1.36	-.02	.27	-.07	-.02
C3: Elementary algebra	6.05**	.62	.12	.59	.35	.09
C4: Two-dimensional geometry	4.72**	.69	.11	.74	.25	.06
C5: Data and basic statistics	3.98**	.73	.07	.12	.20	.05
S2: Number sense	5.28**	.94	.08	.26	.21	.05
S3: Figures, tables & graphs	8.52**	1.60	.06	.35	.20	.05
S4: Approximation & estimation	3.33**	.65	.06	.07	.19	.05
S5: Evaluate and verify options	-10.29**	1.97	-.07	.30	-.20	-.05
S6: Recognize patterns	.72	.52	.02	.31	.05	.01
S7: Proportional reasoning	9.78**	1.32	.09	.37	.27	.07
S8: Novel/unfamiliar problems	5.51**	1.63	.04	.25	.13	.03
S10: Open-ended items	8.87**	.39	.28	.71	.66	.21
S11: Word problems	4.72*	2.10	.03	.26	.09	.02
P1: Translate	-3.90	2.38	-.03	.30	-.06	-.02
P2: Computation application	-5.68**	1.36	-.06	.40	-.16	-.04
P3: Judgmental application	1.59	1.49	.01	.34	.04	.01
P4: Rule application in algebra	5.33**	.68	.13	.71	.28	.07
P5: Logical reasoning	8.94**	.45	.27	.72	.60	.18
P6: Solution search	-3.77**	1.15	-.05	.58	-.12	-.03
P7: Visual figures and graphs	5.78**	.55	.17	.77	.37	.09
P9: Data management	6.68**	.68	.13	.56	.35	.09
P10: Quantitative reading	4.20**	.53	.13	.56	.29	.07

R^2 = .945, adjusted R^2 = .943, $F(23, 691) = 513.528$, $p<.01$

Classification Rate with $D^2 < 4.5$ = 99.6%

* $p<.05$ ** $p<.01$

Table 3

Regression Coefficients of Booklet 5 for Total Score Model

Attribute	*B*	SE_B	β	r	$r_{partial}$	r_{part}
C1: Whole numbers and integers	-.76	.99	-.01	.53	-.03	-.01
C2: Fractions and decimals	11.46**	.41	.22	.46	.73	.18
C3: Elementary algebra	-	-	-	-	-	-
C4: Two-dimensional geometry	6.47**	.52	.09	.37	.43	.08
C5: Data and basic statistics	2.48*	1.09	.02	.19	.09	.01
S2: Number sense	3.44**	.29	.10	.39	.42	.08
S3: Figures, tables & graphs	3.65	10.59	.00	.12	.01	.00
S4: Approximation & estimation	3.89**	.35	.08	-.06	.39	.07
S5: Evaluate and verify options	1.23	1.69	.01	.26	.03	.01
S6: Recognize patterns	1.85**	.37	.07	.51	.19	.03
S7: Proportional reasoning	8.89**	.47	.20	.60	.59	.12
S8: Novel/unfamiliar problems	5.99**	.36	.13	.38	.53	.11
S10: Open-ended items	7.65**	.40	.25	.65	.59	.12
S11: Word problems	6.70**	1.06	.05	.27	.23	.04
P1: Translate	6.05**	.60	.11	.54	.36	.06
P2: Computation application	5.69**	.91	.06	.47	.23	.04
P3: Judgmental application	1.90**	.59	.02	.29	.12	.02
P4: Rule application in algebra	4.06**	.39	.11	.70	.37	.07
P5: Logical reasoning	9.45**	.37	.19	.45	.70	.16
P6: Solution search	.34	.53	.01	.69	.02	.00
P7: Visual figures and graphs	5.49**	.68	.06	.36	.29	.05
P9: Data management	9.68**	.43	.18	.48	.65	.14
P10: Quantitative reading	5.18**	.25	.17	.39	.62	.13

R^2 = .972, adjusted R^2 = .971, $F(22, 695)$ = 1105.708, p<.01

Classification Rate with D^2 < 4.5 = 99.9%

* p<.05 ** p<.01

a: Attribute C3 was excluded because tolerance is 0

Table 4

Regression Coefficients of Booklet 7 for Total Score Model

Attribute	*B*	SE_B	β	r	$r_{partial}$	r_{part}
C1: Whole numbers and integers	-3.69**	.88	-.03	.38	-.16	-.02
C2: Fractions and decimals	7.82**	.85	.06	.29	.33	.05
C3: Elementary algebra	2.63**	.30	.08	.72	.32	.05
C4: Two-dimensional geometry	3.63**	.31	.09	.64	.41	.06
C5: Data and basic statistics	3.99**	.49	.07	.46	.30	.05
S2: Number sense	3.07**	.29	.08	.63	.37	.06
S3: Figures, tables & graphs	5.45**	.88	.04	.35	.23	.03
S4: Approximation & estimation	3.16**	.29	.07	.20	.39	.06
S5: Evaluate and verify options	8.17**	2.23	.02	.27	.14	.02
S6: Recognize patterns	3.50**	.27	.12	.77	.44	.07
S7: Proportional reasoning	5.09**	.77	.05	.36	.25	.04
S8: Novel/unfamiliar problems	5.76**	.27	.15	.53	.63	.12
S10: Open-ended items	4.24**	.31	.13	.72	.46	.08
S11: Word problems	11.20**	.81	.09	.34	.47	.08
P1: Translate	6.09**	.78	.06	.37	.29	.04
P2: Computation application	8.97*	3.44	.02	.10	.10	.01
P3: Judgmental application	4.87**	.51	.07	.41	.34	.05
P4: Rule application in algebra	1.85**	.39	.05	.65	.18	.03
P5: Logical reasoning	5.73**	.19	.24	.69	.76	.17
P6: Solution search	-.12	.63	.00	.54	-.01	.00
P7: Visual figures and graphs	2.52**	.43	.05	.61	.22	.03
P9: Data management	6.04**	.31	.13	.42	.60	.11
P10: Quantitative reading	5.77**	.28	.16	.62	.62	.12

R^2 = .979, adjusted R^2 = .979, $F(23, 685) = 1405.790$, $p<.01$

Classification Rate with $D^2 < 4.5$ = 99.7%

* $p<.05$ ** $p<.01$

Table 5

Descriptive Statistics of Attribute Probabilities for Each Booklet

Attribute	Booklet 1 (*N*=721)		Booklet 3 (*N*=715)		Booklet 5 (*N*=718)		Booklet 7 (*N*=709)		Range
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
C1: Whole numbers and integers	.98	.08	.94	.15	.98	.09	.99	.07	.05
C2: Fractions and decimals	.98	.11	.99	.08	.96	.17	.99	.06	.03
C3: Elementary algebra	.84	.25	.92	.18	.91	.23	.86	.26	.06
C4: Two-dimensional geometry	.95	.15	.91	.23	.97	.12	.87	.20	.10
C5: Data and basic statistics	.98	.08	.86	.17	.99	.05	.96	.14	.13
S2: Number sense	.87	.19	.92	.14	.69	.25	.81	.22	.23
S3: Figures, tables & graphs	.99	.04	.99	.07	1.00	.01	.99	.06	.01
S4: Approximation & estimation	.82	.21	.84	.18	.80	.19	.82	.19	.04
S5: Evaluate and verify options	.99	.07	.99	.07	1.00	.05	1.00	.02	.01
S6: Recognize patterns	.73	.24	.47	.25	.66	.30	.74	.29	.27
S7: Proportional reasoning	.96	.11	.98	.09	.94	.19	.99	.08	.05
S8: Novel/unfamiliar problems	.94	.14	.99	.07	.89	.18	.87	.21	.12
S10: Open-ended items	.90	.20	.82	.30	.87	.28	.89	.25	.08
S11: Word problems	.99	.06	.99	.06	.99	.06	.99	.07	.00
P1: Translate	.97	.10	.99	.06	.96	.15	.99	.08	.02
P2: Computation application	.95	.15	.98	.10	.98	.09	1.00	.01	.05
P3: Judgmental application	.97	.10	.99	.08	.97	.11	.97	.13	.02
P4: Rule application in algebra	.78	.29	.91	.23	.91	.23	.91	.22	.13
P5: Logical reasoning	.85	.28	.85	.30	.93	.17	.73	.35	.20
P6: Solution search	.87	.18	.97	.12	.87	.23	.97	.10	.10
P7: Visual figures and graphs	.95	.16	.87	.28	.98	.09	.95	.15	.11
P9: Data management	.94	.15	.94	.19	.95	.16	.93	.18	.02
P10: Quantitative reading	.89	.17	.75	.29	.77	.29	.89	.22	.14

Table 6

Frequencies of Appearances of Attributes Required in Items for Each Booklet

Attribute	Booklet 1	Booklet 3	Booklet 5	Booklet 7	Total Item	Range[a]
C1: Whole numbers and integers	14	6	10	10	40	.05
C2: Fractions and decimals	15	19	19	17	70	.03
C3: Elementary algebra	10	9	4	4	27	.06
C4: Two-dimensional geometry	13	14	10	6	43	.10
C5: Data and basic statistics	10	6	10	7	33	.13
S2: Number sense	8	6	5	4	23	.23
S3: Figures, tables & graphs	19	17	22	15	73	.01
S4: Approximation & estimation	4	6	5	5	20	.04
S5: Evaluate and verify options	17	11	15	17	60	.01
S6: Recognize patterns	6	1	3	4	14	.27
S7: Proportional reasoning	11	12	14	12	49	.05
S8: Novel/unfamiliar problems	10	12	10	8	40	.12
S10: Open-ended items	13	10	12	9	44	.08
S11: Word problems	25	18	18	18	79	.00
P1: Translate	15	16	13	13	57	.02
P2: Computation application	17	17	19	19	72	.05
P3: Judgmental application	9	10	7	8	34	.02
P4: Rule application in algebra	7	9	4	4	24	.13
P5: Logical reasoning	17	13	11	7	48	.20
P6: Solution search	9	10	5	8	32	.10
P7: Visual figures and graphs	13	10	13	8	44	.11
P9: Data management	20	14	13	8	55	.02
P10: Quantitative reading	11	6	9	10	36	.14

a. The values of range, which are probability differences from the highest to lowest values, are adopted from Table 5.