Asia Pacific Journal of Educators and Education, Vol. 26, No. 1, 33–50, 2011

THE DEVELOPMENT AND CONSTRUCT VALIDATION OF THE MATHEMATICS PROFICIENCY TEST FOR 14-YEAR-OLD STUDENTS

Ahmad Zamri Khairani*
School of Educational Studies
Universiti Sains Malaysia
11800 Pulau Pinang
*Corresponding author: emel_zamri@yahoo.com

M. Sahari Nordin
Institute of Education, International Islamic University Malaysia
Jalan Gombak 53100 Kuala Lumpur
msahari@iiu.edu.my

Abstract: Mathematics education of yesterday, where the emphasis of procedural rote learning took centre stage, is no longer viable. Today, students must understand the Mathematics that they are learning; that is, Mathematics involves various components or *strands* that are interdependent and interwoven. The purpose of this study was to assess three of the strands, namely, conceptual understanding, procedural fluency, and strategic competence among 14-year-old students. This study also sought empirical evidence for how the Mathematics proficiency construct contributed to each strand. Results from Rasch Model calibration showed that students were most proficient in conceptual understanding followed by strategic competence and procedural fluency. Confirmatory factor analysis confirmed that Mathematics proficiency was a significant determinant for each strand. Lastly, this study reported several implications calling for future research.

Keywords: conceptual understanding, procedural fluency, strategic competence, Rasch model, confirmatory factor analysis

Abstrak: Pendidikan matematik yang menekankan hafalan prosedur-prosedur tidak lagi relevan pada masa ini. Pelajar sebaliknya perlu memahami bahawa matematik terdiri daripada beberapa komponen yang saling berkaitan dan saling bergantungan antara satu sama lain. Tujuan kajian ini adalah untuk menilai tiga daripada komponen ini, iaitu, pemahaman konsep, kelancaran prosedur, dan kecekapan penyelesaian masalah dalam kalangan pelajar-pelajar berusia 14 tahun. Kajian ini juga cuba mendapatkan bukti empirik tentang hubungan di antara konstruk profisiensi matematik dengan ketiga-tiga komponen ini. Dapatan kajian yang menggunakan tentukuran Rasch menunjukkan pelajar lebih profisien dalam pemahaman konsep diikuti dengan kecekapan penyelesaian masalah dan kelancaran prosedur. Analisis faktor pengesahan mengesahkan bahawa profisiensi Matematik merupakan penentu yang signifikan bagi ketiga-tiga komponen. Beberapa implikasi kajian yang memerlukan kajian lanjut turut dilaporkan.

Kata kunci: pemahaman konsep, kelancaran prosedur, kecekapan strategik, model Rasch, analisis faktor pengesahan

INTRODUCTION

Mathematics is important since every human being must be able to perform some basic mathematics in order to participate effectively within his or her society. Knowledge in mathematics enhances the capabilities of human mind, which in turn, facilitates the development of science and technology. Capacity for logical thought, explanation, and justification enables mathematics to function as a model of deductive reasoning, which is essential in bringing order to human affairs. The importance of both practical and theoretical aspects has earned the subject of Mathematics a pivotal place in teaching and learning.

Today, students must face new challenges in which mathematics is no longer limited to a few selected and isolated areas. The National Council of Teachers of Mathematics (1991) clearly emphasise the need for students to spend more time on reasoning and problem solving, communicating ideas, exploring relationships among representations of mathematical forms, and making connections between concepts. For example, while proficiency with numbers is important in mathematics, this proficiency can also be found throughout a mathematics curriculum. The domain of numbers both supports and is supported by other branches of mathematics, such as shape and space and relations (Ministry of Education, 2002).

In order to cope with the challenges, students must develop proficiencies essential to learning Mathematics, that is, aspects of expertise, competence knowledge, and facility. Since it is not possible to capture all components of Mathematics proficiency, researchers turn their attention to modelling the aspects to describe the complex process as well as to develop a wide range of ideas regarding the measurement of the construct (Bratina, 2004). One of the most highly cited mathematics proficiency models was proposed by Kilpatrick, Swafford and Findell (2001). The multidimensional model consists of five interwoven and interdependent strands: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition. The first three strands feature rudimentary mathematics abilities, whereas the latter two represent specifications for mathematics reasoning (ability to think logically) and mathematics communication (how students present their answer to solve problems). The strands are seen as "reflecting a firm, sizable body of scholarly literature both inside [such as Mathematics educators] and outside [for example, cognitive scientists] mathematics education" (Kilpatrick et al., 2001, p. 118, emphasis added).

Conceptual understanding features strands regarding comprehension of mathematics concepts, operations, and relations, while procedural fluency represents the skills to carry out procedures appropriately, efficiently, and accurately (Kilpatrick et al., 2001). Strategic competence, on the other hand, is similar to problem-solving. Capacity to think, reflect, explain, and justify is captured under adaptive reasoning while productive disposition involves the tendency to make sense in mathematics. This disposition helps students be more confident in their knowledge and ability (Kilpatrick et al., 2001; Resnick, 1987). It is important to acknowledge that the strands are interwoven and interdependent. For example, in the process of acquiring conceptual understanding, a certain level of procedural fluency is required to develop and strengthen that understanding. When solving non-routine problems (strategic competence), students' attitudes and beliefs as learners become more positive (productive disposition). In short, students with proficiency in mathematics "understand basic concept, are fluent in performing basic operations, exercise a repertoire of strategic knowledge, reason clearly and flexibly, and maintain a positive out-look toward mathematics" (Kilpatrick et al., 2001, p. 409).

The present study, however, intends to discuss only the first three strands as based on the following justifications. Firstly, it is important to recognise that size and capacity of the strands should not be confounded because the model provides a description of holistic mathematics proficiency strands. Some strands, however, may be more important at a certain age level as compared to the other strands. For example, while conceptual understanding and procedural fluency have fully developed for 14-year-old students, their adaptive reasoning may be quite limited (Inhelder & Piaget, 1958; Sternberg & Rifkin, 1979, as cited in Kilpatrick et al., 2001). Secondly, construction of test items to measure conceptual understanding, procedural fluency, and strategic competence are relatively easy using a standardised achievement test. However, a more thorough form of a test, such as performance assessment, is needed to scale productive disposition and adaptive reasoning.

The purpose of this study was twofold. First, the study aimed to investigate the adequacy of a purposely-developed proficiency test that served as a foundation to assess conceptual understanding, procedural fluency, and strategic competence among 14-year-old students. Evidence of the validity of the test was the essence of this study. Secondly, this study also aimed to develop a mathematics proficiency model and sought empirical evidence for how the mathematics proficiency construct contributed to each strand.

METHOD

Samples

The sample for this study consisted of 588 14-year-old students from the district of Lower Perak. Demographic features of the sample included the following: (a) 234 males (39.8%) and 354 females (60.2%), and (b) 384 Malays (65.3%), 59 Chinese (10.0%), and 145 Indians (24.7%). The sample represented three equally distributed levels of mathematics ability: high, moderate and low, as based on information from their respective schools. The sample size was considered adequate for performing a Rasch model analysis (Hambleton & Cook, 1983; Tang, Way, & Carey, 1993), as well as for measurement modelling (Chou & Bentler, 1995; Hoyle & Kenny, 1999). This study adopted a purposive sampling procedure because the researcher had personal knowledge about the sample schools, especially in terms of their mathematics achievement (Gay, Mills, & Airasian, 2006).

Instrument

The instrument used in this study was a self-developed, 50-item Mathematics Proficiency Test (MPT) with the following proportion of strands: Conceptual Understanding (50%), Procedural Fluency (32%), and Strategic Competence (18%). In the development of the MPT, three fundamental test parameters were identified: test content, learning outcomes and item difficulty. The first two parameters were outlined in the *Curriculum Specifications* (Ministry of Education, 2002), whereas the difficulty for every item was conceptually determined by the researcher. Test content included the strands, topics and subtopics. Learning outcomes identified knowledge, skills and abilities that students needed to demonstrate at the end of every topic or subtopic. Items were then developed to operationalise these learning outcomes in terms of item scores. Three levels of difficulty (easy, moderate and difficult) were applied to target the learning outcomes.

The process of the MPT content validation involved a discussion with three experienced teachers regarding the table of specifications and item difficulty. The weight of the topics in the table of specifications was given prime attention at the initial stage of the discussion. Several criteria were employed during the process. The criteria included: (1) the coverage given in the Curriculum Specifications, (2) the suitability of the topics for multiple-choice format, and (3) the rating of the topics by the panel. Topics with the weight of 3 were considered to be important topics for Form 2 since they were given the widest coverage in the Curriculum Specifications. The panel believed that these were important topics because they involved a lot of concepts to be understood, procedures to be mastered, and problems to be solved. The topics included Linear Equation, Algebraic Expressions II, Ratios, Rates and Proportions I, and Coordinates and Circles I. In addition, these topics were generally popular in the high-stake, national-level of *Penilaian Menengah Rendah*. Directed Numbers was considered to be the most important topic in the number sense strand.

However, the panel decided that the topic should be given the weight of 2 instead of 3 for a couple of reasons. First, some of the learning outcomes were addressed in Form 1. In Form 2, it was the expansion of the same knowledge, skills, and ability toward integers, fractions and decimals. Second, the same skills (multiplication and division) were addressed across the topics for integers, fractions and decimals. In short, regarding Directed Numbers, no new concept was introduced in Form 2. Similarly, although the topic of Transformation 1 covered a wide range of concepts, the panel agreed that it involved much easier concepts. Moreover, some of the content was repeated in Form 3. Squares, Square Roots, Cubes and Cube Roots were given the weight of only 1, although they were widely covered in the specifications. Those topics were considered to be more suitable in a constructed-response format instead of a multiple-choice format. Likewise, topics of Pythagoras' Theorem, Construction, Loci in Two Dimensions, Solid Geometry II and Statistics II were also more suited to be tested in a constructed-response format. Furthermore, these topics were only cited briefly in the Curriculum Specifications.

The next stage was to develop test items according to the table of specifications. At the same time, the difficulty level of each item was also conceptualised by the researcher. In this study, items for the MPT were developed through various procedures. Most of the items, especially in the conceptual understanding and procedural fluency, were adopted from the pool of items available. Also, a number of questions, especially in the strategic competence strand, were developed specifically for this study. After the items had been developed, they were again given to the same panel of experienced teachers for validation. One major concern was that the panellists tended to have opposing opinions as to the difficulty of some of the items. An easy item for a teacher might be considered as moderate or even hard for another. Similarly, a moderate item might be considered difficult to others. Thus, the panellists failed to reach an agreement regarding the distribution of item difficulty on the MPT. This situation was expected since conceptualisation of item difficulty had been identified as a potential problem from the item development stage. Hence, empirical data from the pilot was used to provide more meaningful information on item difficulty.

Data Analyses

The Rasch model analysis enabled the present study to determine the validity of the MPT in measuring mathematics proficiency. Rasch model analysis is a method of obtaining objective, fundamental and linear measures from stochastic observation of an ordered category (Linacre, 2005). In this study, the WINSTEPS version 3.57 Rasch Model software (Linacre, 2005) was used. In WINSTEPS, the test scores (called *measures* and reported in *logits*) were determined through an iterative calibration of both the person and item using Joint Maximum Likelihood Estimation (JMLE). The measures from the Rasch Model calibration had properties that met the requirements of fundamental measurement (Hambleton & Swaminathan, 1985; Wright & Masters, 1982). These properties include: (1) the measures are invariant of items and persons used in the calibration process, and (2) the inclusion of quality-control fit statistics for both item and person for precision of estimation. Most relevant to this study, however, is the fact that through verification of empirical data and theory, the Rasch Model allowed the test results to be directly referenced to the measured construct, thus facilitating interpretation of students' achievement (Masters, Adams, & Lokan, 1994).

In this study, the adequacy of the MPT in measuring the mathematics proficiency construct was examined in two aspects: (1) validity of test items and (2) construct validity of mathematics proficiency. Evidence of validity of the test items was demonstrated through three criteria: item polarity, fit statistics, and dimensionality (Wright & Stone, 1979). Item polarity gives an indication of whether the items are working together in the same direction to define a construct. Fit statistics, the infit mean-square (MNSQ) and outfit MNSQ, help detect discrepancies between the data and Rasch Model expectation. Only when a test fits the model expectation can be considered as having the property of fundamental measurement. Meanwhile, investigation of dimensionality using Principal Component Analysis (PCA) was carried out to ensure that the MPT was measuring only a single construct, the Mathematics proficiency construct.

One essential aspect of construct validity is related to the score's meaning and interpretation. Messick (1993) points out two major threats to construct validity, namely, *construct-irrelevant variance* and *construct under-representation*. Construct-irrelevant variance relates to the irrelevant sub-dimensions that contaminate measurement of the focal construct by producing reliable variance in the test scores. Construct under-representation identifies whether there were significant gaps between item distributions. Baghaei (2008) argues that within the framework of Rasch Model analyses, items that do not fit the model's expectation are instances of construct-irrelevant variance, whereas significant gaps between items along the continuum of measured scale are indications of construct under-representation.

Confirmatory factor analysis (CFA), AMOS version 7.0, was used in this study to verify that mathematics proficiency is a significant determinant for the three strands. The data-fitting software adopted maximum likelihood procedure for parameter estimation in the measurement model, where the proficiency strands are linked to their underlying factor. Thus the primary interest of CFA is observation of strengths of the regression path (factor loadings) between the underlying factor and the strands. The unit of analysis for the three strands hypothesised was the *logit* measures from the Rasch Model calibration of the MPT. Figure 1 shows the hypothesised model where Mathematics proficiency (MATHPROF) was the underlying factor manifested by the three indicators: conceptual understanding (CONCEPT), procedural fluency (PROCEDUR), and strategic competence (STRATEGI). Error terms that usually related to measurement error are labelled e1, e2 and e3.

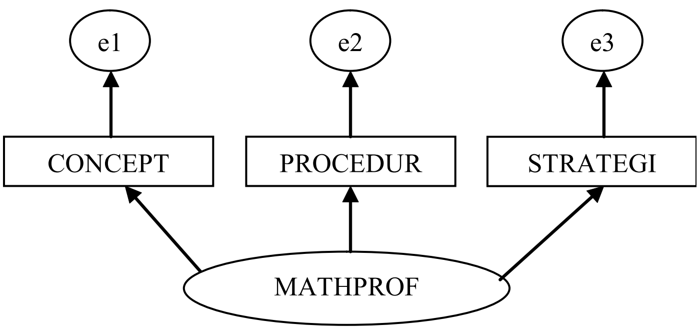

Figure 1. The hypothesised mathematics proficiency model

RESULTS AND DISCUSSION

Summary of item difficulty and student ability measures provided initial information as to the adequacy of the MPT. In Figure 2, both the items and the students were located along the proficiency scale. The items on the top were more difficult, and the students at the top displayed higher ability. As we went down the line, the items became easier, and the students displayed less ability. Item difficulty measures spread approximately 4 *logits* (from 1.65 to +2.18), while student ability measures spanned approximately 8 *logits* (from 3.09 to +5.09). The mean for item difficulty was 0.00 (standard error = .89), while the mean for student ability was .10 (standard error = 1.22).

The small difference in mean measures of the students and the items indicated that the MPT targeted the student well. Reliability of item difficulty measures was very high (.99), suggesting that the ordering of item difficulty was highly replicable with another comparable sample of students. Internal consistency of the student ability measure was also high at .90, indicating that it was highly likely that the ordering of student ability could be replicated because most of the variance in the measured scores was attributed to true variance of the mathematics proficiency construct.

As depicted in Table 1, the point measure correlation (PTMEA CORR.) ranged from .25 to .60, with no item containing zero or negative values. This correlation indicated that all items were working together in the same way in defining the mathematics proficiency construct. The means of the infit and outfit MNSQ of 1.00 and 1.01, respectively, were close to the value expected by the model (1.00). This suggests that the amount of distortion of the measurement was minimal. Although the standard deviation of both the infit and outfit MNSQ (.10 and .16, respectively) were slightly higher than the expected value, these discrepancies were small and showed that the data demonstrated little variation from the Rasch Model expectation. Individual items demonstrated infit MNSQ values from 0.81 to 1.09, while outfit MNSQ were between 0.76–1.3, which were within the acceptable range of 0.7–1.3 (Bond & Fox, 2001). Results of the PCA of the residuals indicated that the largest factor extracted from the residuals was 2.1 units, which has the strength of about 2 items and is well below the 5 items needed for consideration as a second factor (Linacre, 2005). In addition, no gaps of .5 *logits* or more (Linacre, 2004) between item distributions on the proficiency scale showed that the items were adequate in accessing important features of the Mathematics proficiency construct. Thus, it can be concluded that the MPT was adequate in measuring the Mathematics proficiency construct.

Empirical scaling indicated that items for conceptual understanding were the easiest (mean = –0.16 *logits*) among the three strands, followed by strategic competence (mean = –0.0045 *logits*). In contrast, procedural fluency (mean = 0.24 *logits*) was the most difficult strand. The scaling of the proficiency strands was within the researcher's expectation, although it is contradictory to popular belief that solving strategic competence (problem solving) was the most difficult task. This can be attributed to the fact that strategic competence involves mostly word-problem items as opposed to the NAEP-like items usually associated with research in problem-solving ability of students.

One important observation from the inspection of the individual item was that the Form 2 students involved in this study were able to answer items related to understanding information that was explicitly stated (e.g., the terminology found in the diagram [item 29], or from the list of answers option [item 47]). This understanding was expected since the items only required a lower level of conceptual understanding in which students only need to locate the information that is explicitly stated. Similarly, the students were able to answer items that require use of some straightforward procedures (e.g., items 17 and 26). However, students had difficulty with items that required them to use their prior knowledge to solve new problems, particularly, making connections between topics. That trend is troublesome because making connections between various forms of mathematical knowledge, especially between concept and procedure and between mathematics and real-life experience, is important to effective mathematics learning and teaching (Lappan, Fey, Fitzgerald, Friel, & Phillips, 2002).

Another important finding was that all strands showed substantial dispersion, with both conceptual understanding and strategic competence having a similar spread (standard deviations SD of 0.97 and 0.98 respectively), while procedural fluency had a smaller spread (SD = 0.73). While the large dispersion for conceptual understanding was within expectation due to the number of items involved (*n* = 23), a relatively small spread in procedural fluency (*n* = 16), as compared to strategic competence (*n* = 9), was quite unexpected. Similarly, empirical scaling demonstrated that students were most proficient in conceptual understanding (mean = 0.16 *logits*) followed by strategic competence (mean = –0.48 *logits*). These students were least proficient in procedural fluency (mean = –1.31 *logits*). In terms of dispersion, student measures spread substantially in both strategic competence (SD = 1.46 *logits*) and conceptual understanding (SD = 1.26 *logits*), but they demonstrated less dispersion in procedural fluency (SD = 0.69 *logits*).

```
STUDENTS MAP OF items
               <more>|<rare>
    5                .  +
                        |
                        |
                        |
                        |
                        |
    4                .  +
                     .  |
                        |
                        |
                     .  |
                     .  |
    3                .  +
                     .  |
                     .  |
                    .#  |
                    .# T|
                        |  Q50
    2                #  +  Q5
                 .###   |T Q4
               .#####   |  Q48
                  ###   |  Q14
                   .#   |  Q37
               ###### S|  Q20    Q38
    1              .#   +
               .#####   |S Q6
                #####   |  Q19    Q8
            .########   |  Q31
           .#########   |  Q30    Q41    Q42    Q43
                .####   |  Q10    Q45
    0         .#######  +M Q18    Q23    Q34    Q36    Q39    Q40
            .######## M|  Q12    Q22    Q25    Q32    Q9
             .#######   |  Q27    Q35    Q46    Q7
             .#######   |  Q26    Q33
                #####   |  Q11    Q15    Q16    Q17    Q24    Q49
             .#######   |S Q2     Q21    Q47
   -1     .##########   +  Q13    Q44
                 .###   |  Q1
             .####### S|  Q3
               ######   |  Q28
                .####   |  Q29
                 .###   |T
   -2             .##   +
                   .#   |
                    .   |
                    # T|
                        |
                        |
   -3                   +
                    .   |
                        |
                        |
                        |
                        |
   -4                   +
               <less>|<frequ>
EACH '#' WAS 4.
```

Figure 2. Wright map

Table 1. The MPT: Item statistics

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	item
20	154	468	1.09	.11	1.19	3.6	1.30	4.0	.25	Q20
48	102	400	1.71	.13	1.14	2.1	1.29	2.9	.25	Q48
10	261	576	.10	.09	1.19	4.6	1.24	4.0	.32	Q10
8	199	537	.67	.10	1.15	3.3	1.23	3.4	.33	Q8
18	269	573	.02	.09	1.16	3.8	1.28	4.4	.34	Q18
45	247	569	.23	.10	1.15	3.5	1.28	4.5	.34	Q45
11	339	568	-.70	.10	1.09	2.3	1.19	3.0	.35	Q11
47	354	564	-.86	.10	1.07	1.8	1.21	3.1	.35	Q47
46	303	573	-.32	.09	1.13	3.3	1.18	3.0	.36	Q46
6	173	493	.90	.11	1.09	1.8	1.15	2.3	.36	Q6
41	240	566	.29	.10	1.13	3.1	1.16	2.6	.37	Q41
25	288	575	-.17	.09	1.11	2.7	1.13	2.3	.38	Q25
50	68	296	2.18	.15	.99	-.1	1.00	.1	.38	Q50
17	333	572	-.60	.10	1.05	1.3	1.10	1.6	.40	Q17
30	243	568	.26	.10	1.09	2.1	1.10	1.6	.40	Q30
4	97	370	1.77	.13	1.00	.0	.96	-.4	.40	Q4
42	233	553	.33	.10	1.06	1.6	1.12	2.0	.41	Q42
37	127	426	1.37	.12	.98	-.4	1.08	1.0	.42	Q37
14	118	382	1.44	.12	.97	-.4	1.01	.1	.42	Q14
31	210	537	.56	.10	1.03	.7	1.07	1.2	.42	Q31
36	264	576	.08	.10	1.06	1.4	1.09	1.6	.43	Q36
19	204	538	.63	.10	1.03	.6	1.09	1.4	.43	Q19
5	78	306	2.00	.14	.95	-.7	.94	-.6	.43	Q5
38	142	467	1.22	.11	.97	-.5	1.09	1.2	.43	Q38
26	325	574	-.52	.09	1.02	.6	1.01	.2	.44	Q26
34	274	569	-.08	.09	1.03	.8	1.02	.4	.44	Q34
35	309	573	-.38	.09	1.01	.2	1.06	1.1	.45	Q35
43	236	565	.30	.10	1.02	.4	1.02	.4	.45	Q43
12	289	570	-.21	.09	.99	-.2	1.02	.4	.46	Q12
1	375	558	-1.13	.10	.95	-1.1	.87	-1.9	.46	Q1
29	386	518	-1.65	.11	.90	-1.9	.82	-2.2	.47	Q29
33	326	574	-.53	.09	.97	-.7	.94	-1.0	.48	Q33
7	297	575	-.26	.09	.98	-.4	.94	-1.1	.48	Q7
40	272	571	-.05	.09	.97	-.7	.98	-.4	.48	Q40
44	354	558	-.92	.10	.93	-1.8	.91	-1.4	.49	Q44
39	268	571	.02	.10	.98	-.5	.96	-.8	.49	Q39
28	388	531	-1.54	.11	.89	-2.2	.79	-2.7	.49	Q28
32	278	573	-.09	.09	.95	-1.3	.95	-1.0	.50	Q32
27	299	576	-.27	.09	.93	-1.8	.96	-.6	.51	Q27
15	331	572	-.59	.10	.92	-2.2	.92	-1.4	.52	Q15
9	288	572	-.19	.09	.93	-1.8	.90	-1.9	.52	Q9
2	351	566	-.82	.10	.90	-2.7	.87	-2.0	.52	Q2
13	361	560	-.98	.10	.89	-2.8	.82	-2.9	.52	Q13
3	388	543	-1.41	.10	.86	-3.1	.74	-3.5	.53	Q3
22	292	576	-.20	.09	.88	-3.2	.87	-2.5	.55	Q22
16	334	573	-.61	.10	.85	-4.1	.79	-3.7	.57	Q16
23	266	575	.07	.10	.86	-3.7	.84	-3.0	.57	Q23
24	337	571	-.64	.10	.84	-4.2	.77	-4.1	.58	Q24
21	349	566	-.80	.10	.81	-5.0	.72	-4.8	.60	Q21
49	337	564	-.71	.10	.81	-5.3	.76	-4.3	.60	Q49
MEAN	267.1	534.9	.00	.10	1.00	-.1	1.01	.1		
S.D.	85.0	69.9	.89	.01	.10	2.4	.16	2.4		

The result of the confirmatory factor analysis was found to be of statistical significance and of practical importance since the standardised structural coefficient was larger than 0.1. This implied the following: (1) mathematics proficiency was a significant determinant for all three strands, and it contributed almost evenly toward conceptual understanding (90%), procedural fluency (81%), and strategic competence (82%); (2) a large amount of variability (81%) in conceptual understanding was accounted by Mathematics proficiency, as compared to 66% and 68% in procedural fluency and strategic competence, respectively; and (3) each strand and its error term had a nonzero loading on mathematics proficiency, meaning that it was not inter-correlated. The result of the confirmatory factor analysis provided more evidence of adequacy of the MPT in measuring strands of mathematics proficiency, where each item was adequate in measuring one strand at a time. Since the model was just-identified (i.e., the number of data variances and covariances equals the number of parameters to be estimated), it is unnecessary to assess the value of fit indices (TLI, CFI, GFI, RMSEA, etc.) because the data will fit the model perfectly. In other words, it is known that the observed measures meet the minimal requirement of the parameters that are estimated. Thus, it can be concluded that the parameters are estimable, whereas the model is testable eventhough it yields a unique solution for all parameters (Byrne, 2001). Results from the Pearson correlations further support the findings. All indicators were positively correlated (correlation coefficients between .705 and .751). In short, the study confirms that the three strands of mathematics proficiency, as proposed by Kilpatrick et al. (2001), were accepted by the data.

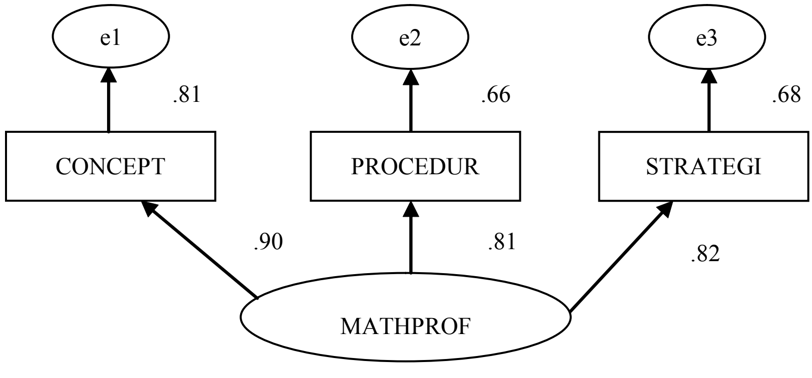

Figure 3. Standardised coefficients of the hypothesised Mathematics Proficiency Model

Table 2. Inter-correlations among constructs

	Conceptual understanding	Procedural fluency	Strategic competence
Conceptual understanding	1.000		
Procedural fluency	.751**	1.000	
Strategic competence	.727**	.705**	1.000

Note: ** $p < .01$, ($N = 588$)

CONCLUSION

Using the framework of the Rasch Model, it is evident that the MPT was adequate in measuring the mathematics proficiency construct regarding the following findings: (a) the individual item provided enough contribution to the overall measurement of Mathematics proficiency construct, (b) the MPT fit the requirement of the Rasch measurement model and demonstrated substantial evidence of construct validity, and as such, the calibrated items were useful in measuring students' proficiency in mathematics, (c) the mathematics proficiency construct measured using the test did not confound with other related constructs, and (d) threats to construct validity, such as construct-irrelevance variances and construct under-representation, were kept to a minimum. In addition, the study found that all indicators in the hypothesised model had strong agreement with the model proposed by Kilpatrick et al. (2001).

As documented throughout the findings, the MPT was developed, validated and scrutinised for empirical evidence of adequacy in measuring the mathematics proficiency construct. Practically, the procedure may be replicated so that commendable results can be obtained from a particular test that would tend to measure any construct. Theoretically, having the capacity to resolve some of the rudimentary issues in measurement, this study has added more evidence in favour of the Rasch model. The present study also extends some theoretical implications regarding the investigation of the mathematics proficiency construct within a local context. The 5-factor model, proposed by Kilpatrick et al. (2001), represents a holistic view of mathematics proficiency. However, the model requires some comprehensive assessments to test the model as well examine whether all the factors are estimable. Thus, this study delimits its investigation into conceptual understanding, procedural fluency, and strategic competence factors. Moreover, these are the rudimentary mathematics abilities that are essential as foundations for related constructs, such as mathematics reasoning and mathematics communication. Hopefully, the findings would trigger more effort in enriching our knowledge of the mathematics proficiency construct within our own local context.

The study also has several practical implications that warrant further investigation into developing proficiency in teaching mathematics. First, assessments should be planned to promote greater mathematics proficiency rather than to rank students. Assessments should be in the form of criterion-referenced, where information about what students know and are able to do are available so that they can learn from the assessment. Teachers, on the other hand, can use the information to make effective instruction decisions sincebecause feedback generally leads to clearer and more effective instruction (Fennema, Carpenter, Franke, Levi, Jacobs, & Empson, 1996; Thompson & Briars, 1989).

Regarding mathematics pedagogy in schools, the findings clearly showed the importance of making connections within and among topics to nurture proficiency. Mathematics cannot and should not be taught as an isolated construct; rather, mathematics should be interwoven and interdependent among topics or strands. Successful learning can be characterised by comprehension of mathematical ideas. As such, teaching Mathematics also requires similar aspects. Mathematics teachers must possess knowledge that is connected: knowledge of Mathematics, students, and pedagogy (Kilpatrick et al., 2001). Integrated knowledge of Mathematics (the content), knowledge on how students develop their understanding (the psychology of learning), and knowledge on how Mathematics should be taught (the teaching method) are the kinds of knowledge that would very much make a difference in teaching and learning.

REFERENCES

Baghaei, P. (2008). Rasch model as a construct validation tool. *Rasch Masurement Transaction, 22*(1), 1145. Retrieved 21 December 2008, from http://www.rasch.org

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn: Brain, mind, experience and schoo*l. Washington, DC: National Academy Press.

Bratina, T. A. (2004). *Reflections on a model for describing what it means to be proficient.* Retrieved 4 December 2006, from www.iste.org/rinterTemlate.cfm?Section=2003_20048&Temlate=/ContentManagement/HTMLDisplay.cfm&ContentID=2335&FuseFlag=1

Byrne, B. M. (2001). *Structural equation modelling with AMOS*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modelling* (pp. 37–59). Thousand Oaks, CA: SAGE.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, *27*, 403–434.

Fuson, K. C. (1992). Research on whole number addition and subtraction. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 243–275). New York: Macmillan.

Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational research* (8th ed.). New Jersey: Pearson Prentice Hall.

Hambleton, R. K., & Cook, L. (1983). Robustness of item response model and effect of test length and sample size in precision of ability estimates. In D. weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing.* New York: Academic Press.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.

Hiebert, J. (Ed.). (1986). *Conceptual and procedural knowledge: The case of mathematics.* Hillsdale, NJ: Erlbaum.

Hiebert, J., & Carpenter, T. (1992). Learning and teaching with understanding. In D. Grouws (Ed.), *Handbook in research in mathematics teaching and learning* (pp. 65–97). New York: Mcmillan.

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability and test of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195–222). Thousand Oaks, CA: SAGE.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping how children learn mathematics*. Washington, DC: National Academy Press.

Lappan, G., Fey, J. T., Fitzgerald, W., Friel, S. N., & Phillips, E. D. (2002). *Connected mathematics series*. Glenview, IL: Prentice Hall.

Linacre, J. M. (2004). From Microscale to Winsteps: 20 years of Rasch software development. *Rasch Measurement Transactions, 17*(4), 958.

Linacre J. M. (2005). *A user's guide to Winsteps* [Program Manual].

Masters, G. N., Adams, R., & Lokan, J. (1994). Mapping student achievement. *International Journal of Educational Research*, *21*, 557–664.

Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg, & T. Ben-Zee (Eds.), *The nature of mathematical thinking* (Studies in Mathematical Thinking and Learning Series, pp. 29–53). Mahwah, NJ: Erlbaum.

Messick, S. (1993). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13–104). Phoenix, AZ: Oryx Press.

Ministry of Education. (2002). *Curriculum specifications for Mathematics Form 2.* Kuala Lumpur: Curriculum Development Centre.

National Assessment Governing Board. (2000). *Mathematics framework for the 1996 and 2000 National Assessment of Educational Progress.* Washington, DC: Author. Retrieved 10 August 2007, from http://www.nagb.org/pubs/96–2000math/toc.html

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.

Nunes, T. (1992). Cognitive invariants and cultural variation in mathematical concepts. *International Journal of Behavioral Development, 15,* 433–453.

Resnick, L. B. (1987). *Education and learning to think.* Washington, DC: National Academy Press. Retrieved 10 July 2006, from http://books.nap.edu/catalog/1032.html

Rittle-Johnson, B., & Siegler, R. S. (1998). The relation between conceptual and procedural knowledge in learning mathematics: A review. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 75–110). East Sussex, UK: Psychology Press.

Saxe, G. (1990). *Culture and cognitive development.* Hillsdale, NJ: Erlbaum.

Schoenfeld, A. H. (1988). Problem solving in context(s). In R. I. Charles, & E. A. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (Research Agenda for Mathematics Education, vol. 3, pp. 82–92). Reston, VA: National Council of Teachers of Mathematics.

Tang, K. L., Way, W. D., & Carey, P. A. (1993). *The effect of small calibration sample size on TOEFL IRT-based equating* (Tech. Rep. No. 7). Princeton, NJ: Educational Testing Service.

Thompson, A. G., & Briars, D. J. (1989). Assesing students' learning to inform teaching: The message in NCTM's evaluation standards. *Arithmetic Teacher*, *37*(4), 22–26.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Wright, B., D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.

APPENDIX: SAMPLE ITEMS

Conceptual Understanding

7 In algebraic term $-\frac{x^2yz}{3}$, which statements are **TRUE**?

I Coefficient of z is $-x^2y$

II Coefficient of y is $-x^2y$

III Coefficient of yz is $-\frac{x^2}{3}$

IV Coefficient of x^2yz is $-\frac{1}{3}$

A I and II **C** II and III

B I and III **D** III and IV

Procedural Fluency

8 $12w^3y^3 \div (-9wxy) \times (-3x^2y^3) =$

A $4w^2xy^4$ **C** $4w^4x^3y^6$

B $4w^4x^2y^3$ **D** $\dfrac{4w^2}{9x^3y^2}$

Strategic Competence

6 A square cardboard has an area of 324 cm^2. It is cut into 9 pieces of equal width and length. If all equal pieces are joined to form a long piece, find the length, in cm, of the long piece.

A 18 **C** 72

B 54 **D** 84