TESTING A DEVELOPMENTAL MODEL OF MEASURING PROBLEM SOLVING SKILLS BASED ON SCHEMA THEORY

Leonidas Kyriakides, George Philippou & Charalambos Charalambous

Department of Education, University of Cyprus

The aim of this study was to provide construct related evidence of a model measuring problem solving (PS) skills based on Marshall's schema theory (ST) using responses from 712, 4th and 5th grade Cypriot students to a battery of tests on PS. The Extended Logistic Model of Rasch was used and a scale was created for the battery of tests and analysed for reliability, fit to the model, meaning and validity. It was also analysed separately for each of the two types of knowledge proposed by Marshall in order to examine the appropriateness of ST in building a model of measuring PS skills. The analysis reveals that the battery of tests has satisfactory psychometric properties and supports the conceptual design of the proposed model. The findings are discussed with reference to intended uses of the assessment of PS skills and suggestions for further research are drawn.

INTRODUCTION

Problem solving (PS) is a central issue in mathematics education theory, as it can be documented by recently published literature on curriculum and assessment in Mathematics (see e.g., NCTM, 2000). In this context, Marshall (1995) developed a comprehensive proposal for teaching and assessing PS, which is grounded on schema theory (ST) and the acquisition of basic schemas by the learners. Mayer (1992) points out that "a schema is an organised structure consisting of certain elements and relations which are related to a situation and it can be used for understanding incoming information" (p. 228). ST was based on the assumption that the external representations used to describe the structure of a problem (i.e. diagrams) can serve in constructing a mental model which can be retrieved and used in solving analogous problems of the same structure (Goldin, 1998). Thus, ST aims to help students systematise PS experiences by providing them with simple diagrams for solving problems of additive and multiplicative structure.

Previous research on PS (Schmidt & Weiser, 1995) revealed that the crucial element in solving a problem lies in its structure and not on characteristics such as its context, content, or the number or type of the operations needed for solving the problem. Consequently, attention should be paid to the construction of the mental schema, which mirrors the structure of the problem. Diagrams or appropriate physical schemata can serve as vehicles for developing a solution plan (Marshall, 1995), since they provide access to similar structure problems that have been encountered in the past, and the means for reformulating or simplifying a problem. It is also recognised that retrieving the appropriate schema facilitates the design of a solution strategy, which is the most important part in the whole endeavour. Nesher and Hershkovitz

(1994) argue that schemas constitute the bridge between the verbal formulation of a problem and its mathematical structure. Schemas are, therefore, of primary importance with respect to cognitive processes involved in PS, since they facilitate the comprehension of the semantic relations in given text and serve as generalised frames for action in given situations (Philippou & Christou, 1999). In brief, schemas help students construct deeper understanding of problems, clarify their thinking and justify their ideas.

Marshall (1995) proposed five distinct problem situations: change, group, compare, restate and vary. The former three situations can be used to solve additive structure problems, while the last two situations are mainly used for solving multiplicative structure problems. For each situation, Marshall (1995) proposed an appropriate diagram, which is expected to help students recognise the problem situation and solve the problem. In this process, students are expected to proceed through four stages, each representing a different type of knowledge: identification knowledge, elaboration knowledge, planning knowledge and execution knowledge. The first stage refers to specific characteristics, features, and facts that help students recognise the schema; identifying the main elements of a problem can be considered as the most important part for schema activation, since it is this understanding that contributes to the initial recognition of a situation. The second stage refers to working out rules and limitations having to do with the use of the schema; at this stage the students recognise the details that are distinct to each schema, and consequently choose or construct the correct schema to solve the problem. The planning knowledge refers to ways through which students make decisions about the schema that is appropriate for the solution of a problem. Finally, the last type of knowledge includes a set of procedures, rules, or algorithms that can be applied to reach sub-goals and the final goal.

To the best of our knowledge and belief, no systematic attempt has so far been made to verify the sequence of these four stages, or to investigate whether these stages, representing different types of knowledge, may help us form a developmental model measuring PS skills based on ST. We consider the development of a model as highly important, since it may help teachers design activities that will enable students to progressively develop the required abilities to solve routine problems.

In this context, the main purpose of this study was to collect empirical data in order to examine the validity of a model measuring pupils' skills in PS. The developed model refers to additive structure problems of three situations, namely: change, group and compare, with regard to the former two types of knowledge that a student is expected to acquire (identification and elaboration knowledge) in order to be able to use appropriate schema(s) to solve a problem. Hence, another purpose of this study was to reconfirm the two aforementioned types of knowledge proposed by Marshall which could be helpful for using ST in teaching and assessing PS.

THE DEVELOPMENT OF THE BATTERY OF TESTS ON PS

To answer our research questions, a battery of tests on PS was constructed guided by existing research and theory on assessment of PS skills in Mathematics and by taking into account ST. Further, we sought a developmental ordering of tasks on a continuum of difficulty, which is an essential concept derived from research on developmental assessments for measuring proficiency in cognitive abilities and content areas (Hambleton, 1995). A final key requirement in designing the tests was its alignment with the mathematics curriculum that was operative in the country where the study was conducted.

The specification table of the tests (Table 1) included eight levels of PS skills which belong to the first two types of knowledge proposed by Marshall (1995); the former four levels referred to the identification knowledge of schemas and the remaining four to the elaboration knowledge.

Types of knowledge	Levels	Items of the battery of tests
Identification knowledge	1. Verbal recognition of problems	1-26
	2. Diagrammatical recognition of problems	27-52
	3. Selection of a problem reflecting the structure of a given diagram	53-78
	4. Posing questions from a mathematical situation	131-139
Elaboration knowledge	5. Filling in the boxes of given diagrams	105-130
	6. Correct placement of the unknown quantity	79-104
	7. Problem Posing based on given contexts	140-159
	8. Problem Posing based on given numbers	160-178

Table 1: Specification table of the tests on PS based on ST

The eight levels were mainly based on tasks proposed by Marshall (1995) for assessing the acquisition of each type of schema-knowledge. Specifically, the first two levels included tasks examining the verbal and diagrammatical identification of the schema needed for solving a problem (i.e. students were requested to identify the structure or the appropriate diagram which could be used to solve a given problem). The third level included tasks examining students' ability to select problems that could be solved using a given diagram, while the fourth level referred to their ability in posing a question to produce a problem of a certain structure. The four remaining levels included tasks such as filling the data in a given diagram to represent the structure of the problem (5^{th} level), placing the unknown quantity in the correct position of a diagram (6^{th} level), and posing a problem given either the context of the problem (i.e., a given diagram including words and numbers) or solely numbers (7^{th} and 8^{th} level, respectively). The specification table guided the construction of a battery of tests with 178 items, representing the eight aforementioned levels. Each level included tasks of all three problem-situations (change, group and compare situation).

METHODS

The items in the final version of the battery of tests were content validated by three experienced primary teachers and two university tutors of Mathematics Education. The "judges" of the tests were asked to mark-up, make marginal notes or comments on or even rewrite the items. Based on their comments, minor amendments were made particularly where some terms used were considered as unfamiliar to primary pupils. The final version of the tests (available on request) was administered to all 4th grade (314) and 5th grade (398) pupils from 21 primary schools selected by stratified sampling (336 of the subjects were boys and 376 were girls).

The Extended Logistic Model of Rasch (Rasch, 1980) was used and the data were analysed by using the computer program Quest (Adams & Khoo, 1996). The data were initially analysed with the whole sample (n=712) for all items together; it was found that all items fit the model. The analysis was repeated with each of the four groups (grade 4, grade 5, boys and girls) of the sample, to investigate whether the battery of tests is used consistently by each group of the sample. By taking into account the item difficulties derived from the analysis of the whole sample, we used the procedure for detecting pattern clustering in measurement designs developed by Marcoulides & Drezner (1999) in order to examine whether levels of PS skills similar to those described in the specification table could be identified. Moreover, separate analysis of the two sub-scales, which refer to the first two types of knowledge mentioned by Marshall, was conducted to analyse the meaning of the general scale and the trait it measures.

FINDINGS

Figure 1 illustrates the scale for the 178 test items with item difficulties and the whole group of pupils' measures calibrated on the same scale. Both figure 1 and the item fit map for the 178 items fitting the model reveal that all the items of the tests have a good fit to the measurement model. Moreover, pupils scores range from –3.44 to 3.58 logits and the item difficulties range from -3.66 to 3.62 logits. This implies that the 178 items of the test are well targeted against the pupils' measures.

Table 2 provides a summary of the scale statistics for the whole sample and for each of the four groups of the sample. We can observe that for the whole sample and for each group the indices of cases and item separation are higher than 0.85 indicating that the separability of the scale is satisfactory (Wright, 1985). We can also see that the infit mean squares and the outfit mean squares are 1 and that the values of the infit t-scores and the outfit t-scores are approximately zero. And since the mean squares are within 30% of the expected values, calculated according to the model, it can be claimed that there is a good fit to the model. Moreover, the analyses of each of the four groups separately revealed that almost all items (176 out of 178) have difficulties, which could be considered invariant among boys and girls, within the measurement error (0.15 logits). The difficulties of 171 out of 178 items are invariant between the two age groups but 7 items vary markedly across the two age groups. Thus, an important aspect of creating a scale (sample-free item difficulties) has not been completely achieved.

```
----------------------------------------------------------------
High Achievement in Problem Solving       Difficult items
  Thresholds                            |
                                  X     |  172 174 177 178 175 176 166
                                 XX     |  155 157 159 162 170 171 165 167
                                 XX     |  146 156 158 160 152 153 164 173
  3.0                           XXX     |  144 145 161 140 147 151 163 168
                               XXXX     |  142 143 150 141 148 149 154 169
                                 XX     |
                                 XX     |
                                XXX     |    80   81   83   93   94 95 96
  2.0                            XX     |    87   89   91   100  97 98 99
                               XXXX     |    79   85   86   92   103 101
                               XXXX     |    82   84   88   90   102 104
                              XXXXX     |
                          XXXXXXXXX     |   126 129 130 123 125
  1.0                    XXXXXXXXXX     |   120 124 128 110 121 115 122
                       XXXXXXXXXXXX     |   107 111 114 112 116 118 119
                     XXXXXXXXXXXXXX     |   105 109 106 108 113 117 127
                    XXXXXXXXXXXXXXX     |
                XXXXXXXXXXXXXXXXXX      |   137    135    139
   .0          XXXXXXXXXXXXXXXXXXX      |   132    136    134
                XXXXXXXXXXXXXXXXXX      |   131    133    138
                 XXXXXXXXXXXXXXXXX      |
                    XXXXXXXXXXXXXX      |
                     XXXXXXXXXXXXX      |    74   78   70   72   75   76
 -1.0                 XXXXXXXXXXXX      |    58   68   60   56   65   66   77
                        XXXXXXXXXX      |    71   55   53   59   64   67   73
                          XXXXXXXX      |    62   54   57   61   63   69
                            XXXXXXX     |
                              XXXXX     |   51   52   50   49   48
 -2.0                           XXX     |   28   30   31   32   41   45   46
                               XXXX     |   29   34   38   33   40   44   47
                                XXX     |   27   35   36   37   39   42   43
                                 XX     |
                                  X     |   6   21   22
 -3.0                            XX     |   1   3    7    13    16    17   19
                                 XX     |   5   11   9    14    18    20   23
                                  X     |   2   4    8    10    12    15   24
                                        |   26   25
                                        |
Weak achievement in Problem solving       Easy items
----------------------------------------------------------------
```

Note: Each X represents 3 pupils

Figure 1: Scale for the battery of tests on PS (N=712, L=178)

The Rasch model was also helpful in analysing the conceptual design of the battery of tests. The indices of cases and item separation of each sub-scale (representing the two types of schema-knowledge) for the whole group and for each of the four groups of the sample are higher than 0.85 indicating that the separability of each sub-scale is satisfactory. Moreover, the infit mean squares and the outfit mean squares for the whole sample and for the four groups are one and the relevant values of the infit t-scores and the outfit t-scores are approximately zero. Thus, both sub-scales have satisfactory psychometric properties.

Statistics	Whole (n=712)	Boys (n=336)	Girls (n=376)	Grade 4 (n=314)	Grade 5 (n=398)
Mean (items)	0.00	0.00	0.00	0.00	0.00
(persons)	-0.18	-0.21	-0.16	-0.48	0.39
Standard deviation (items)	1.67	1.84	1.61	1.84	1.32
(persons)	1.19	1.45	1.06	0.96	1.22
Separability* (items)	0.99	0.99	0.99	0.96	0.99
(persons)	0.89	0.88	0.89	0.86	0.91
Mean Infit mean square (items)	1.00	1.00	0.99	1.00	1.00
(persons)	1.00	1.00	1.00	1.00	0.99
Mean Outfit mean square (items)	1.02	1.00	1.03	1.03	1.00
(persons)	1.02	1.01	1.04	1.02	1.01
Infit t (items)	-0.04	-0.05	-0.03	0.02	-0.02
(persons)	-0.01	-0.01	-0.01	-0.03	-0.01
Outfit t (items)	-0.05	-0.06	0.01	0.02	0.07
(persons)	0.04	0.08	0.05	0.05	0.02

Separability* (reliability) represents the proportion of observed variance considered to be true.

Table 2: Statistics relating to the scale for the whole sample and the four groups

Comparing the difficulties of the items of the two sub-scales, we can observe that the measurement model places a significant number of the items concerning identification knowledge at the easiest part of the scale and a significant number of items on elaboration knowledge at the harder part of the scale. In order to examine further this finding, the procedure for detecting pattern clustering in measurement designs developed by Marcoulides & Drezner (1999) was used. The cumulative D for the seventh cluster solution is 86% and the eighth gap adds only 2%. Moreover, all the gaps after the seventh gap are very small and this indicates that the 178 items are separable into seventh clusters. Thus, seven levels of PS skills based on ST can be identified. These levels are similar to the levels mentioned at the specification table of the test. More specifically, pupils who are at the first level (i.e. below –2.60 logits) are able to recognise verbally problems. The second level (-2.47 up to –1.78) refers to the diagrammatical recognition of problems. Pupils who are at the third level (-1.54 up to –0.85) are able to select a problem which reflects a given diagram. After the third level, there is a relatively big area where none item is included. This implies that there is a gap between the third and the fourth level (-0.24 up to 0.36) which refers to pupils' ability to pose questions from a mathematical situation. This gap can be attributed to the fact that although the skills of the first four levels refer to the first type of knowledge mentioned by Marshall (1995) pupils have to make an important progress in order to be able to pose questions. At the fifth level (0.45 up to 1.20), pupils are able to fill the data in a given diagram to represent the structure of a relevant problem and at the sixth level (1.50 up to 2.16) they are able to place the unknown quantity in the correct position of a diagram. Then there is a relatively big gap between the sixth and the seventh level (2.80 up to 3.62), which refers to pupils' ability to pose a problem based on a given diagram including either words and/or numbers.

DISCUSSION

The findings of this study provide support to the conceptual design of the proposed model of measuring PS skills. The underlying trait, that is pupils' abilities to use appropriate schemas to solve routine problems, seems to be an overarching concept comprised of the two types of knowledge mentioned by Marshalll and upon which the specification of the test was based. It would theoretically be expected that primary pupils would find it easier to develop skills concerning the identification knowledge rather than the elaboration knowledge. The findings of this study provide further support to this argument. Moreover, the procedure for detecting pattern clustering in measurement designs developed by Marcoulides & Drezner (1999) was found useful in supporting the theoretical background upon which the construction of the battery of tests was based. The seven levels of PS skills, which were identified, were similar to those described in the specification table of the test. However, this technique did not identify two different levels in relation to the problem posing skills concerning the elaboration knowledge, as it was expected. Moreover, a gap among the levels of each sub-scale was identified which revealed that problem posing skills are more difficult to be achieved than any other skill concerning each type of knowledge. This finding provides support to the assertion that problem-posing tasks are more complex and difficult for students than PS tasks (Silver, 1994; English, 1997).

The battery of tests on PS skills based on ST and its Rasch scale may help teachers decide how to identify and meet pupils' learning needs in relation to the seven levels of PS thinking and how to use their teaching time and their resources. An important implication of the identification stage is that it works as the first in sequence, facilitating pupils to make decisions, improve their abilities and move to the next stage (i.e. elaboration knowledge). Teachers should also be aware of the fact that the aforementioned two stages consisting of seven levels of thinking follow a linear sequential hierarchy. However, some pupils could be at the same level even though their abilities may differ. There is no clear distinction between consecutive levels, except between the levels concerning problem posing skills and those which are lower than them.

It goes without saying that further research is needed regarding the levels of PS skills based on ST. Specifically, further studies could explore whether the developmental model which emerged from this study and the seven levels, which were identified, may also derive from a study measuring pupils skills in solving problems of multiplicative structure. Furthermore, pupils' skills in solving two or three step-problems should be examined. Finally, there is a need to expand the model of measuring PS skills in order to refer to the next two types of knowledge (i.e. planning knowledge and execution knowledge). The findings of these studies may contribute in building a comprehensive model for PS skills based on ST.

REFERENCES

Adams, R.J. & Khoo, S.T. (1996). *Quest: The Interactive Test Analysis System.* Camberwell, Victoria: ACER.

English, L.D. (1997). Development of seventh-grade students' problem posing. In E. Pehkonnen (ed.) *Proceedings of the PME 21* (pp. 241-248). Finland.

Goldin, G.A. (1998). Representational systems, learning and problem solving. *Journal of Mathematical Behavior, 17* (2), 137-165.

Hambleton, R. K. (1995). Meeting the measurement challenges of the 1990s and beyond: New assessment models and methods. In T. Oakland, & R. K. Hambleton (Eds.), *International Perspectives on Academic Assessment* (pp. 83-104). Boston: Kluwer Academic Publishers.

Marcoulides, G.A., & Drezner, Z. (1999). A procedure for detecting pattern clustering in measurement designs. In M. Wilson, & G. Engelhard, Jr. (Eds.). *Objective measurement: Theory into practice (Vol. 5).* Ablex Publ.Corporation.

Marshall, S.P. (1995). *Schemas in Problem Solving.* New York: Cambridge University Press.

Mayer, R. (1992). *Thinking, Problem Solving and Cognition.* NY: Freeman.

Nesher, P., & Hershkovitz, S. (1994). The role of schemes in two-step problems: analysis and research findings. *Educational Studies in Mathematics, 26,* 1-23.

NCTM (2000). *Principles and Standards for School Mathematics*. Reston, V.A.

Philippou, G., & Christou, C. (1999). A schema-based model for teaching problem solving. In O. Zaslavski (Ed.), *Proceedings of the PME23* (57-64). Haifa: Technion–Israel Institute of Technology.

Rasch, G. (1980). *Probabilistic Models for some Intelligence and Attainment Tests.* Chicago: University of Chicago Press.

Schmidt, S. & Weiser, W. (1995). Semantic structures of one-step word problems involving multiplication or division. *Educational Studies in Mathematics, 28,* 55-72.

Silver, E.A. (1994). On mathematical problem posing. *For the Learning of Mathematics, 14* (1), 19-27.

Wright, B. D. (1985). Additivity in psychological measurement. In E. E. Roskam (Ed), *Measurement and personality assessment, (101-112).* Amsterdam: Elsevier Science.