TESTING A COMPREHENSIVE MODEL FOR MEASURING PROBLEM SOLVING AND PROBLEM POSING SKILLS OF PRIMARY PUPILS

Charalambos Charalambous, Leonidas Kyriakides, & George PhilippouDepartment of Education, University of Cyprus

The study reported in this paper is an attempt to develop a comprehensive model of measuring problem solving and posing (PSP) skills based on Marshall's schema theory (ST). A battery of tests on PSP skills was administered to 5th and 6th grade Cypriot students (n=2519). The Rasch model was used and a scale was created for the battery of tests and analyzed for reliability, fit to the model, meaning and validity. The analysis revealed that the battery of tests has satisfactory psychometric properties. The identified scale verifies previous findings suggesting that a number of variables are interwoven in the problem solving process. Yet, problem representation possesses a critical role in the process. The scale also suggests that achievement in posing problems is affected by the type of given information. The findings are discussed with reference to intended uses of teaching mathematics and suggestions for further research are drawn.

INTRODUCTION

Though problem solving (PS) has always consisted an integral part in mathematics education, it was only after the evolutionary work of George Polya that researches and mathematics educators realized the importance of elaborating on the process of solving problems. As a consequence, a number of models have been proposed to describe the cognitive elements involved in that process (i.e., Anderson, 1993; Mayer & Hegarty, 1996; Verschaffel, Greer & De Corte, 2000). Most of the aforementioned models provide for general approaches and strategies for PS, irrespective of the problem type. On the contrary, ST proposed by Mashall (1995), elaborates on routine problems presenting a comprehensive PS approach. ST aims to provide solvers with a number of cognitive schemata that can be used as guides during the PS process. It also employs the idea of using simple external representations (diagrams) which act as learning aids in retrieving and enhancing cognitive schemata (Goldin, 1998; Diezmann & English, 2001).

ST focuses mainly on the structure of the problems, providing five distinct problem structures (change, group, compare, restate and vary) that capture most routine problems that are usually presented to primary students. The former three problem structures can be used to solve additive problems, while the last two are mainly used for solving multiplicative structure problems. For each situation, Marshall (1995) proposed an appropriate diagram, which is expected to help students recognize the problem situation and solve the problem. Combinations of the above-mentioned structures could be helpful in solving more complex problems (two or three step problems).

Marshall (1995) also, identified four main elements (types of knowledge) involved in the PS process: identification, elaboration, planning and execution knowledge. The first type of knowledge refers to identifying the structure of a problem, and thus, can be considered as the most important part for schema activation. The second type of knowledge refers to recognizing the details that are distinct to each schema. Selecting the appropriate

diagram, placing data in it and drawing equations from it can be considered as elements of this type of knowledge. The planning knowledge refers to setting a solution plan for solving a given problem and it is usually conceived as unifying all needed decisions in order to arrive at a solution (thus, it includes elements of the two aforementioned types of knowledge). This type of knowledge is more prevalent in solving multiple step problems. Finally, the last type of knowledge includes executing algorithms.

The model described above was first introduced in upper elementary grades (4th to 6th) in Cyprus in 1998, with minor amendments. Specifically, only four problem structures were introduced, given that restate problems were embodied in comparison problems. Problem-posing (PP) activities were also included, since the significance of PP is nowadays well accepted (Silver & Cai, 1996). The present study builds on a previous study that investigated whether the first two types of knowledge mentioned in the model in relation to additive problem structures might help us form a developmental model measuring PS skills based on ST (Kyriakides, Philippou & Charalambous, 2002). In this paper, we report on testing a more comprehensive model including: (a) all problem structures, (b) one-step and multiple step problems (2 and 3 step problems), and (c) the former three types of knowledge, since execution knowledge refers mainly to executing algorithms. In this context, the main aims of this study were: (a) to develop a comprehensive model for measuring pupils' skills in problem solving and posing (PSP) one-step and multiple step problems, and (b) to collect empirical data in order to examine its validity.

THE DEVELOPMENT OF THE BATTERY OF TESTS ON PS

To answer our research questions, a battery of 48 tests on PSP was constructed guided by existing research and theory on assessment of PSP skills in Mathematics and by taking into account ST. Furthermore, a key requirement in designing the tests was its alignment with the mathematics curriculum that was operative in Cyprus. Thus, items were mainly based on ideas presented in ST as well as on activities included in the curriculum of Cyprus primary schools.

The specification table of the tests (Table 1) included fourteen levels of PSP skills related to three types of knowledge. Levels 1-3 referred to identification knowledge. Specifically, the first two levels included tasks examining the verbal identification of the schema needed for solving a problem (i.e., students were requested to identify the structure of a given problem or select a problem representing a given structure). The third level included tasks examining students' ability to select information and pose questions in order to produce problems of a given structure. The following four levels (levels 4-7) included tasks related to elaboration knowledge, which is linked to the use of diagrams. Namely items included choosing the correct diagram representing the structure of a given problem or selecting a problem that could be represented by a given diagram (4th level), placing the data and the unknown quantity of a problem in the correct position of a given diagram (5th level), setting equations for given diagrams (6th level), and posing problems based on specified diagrams (7th level). Items related to planning knowledge (levels 8-14) were similar to the above described, although they mainly referred to multiple step problems. Specifically, the items of the 8th level were similar to those of the 1st level (thus, these items included elements of the identification knowledge).

Types of knowledge	Levels	Items of the batte
Identification knowledge	*1. Verbal recognition of problems**	*1-12*
	*2. Selection of problems based on a given structure**	*13-24*
	*3. Posing problems of a given structure**	*25-40*
Elaboration knowledge	*4a. Diagrammatical recognition of problems**	*41-52*
	*4b. Selection of problems based on given diagrams**	*53-64*
	*5. Filling in data and unknown in given diagrams**	*65-100*
	*6. Setting equations based on given diagrams**	*101-127*
	*7. Posing problems based on given diagrams**	*128-151*
Planning knowledge	*8. Verbal recognition of problems** (I)*	*164-183*
	*9a. Diagrammatical recognition of problems** (E)*	*184-213*
	*9b. Selection of problems based on given combin structures** (E)*	*214-223*
	*10. Filling in data and unknown in given diagrams** (E)*	*224-263*
	*11. Setting equations based on given diagrams** (E)*	*264-338*
	*12. Posing multiple step problems** (E)*	*339-378*
	*13. Recognizing, representing and solving problems**	*152-163*
	*14. Recognizing, representing and solving problems***	*379-398*

* one-step problems, ** multiple step problems,

(I)=identification, (E)= elaboration knowledge is also prevalent

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

Similarly, levels 9-12 were analogous to levels 4-7 (thus, these items included elements of the elaboration knowledge). The remaining two levels referred to setting and carrying out all needed actions to solve either one-step problems (13th level) or multiple step problems (14th level). The specification table guided the construction of a battery of tests with 398 items, representing all levels. Levels 1-7 and 13 included tasks of all four problem structures, while levels 8-12 and 14 included combinations of the four problem structures.

METHODS

The items in the final version of the battery of tests were content validated by four experienced primary teachers, two mathematics textbooks writers, 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, amendments were made, particularly where terminology used was considered as unfamiliar to primary pupils. The final version of the battery of tests (available on request) was administered to all 5th grade (1184) and 6th grade (1335) pupils from 27 primary schools selected by stratified sampling (1298 of the subjects were boys and 1221 were girls). The Extended Logistic Model of Rasch (Andrich, 1988; Rasch, 1980) was used and the data were analyzed by using the Quest program (Adams & Khoo, 1996). The data were initially analyzed with the whole sample (n=2519) for all items together. The analysis was repeated with each of the four groups (grade 5, grade 6, boys and girls) of the sample, to investigate whether the battery of tests was consistently used by each group of the sample.

FINDINGS

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 equal or 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 close to 1 and that the values of the infit t-scores and the outfit t-scores are approximately zero. It can be claimed that there is a good fit to the model. The comparatively high value of outfit t-scores for persons can be seen as an indication of the relatively low separability of the persons scale and this can be attributed to the fact that the test was administered to children of a limited age span (only children of the two upper grades were included in the survey) and thereby the variation among their abilities was relatively low.

STATISTICS	Whole (n=2519)	Boys (n=1298)	Girls (n=1221)	5th grade (n=1184)	6th grade (n=1335)
Means (items)	0.00	0.00	0.00	0.00	0.00
(persons)	0.13	0.05	0.23	0.07	0.19
Standard deviation (items)	0.98	0.96	1.01	0.99	0.98
(persons)	0.97	1.00	0.95	0.95	1.01
Seperability* (items)	0.94	0.88	0.88	0.88	0.89
(persons)	0.86	0.86	0.85	0.85	0.86
Mean Infit mean square (items)	1.00	1.00	1.00	1.00	1.00
(persons)	1.00	0.98	1.00	0.98	1.00
Mean Outfit mean square (items)	1.02	1.02	1.03	1.03	1.02
(persons)	1.02	1.02	1.03	1.03	1.02
Infit t (items)	-0.05	-0.03	-0.02	-0.03	-0.03
(persons)	0.02	0.00	0.04	0.00	0.03
Outfit t (items)	0.03	0.04	0.05	0.05	0.04
(persons)	0.11	0.11	0.11	0.11	0.11

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

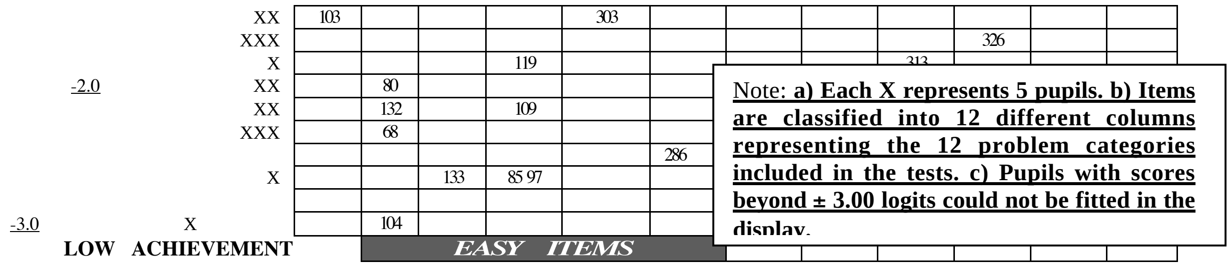
Table 2: Statistics relating to the scale for the whole sample and the four groups

Figure 1 illustrates the scale for the 398 test items with item difficulties and the whole group of pupils' measures calibrated on the same scale. The items appear in twelve columns. The first four represent the four problem structures (1=change, 2=group, 3=vary, 4=compare situation), while the remaining eight represent combinations of the four problem structures. Namely, these columns include items involving two additive structures, one additive and one multiplicative, one multiplicative and one additive, two multiplicative structures, three additive structures, two additive and one multiplicative, two multiplicative and one additive and three multiplicative structures (columns 5-12, respectively). Both figure 1 and the item fit map for the 398 items fitting the model reveal that all the items of the tests have a good fit to the measurement model.

.

	1	2	3	4	5	6	7	8	9	10	11	12
HIGH *ACHIEVEMENT*	*DIFFICULT ITEMS*											397
3.0				87				385			375	398

												376	
	X				86								
											373		
	X												
	X				74								
											353		
	XX												
		37				364							
	XXX					360 340			230				263
	XX	152				390							378
	XXXX				99			367				395	262
2.0	XX	89											
	XXXX								250				
	XXX				75 161		361			257			
	XXXXXX				63		255			371			
	XXXXXX				36 151	344	370					220	
	XXXXX				28	209	289	347				240	
	XXXXXX			32						392			222
	XXXXXX				30					391 217			223
	XXXXXXX	153				380			210	372		200 396	282
	XXXXXXXXX						355 350	386					182
	XXXXXXXXXXXX	65		39				212		351	394		336
	XXXXXXXXXX	90	31 35		64		288	387		237	393 354	356	358
1.0	XXXXXXXXXX	128		71			363 235		274		179 219	327	243 377
	XXXXXXXXXXXXX	33 77	38		150			171		216		261 330	242
	XXXXXXXXXXXXX	13 120				368		211		197		280	338
	XXXXXXXXXXXXXX				159	245	381	292	270 305			180	335 333
	XXXXXXXXXXXXXXXXX		5		40 76	348 225	275	294	365 170	352		281 332	
	XXXXXXXXXXXXXXXXXXXX	53	81	70 107	149	164 184	341		389 312	277 318	259		
	XXXXXXXXXXXXXXXX	29		158		169	175	172 192	190	177	199	221	203
	XXXXXXXXXXXXX	1		83	72 112	388 249	195		174	256 316	279		
	XXXXXXXXXXXXXXXXXXX	14 66			51 163	193 229 265	383			176		241	357 183
	XXXXXXXXXXXXXXX	78 114	69	19	139	165	268 266				278		
	XXXXXXXXXXXXXXXXXXX	41	17 45		12	189	208	251	234 254		238	328	202
	XXXXXXXXXXXXXXX				52 147	269		231 293		236		181 201	
0.0	XXXXXXXXXXXXXXXXXX	25			24 1 11	382 207		191					
	XXXXXXXXXXXXXXX	101		157	84 49	362 167	215	252					
	XXXXXXXXXXXXXXX	141 140		27	23 160	173 187 300		232 271	349	196			
	XXXXXXXXXXXXXXX	15 54	57		60 138	213	246 284	272	345				
	XXXXXXXXXXXXXX	2 42	113	18	115 110	359 204	226		214		276		337
	XXXXXXXXXXX	118	155	7	50	224 227	228	346			178		
	XXXXXXXXXX		144	59 146	11	339 185	188 285			315	322		
	XXXXXXXXXXXXXX		16	82 95	2 1 96	233 253	287						
	XXXXXXXXXX				9 10 48		168				218		
	XXXXXXXXXXXXX			47	61	342 267	166		194				
	XXXXXXXXX	55 129			87 388	247	343 186						
	XXXXXXXXXXXX			106	20 148	299	264					331	
	XXXXXX			6 58	137	273							
-1.0	XXXXXXXXXXXX	67 130		122 124		206	206		307	319			
	XXXXX	102	116		108				310				
	XXXXXXXXX		143		22 123	301		291		317			
	XXXXXX	43		94	62 127	304							
	XXXXXX		44	145	136 121				308				
	XXXX	117			125		290			314		329	
	XXXXXX	79 142				298		296					
	XXXX	3 91	4	46	100				311				
	XXXX										323		
	XX										321		



Moreover, pupils' scores range from –3.62 to +3.58 logits and the item difficulties range from -3.20 to +2.99 logits. This implies that the 398 items of the test are relatively well targeted against the pupils' measures, though a set of both more and less difficult items could be given to 19 students placed at the two opposite ends of the ability scale (12 pupils' scores were over +2.99 logits, and 7 pupils had lower scores than –3.20 logits).

The following observations arise from both Figure 1 and Table 1. Firstly, as concerns posing one-step problems (columns 1-4), items 25-40 (PP by selecting the needed data and posing a proper question to reflect problems of a given structure) are among the most difficult items of the test. In contrast, PP based on complete diagrams provides adequate guideline, and thus PP items of this type (items 128-151) turn out to be easier than items of the previous type and of many PS items, as well. In the case of multiple step problems, only the second type of PP was included (items 339-378). Figure 1 reveals that PP of this type is more difficult than solving problems of the analogous structure. There is only one exception in the 12th category (problems involving three multiplicative structures), where students had more difficulties in solving rather than posing problems.

As regards solving one-step problems, columns 1-4 reveal that the three types of knowledge cover a wide spectrum of PS abilities. At the one end of the spectrum (difficult items end) one may observe items related to the planning knowledge. This is more obvious for non-consistent problems (i.e., problems with inconsistency between their wording and the operation needed to arrive at a solution), such as items 152, 153, 159 and 161. The "difficult items end" is also occupied by items related to the 5th level (the second type of elaboration knowledge). Specifically, these items concern filling in the proper diagram in order to represent the structure of a given problem sufficiently. Items related to choice of the proper representation (items 41-64) appear somehow lower rather than the previous items, even lower to items related to identifying the structure of a given problem (levels 1-2). Items linked to setting the proper equations appear at the lower end of the scale, except of those connected to inconsistent problems (such as items 107, 111, 112, 114, 120). Finally, the distribution of items in columns 1-4 suggests that the problem structure interacts with the three types of knowledge, since change and compare problems cover a wider spectrum of abilities, in comparison to vary and group problems.

Regarding multiple step problems, columns 5-12 suggest that planning knowledge items (379-398) can be considered as lying at the hardest end of the ability scale, as in the case with one-step problems. Likewise, filling in the proper representation items (224-263) appear above items related to the identification of the problem structure (items 164-183) or to the selection of the most suitable representation (items 184-243). Moreover, items related to setting the correct equation for a given diagram (items 264-338) appear

somehow below items of the aforementioned levels. Finally, the distribution of items in the two final columns suggests that problems involving more than one multiplicative structure can be considered as more difficult than those involving mainly additive structures.

DISCUSSION

The findings of the present study provide support to results of relevant studies related to problem solving and posing (Mayer & Hegarty, 1996; Goldin, 1998; English, 1997; Silver & Cai, 1996; Kyriakides, Philippou & Charalambous, 2002). Analytically, achievement in problem posing seems to be influenced by the type of given information.

Complete diagrams aid the construction of problems in contrast to PP by selecting and combining given statements. However, in the case of multiple step problems, even though pupils were provided with complete diagrams, PP activities turned out to be harder than PS items. The distribution of items in Figure 1 also suggests that a number of variables are interwoven in the PS process. The structure of the problem, the cognitive processes involved in solving problems (i.e., types of knowledge), the consistency between the wording of the problem and the suitable operation, as well as the number of needed steps for solving a problem (one vs. multiple steps) are some of the variables affecting PS achievement. However, a relatively consistent pattern concerning the type of knowledge involved in the PS process emerges from Figure 1, both for one-step and for multiple step problems. Planning knowledge items are the most difficult, as it was expected, since achievement in these items demands the presence of the previous two types of knowledge. Using the correct representation properly also appears to be a critical element in the PS process. However, the selection of the proper representation is not sufficient in the PS process. Solvers need to place the given data and the unknown quantity in the correct position to form a complete representation that will guide the selection of the proper operation(s). Indeed, the present study suggests that setting the correct equation for solving a problem is of less importance than constructing a proper representation for a given problem.

It goes without saying that teachers should help students pay attention to the construction of proper representations. Teachers should also be aware that a number of variables are involved in the PSP process. Awareness of these variables can be helpful in both designing teaching interventions for eliminating related difficulties and measuring pupils' skills in PSP. Further research is also needed in order to specify the importance of each variable in the PS process. Item Response Theory models involving two or three parameters might be helpful in this direction since discontinuities in the levels of the specification table of the test can be assessed.

References

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

Anderson, J.R. (1993). Problem Solving and Learning. *American Psychologist, 48* (1), 35-44.

Andrich, D. (1988). A general form of Rasch's Extended Logistic Model for partial credit scoring. *Applied Measurement in Education, 1* (4), 363-378.

Diezmann, C., & English, L. D. (2001). Promoting the Use of Diagrams as Tools for Thinking. In A. A. Cuoco, & F.R. Curcio (Eds.), *The Roles of Representation in School Mathematics* (pp. 77-89). NCTM: VA, Reston.

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.

Kyriakides, L., Philippou, G., & Charalambous, C. (2002). Testing a developmental model of measuring problem solving skills based on schema theory. In A. C. Cockburn & E. Nardi (Eds.), *Proceedings of the 26th PME, vol. 3* (pp. 257-264). Norwich: University of East Anglia.

LeBlanc, M.D., & Weber-Russell, S. (1996). Text Integration and Mathematical Connnections: A Computer Model of Arithmetic Word Problem Solving. *Cognitive Science, 20,* 357-407.

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

Mayer, R.E., & Hegarty, M. (1996). The Process of Understanding Mathematical Problems. In R. Sternberg & T. Ben-Zeev (Eds.), *The Nature of Mathematical Thinking* (pp. 29-53). New Jersey: Lawrence Erlbaum Associates.

Rasch, G. (1980). *Probabilistic Models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Silver, E.A. & Cai, J. (1996). An Analysis of the Arithmetic Problem Posing by Middle School Students. *Journal for Research in Mathematics Education, 27* (5), 521-539.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making Sense of Word Problems*. Netherlands: Swets & Zeitlinger.

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