

# *Problem Complexity: A Measure of Problem Difficulty in Algebra by Using Computer*

FONG-LOK LEE

*Department of Curriculum and Instruction, The Chinese University of Hong Kong*

REX HEYWORTH

*Department of Educational Psychology, The Chinese University of Hong Kong*

*The present study describes how a measure of problem difficulty, called the problem complexity, can be developed. Four problem difficulty factors for a topic in mathematics were identified, namely, the perceived number of difficult steps, the number of steps required to finish the problem, the number of operations in the problem expression, and students' degree of familiarity with the question. These factors are called the complexity factors and can be used to calculate the problem complexity by using a multiple regression equation. This measure of problem difficulty has the advantage that it can be obtained when a problem is created and is thus suitable for use in computer-assisted instructional systems.*

## **Introduction**

It is generally agreed that students should be able to score higher in a test if the items or exercises are arranged according to their difficulty levels. The reasoning behind this is that students gain a sense of mastery after solving easier problems and are then more motivated to solve subsequent harder ones. Although the statistical measure of item difficult ratio provides a convenient measure of problem difficulty, it is however, not generally agreed that it adequately represents the degree of cognitive challenge an item is to students. Furthermore, with the increasing popularity of computer-assisted instructional systems, most of which would include

some kind of tests after didactic sections, arranging problems according to problem difficulty would be an important aspect of the instruction. Although the selection and arranging of problems can, in many cases, be done before the system is actually used, the items used would need to be tested to obtain their item difficulty ratios. However, in some cases (see for example, Lee, 1996), the computer system allows users, who may be teachers or students, to enter the problems which are then administered to students without any kind of pre-testing. If the computer system is to arrange the problems in terms of difficulty, this measure has to be calculated on the fly (when a problem is created). Measures such as item difficulty ratio are thus inapplicable in such a situation. The present study aims to identify different factors that affect the difficulty of mathematics problems and to develop a measure of difficulty which can be obtained without the need for testing among students.

## **Measures of Problem Difficulty**

### **Item Difficulty Ratio**

Traditionally, the difficulty level of a problem is measured by a ratio called the item difficulty ratio, which is the ratio of the number of respondents who answer correctly to the total number of responses to the problem (Gronlund, 1981). This gives a convenient measure of problem difficulty since it is not difficult to determine the two quantities required for the calculation once the problem has been administered to students. A problem is classified as easy or difficult according to a pre-determined proportion of students who can correctly solve the problem. However, a problem not solved correctly by all the students may not really be a difficult problem, and that a problem solved by all may not be an easy problem. The validity of using item difficulty ratio to represent problem difficulty can only be determined when the correlation between item difficulty ratio and problem difficulty can be clarified.

Concerning the validity of the item difficulty ratio, Mason, Zollman, Bramble, and O'Brien (1992) point out that,

This definition (difficulty in terms of item difficulty ratio) implies that an easy item also would be easy in terms of the cognitive challenge it presents to a respondent. Such a conclusion might be incorrect. Easy items might be answered correctly for the wrong reasons (e.g., there may be a wording clue that points to the correct answer, or the answer might be given on the basis of

automatic grammatical responding rather than thoughtful reply); similarly, a difficult item might not represent a difficult concept, but it might be so poorly phrased as they encourage incorrect responding. (p. 41)

Thus, there may be many reasons why a problem is difficult. Problems which do not require the student to understand difficult concepts or perform complex calculations may be difficult due to poor or misleading phrasing. Once the problem is rephrased or clarified, it may be easily solved. For example, Linville (1970) reported that students might solve more problems when the syntax of the problem statement is not complex. For problems which are difficult due to reasons such as poor phrasing, arranging them in terms of item difficulty ratio is meaningless. This assumption was supported by evidence given by Newman, Kundert, Lane, and Bull (1988) that students' scores are not improved in a test of multiple choice items when the items are arranged in increasing order of statistical difficulty (item difficulty ratio).

### **Cognitive Difficulty**

Other researchers have attempted to establish relationships between problem difficulty measures and variables that affect the problem solving process. For example, Mason, Zollman, Bramble, and O'Brien (1992) found that students needed a longer reaction time in mathematics exercises that required estimation or computation than those that required no computation. Lane (1991) found that difficulty levels of algebraic word problems depend on factors such as the number of assignments and relational propositions, the number of values that need to be derived, the amount of integration required for equation construction, whether the value of the unknown required manipulation to answer the question that was posed in the problem and whether the story context was familiar. These variables represent, in the words of Hornke & Habon (1986), the "cognitive demand of the item." It is thus natural to refer to this measure of item difficulty as the "cognitive difficulty" and the variables affecting the difficulty as the "cognitive variables."

Evidence for ordering problems in terms of cognitive difficulty was given by Newman, Kundert, Lane, and Bull (1988) who categorized 40 multiple choice examination items in educational psychology into three levels of Bloom's taxonomy: knowledge (15 items), comprehension (11 items) and application (14 items). Items belonging to different levels were considered by the researchers as of different levels of cognitive difficulty.

The results showed that students obtained higher scores in harder problems when the problems were arranged in increasing cognitive order (knowledge, comprehension, application). No such effect was found in medium and easy problems.

### **Complexity of Problems**

Holzman, Pellegrino, and Glaser (1983) grouped variables that affect the solution difficulty of number series completion into two categories: the processing dimension and the content-knowledge dimension. The processing dimension referred to the manipulation and management of rule-related information, while the content-knowledge referred to those that tied specifically to numerical skills. An example of the processing dimension is the amount of information to be coordinated in working memory and is measured by the number of memory placekeepers required. An example of the content-knowledge dimension is the facility of using arithmetic operators. Both were found to significantly contribute to the difficulty of number series problems. This suggests that the nature of cognitive variables that contribute to problem difficulty may differ: some variables are related to the subjects' general problem solving ability and others are related to content knowledge abilities. Although such a division is reasonable, it should also be true that both categories of variables represent the demand of the problem to the subjects, whether it is of general problem solving ability or of abilities related to content knowledge. This demand of problems, as reflected in the above examples, might be observable solely from the problem expressions. Variables such as the number of placekeepers required or the facility of operators may well be represented by how complex a problem is since a complex problem will require more memory placekeepers to hold the problem information as well as the use of more operators.

Regarding the complexity of problems, a number of researchers (Jerman, 1983; Lester, 1980; Silver & Thompson, 1984; Zweng, Turner, & Geraghty, 1979) have suggested that mathematics problems are more complex and more difficult to solve when they require several steps to obtain a solution, when subgoals must be reached before a solution can be obtained, and when the problems contain numbers that are of high computational complexity. Although there is no firm conclusion as to whether this complexity of problems is related to the item difficulty ratio (Hornke & Habon, 1986; Marzano & Jesse, 1987), the use of such complexity

factors seems to suggest a possible way of representing problem difficulty provided that these factors can be measured.

The use of such complexity factors as the above to represent problem difficulty has two further advantages. First, complexity factors are machine encodable and second, they can be obtained before a test is administered. Among the different problem complexity factors, the numerical complexity of an expression can be observed just by looking at the problem expression, while the number of steps and the number of subgoals can be obtained by having the computer solve the problems. Thus, measures of these factors can be obtained by just giving the problem expression to the computer. There is no need for a human expert to estimate the problem difficulty and also no need for the administration of tests to obtain the item difficulty ratio. Thus if we can identify the complexity factors and find ways to measure them, we are then in a position to develop a problem difficulty measure. The present study aims to identify the various factors that would contribute to the complexity of problems, and to develop a measure of problem complexity based on these factors.

### **Students' Perception of Problem Difficulty**

In order to develop a measure of the cognitive difficulty of problems, it is necessary that a precise measure of problem difficulty be developed to act as a frame of reference. If the problem difficulty measure is for the purpose of arranging problems in an order that maintains students' motivation, it is the students' perceptions of difficulty during the problem solving process that is important. Whether or not students' perceptions truly reflects problem cognitive difficulty was also investigated in the present study. To contrast the validity of the newly-developed difficulty measure, other measures of problem difficulty were also obtained. The details are described in the following sections.

### **Procedure and Results**

In order that the validity of the newly-developed problem complexity measure can be verified, the item difficulty ratios of the problems used were measured. In addition, teachers and students were required to estimate the difficulty of the problems. Collecting students' perceptions was done while the students were solving the problems. Based on the reasons described in the previous paragraph, students' perception is treated as the

frame of reference for other difficulty measures. On the other hand, as teachers frequently estimate problem difficulty when assigning exercises to students, their estimations would to a certain extent also reflect the difficulties of problems. Besides collecting these estimations, how teachers estimate the difficulty was also investigated in order to identify the problem complexity factors. Initially six problem complexity factors were assumed and teachers were asked to rate their relative importance. Teachers were also asked to suggest any further factors they would use to predict problem difficulty. The factors collected were used to estimate problem complexity and the problem complexity measure developed was compared to other difficulty measures. The following subsections describe how the different measures of problem difficulty were obtained and how the problem complexity factors were measured. All measures were done with a mathematics test paper containing logarithms problems selected to represent the different types of problems commonly found in textbooks. The test paper can be found in Appendix A.

### **Item Difficulty Ratio**

One hundred and twenty-five Secondary 4 (Grade 9) school students in Hong Kong participated in the test. The test was marked and the total number of correct responses for each question in the tests was counted. The item difficulty ratio for each question was calculated using this formula:

$$\text{Item difficulty ratio} = \frac{\text{Number of students who correctly answered the item}}{\text{Total number of students}}$$

The data obtained are shown in Table 1.

### **Students' Perceptions of Problem Difficulty**

In the mathematics test (Appendix A), together with each item, there is a five-point scale indicating how difficult the student thought the item was. This estimation scale ranges from 1 (very easy) to 5 (very difficult). Students were asked to mark on the scale as they were solving each problem. The mean estimations of item difficulty of the 125 students were then collected; the data are also given in Table 1.

### **Teachers' Estimations of Problem Difficulty**

Twenty-nine Hong Kong mathematics teachers completed a questionnaire

**Table 1. Table Showing Various Measures of Item Difficulty**

Paper	Questions	Item difficulty ratio	Students perception	Teachers' estimation	Predicted complexity (clevel)
1	1	0.82	1.40	1.63	-17.05
	2	0.69	1.80	2.85	-8.56
	3	0.60	2.22	2.78	1.83
	4	0.78	1.90	2.11	-12.18
	5	0.74	1.71	1.81	-12.18
	6	0.62	1.85	2.26	-8.68
	7	0.52	2.28	2.56	-6.16
	8	0.82	1.62	2.74	-11.27
	9	0.60	1.84	3.11	-0.22
	10	0.49	2.46	2.85	-4.56
	11	0.62	2.12	2.44	-8.41
	12	0.38	2.66	2.78	1.91
	13	0.63	2.38	3.63	-2.88
	14	0.70	1.87	3.11	3.058
	15	0.28	2.77	3.22	22.77
	16	0.71	1.94	3.70	-13.52
	17	0.57	2.36	3.56	6.672
	18	0.10	3.17	3.74	21.54
	19	0.54	2.54	4.00	11.68
	2	20	0.40	2.53	3.26
A1		0.43	2.28	3.33	-7.50
A2		0.78	1.97	2.96	-9.90
A3		0.73	1.86	2.56	-9.90
A4		0.55	2.46	3.48	-1.62
A5		0.70	2.33	3.67	15.18
A6		0.59	2.56	3.89	24.02
B1		0.34	2.87	3.44	-2.24
B2		0.61	2.44	3.23	-0.87
B3		0.64	2.45	3.11	0.50
B4	0.33	2.45	3.37	7.12	
B5	0.37	3.17	4.00	9.59	
B6	0.27	3.36	4.33	27.45	

containing the same items and the same estimation scale in the mathematics test papers but did not solve the problems. Most of the teachers who participated in this test were studying for the Diploma in Education<sup>1</sup> in The Chinese University of Hong Kong. All except one were part-time students having full-time jobs and had taught for several years. Colleagues of some of these student teachers also participated in the test. Their estimations of problem difficulty are given in Table 1, while the profile of the participating teachers is given in Table 2.

Based on the data in Table 2, the teachers who participated in the test had an average of 5.03<sup>2</sup> years of teaching experience and of these, 4.70

**Table 2. Profile of Teachers Participating in the Estimation of Problem Difficulty**

Teacher characteristics	No of teachers
Sex	
Male	25
Female	4
Age group	
20–25	4
26–30	14
31–35	5
36–40	2
>40	3
Education level	
Secondary	0
Post-Secondary	0
University degree	15
University degree + Diploma of Education	7
Master or above	6
Mathematics as major subject studied	
Yes	26
No	2
Teaching experience	
0–2 years	3
3–4 years	14
5–6 years	5
7–8 years	3
More than 9 years	4
Teaching experience (Secondary 3, 4, 5/Grades 8, 9 10)	
0–2 years	5
3–4 years	13
5–6 years	5
7–8 years	3
More than 9 years	3

years were in teaching Secondary 3, 4 or 5 (Grades 8, 9 or 10) mathematics in which logarithms are taught. Also, all 29 of the teachers were university graduates with 13 of them holding a Diploma of Education or a master's degree. Hence, these teachers have the background to rate the different factors as well as problem difficulty. The estimations of the teachers on the difficulty level of each item were averaged and are reported in column 5 of Table 1.

### Factors Affecting Problem Difficulty

The next part of the study was to contrast the above three measures of



problem difficulty with that obtained by the computer. To find ways for the computer to calculate this problem difficulty, knowledge on how human experts do this should provide valuable insight since experienced teachers practice this whenever they assign homework to their students. Thus, the questionnaire on teachers' estimations of problem difficulty not only asked teachers to predict the item difficulty, but also to identify factors they thought would be important in predicting problem difficulty. Six factors, called the complexity factors, were assumed to affect how teachers predict the problem difficulty. Each factor is separately described below.

*Perceived Number of Difficult Steps During the Problem Solving Process ( $f_1$ )*

This measure reflected whether the students would encounter any difficulties in the solving process. Difficult steps were assumed to be those at which students usually made non-trivial errors. Experienced teachers are normally capable of identifying these steps by just looking at the problem expressions and so estimate the problem difficulty. For computer systems, this can be done by counting the probability of occurrence of frequent errors. Frequent errors can be identified by having the computer system solve a set of sample problems and then choose the ones that occur frequently. For example, in a computer system called Electronic Homework (Lee, 1996), errors that happened more than five times in a test were considered as frequent errors. A list of these errors in the form of mal-rules (incorrect rules) is shown in Appendix B. For each question, the number of difficult steps was then counted. These are shown in Table 3.

**Table 3. Teachers' Rating on Importance of Factors Affecting Problem Difficulty**

Factor	Level of importance
( $f_1$ ) Perceived no. of difficult steps during the problem solving process	( $r_1$ ) 4.00
( $f_2$ ) No. of steps required to finish the problem	( $r_2$ ) 3.43
( $f_3$ ) Numerical complexity	( $r_3$ ) 3.86
( $f_4$ ) No. of occurrences of log'	( $r_4$ ) 2.96
( $f_5$ ) No. of operations in the question	( $r_5$ ) 3.21
( $f_6$ ) Students degree of familiarity with the question	( $r_6$ ) 3.93

*Number of Steps Required to Finish the Problem ( $f_2$ )*

This factor is defined as the number of steps that an expert would require to finish a problem. Since it is possible that there may be more than one solution path to each problem, it was decided to count the number of steps in the shortest paths. As the present study seeks to obtain a computer-generated level of problem difficulty, the number of steps required by a computer system also has to be considered. Hence, for each problem, the number of steps required by human experts (the participating teachers) was obtained. The number of steps required by a computer was determined in the computer system Electronic Homework (Lee, 1996). The results obtained were then compared to see whether there is any significant difference between the two and whether either one of them is a good predictor of problem difficulty. The results of this comparison are given in Table 3.

*Numerical Complexity ( $f_3$ )*

A measure of numerical complexity was also developed. An intuitive expression of numerical complexity would be the larger a number is, the more complex it should be since it is harder to do calculations involving larger numbers. However, to avoid using too detailed a scale, which might not be necessary, the numerical complexity was measured by assigning weights to the numerical values instead of using the numerical values themselves. Every value between one and ten was assigned a weight of 1, while decimals and numerical values greater than ten were assigned weights of 2. The sum of such weights then gave the value of the numerical complexity of the problems which are shown in Table 3.

*Number of Occurrences of "Log" ( $f_4$ )*

This factor is simply the number of logarithmic functions that can be found in the problem. Such numbers were counted and are listed in Table 3.

*Number of Operations in the Question ( $f_5$ )*

This again was obtained by simply counting the number of operations in the problem. An operation is any one of the following: addition, subtraction, multiplication, division, and exponentiation. The results of the counting are listed in Table 3.

*Degree of Familiarity of the Student to the Question ( $f_o$ )*

Students might find that some problems are more familiar than others and it is possible that they would find the familiar problems easier to solve. Students normally learn the topic of logarithms in three stages: firstly, the simplification of numerical expressions, secondly the simplification of expressions involving variables, and thirdly the solving of logarithmic functions. Further, knowledge learned at the earlier stages is used at later stages. It is therefore reasonable to assume that problems learned at earlier stages should be more familiar to students. This forms the basis for the value of the degree of familiarity assigned to each problem. For simplicity, all problems on the simplification of numerical expressions were assigned a value of 1, those on simplification of expressions involving variables were assigned a value of 2, and problems on solving of logarithmic equations were assigned a value of 3. Values assigned to the problems in the test can be found in Table 3.

**Teachers' Rating of the Relative Importance of the Complexity Factors**

Teachers were requested to rate the importance of each of these factors on a five-point scale before they estimated problem difficulty. Besides rating these suggested factors, the teachers were also required to add any other factors which they thought were important. However, no additional factors were suggested by the teachers. In order to validate whether these ratings actually reflect how teachers estimate problem difficulty, the weighted mean of the complexity factor measures, called *clevel*, was developed based on the following formula:

$$clevel = \frac{\sum_{i=1}^6 f_i r_i}{\sum_{i=1}^6 r_i}$$

where  $f_i$  is the value of the  $i$ th factor and  $r_i$  is the corresponding relative importance. Values of the relative importance of the complexity factors as rated by the teachers can be found in Table 3 whilst values of the complexity factors as well as the *clevel* values for each of the problems can be found in Table 4.

**Table 4. Problem Difficulty as Predicted by the Complexity Factors**

Paper no	Q no	Factors						Predicted complexity level (clevel)	
		$f_1$	$f_2$ (human)	$f_2$ (computer)	$f_3$	$f_4$	$f_5$		$f_6$
1	1	1	3	25	0	2	2	1	-17 05
	2	2	3	30	1	2	2	1	-8 56
	3	4	7	65	2	1	1	1	1 83
	4	1	3	30	1	2	2	1	-12 18
	5	1	4	30	1	2	2	1	-12 18
	6	1	4	50	1	2	2	1	-8 68
	7	3	6	40	2	1	1	1	-6 16
	8	2	3	35	1	2	1	1	-11 27
	9	2	3	15	2	3	3	1	-0 22
	10	2	5	30	2	2	2	1	-4 56
	11	3	6	50	1	1	1	1	-8 41
	12	2	3	50	1	3	3	1	1 91
	13	2	4	25	0	3	4	1	-2 88
	14	3	3	30	3	2	2	1	3 05
	15	3	6	80	4	3	3	1	22 77
	16	1	2	20	2	2	1	1	-13 52
	17	4	3	30	3	2	2	1	6 67
	18	5	6	38	2	4	4	1	21 54
	19	3	4	60	3	3	2	1	11 68
	2	20	1	4	50	1	1	1	1
A1		2	3	25	2	1	1	2	-7 50
A2		1	3	15	1	2	2	2	-9 90
A3		1	2	15	1	2	2	2	-9 90
A4		1	4	23	1	3	3	2	-1 62
A5		2	4	75	2	3	3	2	15 18
A6		2	4	40	4	4	4	2	24 02
B1		1	6	50	1	1	2	3	-2 24
B2		1	4	35	2	1	2	3	-0 87
B3		2	5	45	1	1	2	3	0 50
B4		2	4	60	2	1	2	3	7 12
B5		1	5	55	2	2	3	3	9 59
B6	4	9	95	2	2	3	3	27 45	

From Table 3, it can be seen that all the levels of importance were greater than 3, the mid-value, suggesting that the complexity factors were considered as quite important by the teachers. In addition, the correlation between teachers' estimations and clevel was 0.73 indicating that the teachers did use these complex factors to estimate the problem difficulty.

### Correlation Among the Various Measures of Problem Difficulty

Before discussing how a measure of problem difficulty can be developed,

it is worthwhile to note that the correlations among the various measures of problem difficulty mentioned above. Table 5 lists the correlation coefficients of the various measures of problem difficulty.

**Table 5. Correlation Coefficients Among the Measures of Problem Difficulty**

	Dratio	Testm	Sper	Ediff (clevel)
Dratio	1 00	– 53**	– 86***	– 61***
Testm	– 53**	1 00	74***	70**
Sper	– 86***	74***	1 00	73***
Ediff (clevel)	– 61***	73**	73***	1 00

*Note* Dratio = Item difficulty ratio Testm = Teachers' estimation of problem difficulty,  
 Sper = Students' perception of problem difficulty

\*  $p < .05$  \*\*  $p < .01$  \*\*\*  $p < .001$

In Table 5, with the exception of the estimated problem complexity (clevel) which was calculated from the various complexity factors, all the other measures of problem difficulty were collected either by directly measuring students' performances (item difficulty ratio) or through teachers' and students' self-reported estimations of problem difficulty (Teachers' estimation and Students' perception of Problem Difficulty). All four measures are highly correlated.

Among the correlations, the highest coefficient was between item difficulty ratio and students' perception (–0.86). As the negative sign in the coefficient represents the condition where the estimation is higher (more difficult), fewer students will answer the question correctly hence causing a low item difficulty ratio. This coefficient is also higher than those between item difficulty ratio and teachers' estimation (–0.53) and between estimated problem difficulty (clevel) and item difficulty ratio (–0.61). As item difficulty ratio has traditionally been used as the measure of problem difficulties, it appears that the student participants were able to better predict their own performances than their teachers. This result is not unreasonable since the estimation of problem difficulty by the students was carried out at the same time as they were solving the problems. It would be natural for these students to rate as difficult a problem they could not solve and to rate the others as easier, which would result in a high correlation between their ratings and the item difficulty ratio.

### **How Students Rate the Problems**

Although it might be true that students would rate as difficult problems that

they could not solve and vice versa, it is also possible that they could rate such problems according to a finer degree of difficulty. For example, an unsolved problem could still be rated as quite easy if the failure was brought about by the failure to remember a key formula. On the other hand, a solved problem could be rated as very difficult by a very bright student with very high self-esteem since he or she might think that only such students could solve the problem. In this way, students' perceptions of problem difficulty involve more than just seeing whether or not a problem was solved. Rather, there might be some reflection of the cognitive difficulty of the problem.

As students' perceptions of problem difficulty were found to be highly correlated with all the other difficulty measures, it would be reasonable to conclude that this measure would be the most acceptable one among all the others. Hence, if this measure can be predicted by using some observable properties from the problem expressions such as the complexity factors described in the previous section, it would be possible to obtain a measure of problem difficulty which could be used before a test was administered. The following sections describe how this measure can be predicted and how the measure of problem difficulty was developed.

### **Predicting the Problem Difficulty Measures**

Table 1 above shows the predicted problem difficulty (clevel) as well as the other three measures collected directly from teachers and students. The statistical method of multiple linear regression was employed to investigate the effects of predicting these four difficulty measures by using the complexity factors. The results of the predictions are shown in Table 6.

Table 6 shows the regression coefficients found in the analyses with the four problem difficulty measures as dependent variables and the six complexity factors as independent variables. Since there were two different sets of data regarding the number of steps required to finish the problem — one carried out by the machine and the other by human experts — there are two separate sets of results showing 6 factors denoting the machine and 6 factors denoting the human experts.

All six complexity factors assumed in the present study were able to predict the four measures of problem difficulty ( $p < .001$ ) though to different degrees of accuracy. Also, although the predicted problem difficulty was found to be predictable to a very high degree ( $R = 1$ ), it is not a valid prediction since it was calculated using the same complexity factors

**Table 6. Summary of Regression Coefficients Found**

Number of complexity factors	Problem difficulty measures			
	Item difficulty ratio	Students' estimate	Teachers' estimate	Predicted problem difficulty
6 factors (machine)	0.52***	0.81***	0.72***	1.00***
6 factors (human)	0.58***	0.80***	0.72***	0.98***
5 factors (machine)	0.52***	0.77***	0.72***	0.97***
5 factors (human)	0.58***	0.80***	0.72***	0.96***

\*  $p < .05$  \*\*  $p < .01$  \*\*\*  $p < .001$

that were used to predict it. For the other three measures, it would be interesting to look at the differences among them so that a suitable measure of problem difficulty can be chosen.

*About the Three Measures*

For the other three difficulty measures, namely, the item difficulty ratio, students' perception and teachers' estimation, it was found that when they were predicted by multiple linear regression with the six complexity factors as independent variables, not all of them needed to be included in the regression equation. Table 7 shows the complexity factors that appeared in the regression equations to predict the various difficulty measures.

While the item difficulty ratio could be predicted with just one variable, viz. the number of steps, to predict students' perception, three additional variables were required. These were the degree of familiarity, the number of operations and the number of perceived errors. This confirms the assertion made earlier that when students are rating problem difficulty, they do more than just rate as difficult those problems they could not solve and rate as easy those they could solve. The calculation of item difficulty ratio was based on the number of students who could complete the problem. Hence, the more steps required in a problem, the greater chance of making errors. That may explain why the number of steps alone can predict the item difficulty ratio. On the other hand, the fact that students' ratings of problem difficulty depend on three additional variables suggests that they based their ratings on firstly, whether the problem was familiar to them, secondly, how complex the problem looked, and finally whether or not there were perceived difficulties which would easily cause errors. The students' perceptions were found to be related more to the cognitive structure of a problem than is the item difficulty ratio.

**Table 7. Variables in the Equations to Predict the Problem Difficulty Measures**

Item difficulty ratio	Students' perception	Teachers' estimation	Predicted problem difficulty
Machstep (or Humanstep)	Familiar Machstep (or Humanstep) Notmfac	Familiar Nolog Numcomp	Machstep (or Humanstep) Familiar Notmfac Pererr Numcomp Nolog

Note Machstep = Number of steps required for the computer to finish the problem; Humanstep = Number of steps required for human expert to finish the problem; Familiar = Familiarity of the problem to the students; Notmfac = Number of operator in the problem expression; Pererr = Perceived no. of difficult steps during the problem solving process; Numcomp = Numerical complexity; Nolog = No. of occurrence of "log".

It has already been pointed out that arranging problems in order of cognitive difficulty would be more helpful to students. It is therefore possible that student estimation of problem difficulty might be the better measure when compared with item difficulty ratio which only represents the number of correct responses.

Another interesting point comes from the factors predicting teachers' estimations. This measure was found to depend on only three factors: degree of familiarity, occurrences of "log" and numerical complexity, all of which are easily observable just from the problem expressions. Those factors that required an in-depth study of the problems, such as the number of steps required to solve the problem, are not found in the equation. This revealed one important thing, viz. that when the teachers estimated problem difficulty, they based their judgments on some easily obtainable and superficial variables. That might explain why their prediction of the student's achievement was not as good as those predicted by the students themselves. Be that as it may, we cannot deduce that this is what teachers usually do when predicting problem difficulties. But as far as the present study is concerned, the teachers' estimations are not as good a measure of problem difficulty.

#### *Machine Steps or Human Steps?*

Separate regressions were done for the number of steps required by computers and by human experts, and the regression coefficients in predicting



students' perceptions were found to be approximately equal and highly correlated (0.81 and 0.80 for machines and human experts respectively). Hence, it is reasonable to say that using either one would yield identical results. For simplicity, only the machine counted number of steps was entered into the regression equation as the primary consideration is for machine use. However, it is believed that even if the number of steps counted by humans were used, accurate predictions could still be obtained.

**Problem Complexity**

A new variable was then developed to represent problem difficulty based on students' perceptions. It has been shown that students' perceptions can be predicted by several complexity factors which roughly correspond to the cognitive difficulty of problems. This predicted students' perception value could thus be a measure of the cognitive difficulty of a problem. As this predicted value should be different from the original value for students' perception, and as this measure was found to reflect the complexity of a problem, it was given the name of "problem complexity." The finding of this problem difficulty depended on how the students' perceptions could be predicted. Table 8 shows the result of multiple regression with students' perception as dependent variable and the six complexity factors as the independent variables. Problem complexity was then defined according to the regression equation obtained.

Table 8 shows that students' perception could be reasonably predicted ( $R = .81$ ) by four complexity factors: the number of steps, the number of operations, the degree of familiarity of the problems to the students and the

**Table 8. Summary of Multiple Regression Analysis for Variables Predicting Students' Prediction of Problem Difficulty (N = 125)**

Variable	B	SE B	Beta
Familiar	24	08	3 12**
Machstep	08	03	32*
Notmfac	16	06	31*
Pererr	13	06	31*
(Constant)	1 00	20	

Note  $R^2 = 65$

Machstep = No. of steps required for the computer to finish the problem, Familiar = Students' familiarity with the problems, Notmfac = No. of operators in the problem expression, Pererr = Perceived no. of difficult steps during the problem solving process

\*  $p < .05$  \*\*  $p < .01$  \*\*\*  $p < .001$

number of perceived errors. Hence, problem complexity, which is the predicted value of students' perceptions of problem difficulty, was constructed according to the regression coefficient shown in Table 8. The equation of problem complexity was then developed as follows:

$$\text{problem complexity} = 0.08 \times \text{Machstep} + 0.16 \times \text{Notmfac} + 0.24 \times \text{Familiar} + 0.13 \text{ Pererr} + 1.00$$

### Validity of Problem Complexity

In order to test whether this newly developed problem complexity measure can really reflect how difficult a problem is, correlation coefficients between this new measure were calculated with each of the other difficulty measures including item difficulty ratio, teachers' estimate and students' estimate. The correlation coefficients are given in Table 9 and a discussion of the implication is given in the following section.

### Discussion and Conclusions

Table 9 shows that the newly developed problem difficulty measure of problem complexity correlates significantly with the other difficulty measures and should, to a great extent, reflect the difficulty level of problems. Further, if the student's estimate is considered as nearest to the true measure of problem difficulty, we can see that among the three measures of item difficulty ratio, problem complexity and teachers' estimate, item difficulty ratio has the highest correlation coefficient with students' estimate while teachers' estimate has the lowest. Hence, although problem complexity cannot predict more accurately when compared with item difficulty ratio, it can predict better than teachers. Considering that the item difficulty ratio can only be obtained after the problem is tested, problem complexity seems to be the best choice if a measure of problem difficulty is needed

**Table 9. Correlation Among the Problem Difficulty Measures (N = 32)**

	Dratio	Testm	Sper	CompX
Dratio	1.00	-.53**	-.86***	-.64***
Testm	-.53**	1.00	.74***	.65***
Sper	-.86***	.74***	1.00	.81***
CompX	-.64***	.65***	.81***	1.00

*Note.* Dratio = Item difficulty ratio; Testm = Teachers' estimation of problem difficulty; Sper = Students' perception of problem difficulty; CompX = Problem complexity.

\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

immediately after the problem is created such as is required in computer-assisted instructional systems.

Although item difficulty ratio has the highest correlation with students' estimates, it was also found in the regression analysis that item difficulty ratio can be predicted by only one complexity factor, the number of steps required to solve the problem. On the other hand, when students' estimates are predicted, three additional factors including number of perceived errors, familiarity of the problem to the student and number of operations in the problem expression, are required. As more steps required to solve a problem lead to a greater chance for students to make errors, the item difficulty ratio thus cannot reflect how cognitively difficult a problem is but only reflects whether or not the students can solve a problem. It should be pointed out that the complexity factor scores used in the present study are rather rudimentary and were designed for investigative purpose only. If these scores can be further refined, it may be possible that a more accurate prediction can be made in the future.

The present study thus accomplishes its purposes of identifying the complexity factors which affect the cognitive difficulty of one kind of mathematics problem and of developing a problem difficulty measure, called problem complexity, based on the complexity factors. The items used in the present study involved simplifying logarithmic expressions and solving logarithmic equations, which deal with algebraic expressions and equations having one variable. The measure developed thus may also be applicable to algebraic problems in general, provided more detailed measures of the complexity factors can be devised. The complexity factors reported in the present study are by no means exhausted and if additional factors are found, prediction will be more accurate. For problems other than in algebra, further investigation is needed in order to identify the factors as well as to develop a problem difficulty measure.

## Notes

1. A postgraduate diploma course for training both pre-service and in-service university graduates to become qualified teachers.
2. In calculating teaching experience, 9 years or more were counted as 11 years for the sake of simplicity.

## References

Gronlund, N. E. (1981). *Measurement and evaluation in teaching*. NY: Macmillan.

- Holzman, T. G., Pellegrino, J. W., & Glaser, R. (1983). Cognitive variables in series completion. *Journal of Educational Psychology, 75*(4), 603–618.
- Hornke, L. F., & Habon, M. W. (1986). Rule-based bank construction and evaluation within the linear logistic framework. *Applied Psychological Measurement, 10*(4), 369–380.
- Jerman, M. E. (1983). Problem length as a structural variable in verbal arithmetic problems. *Educational Studies in Mathematics, 5*, 109–123.
- Lane, S. (1991). Use of restricted item response models for examining item difficulty ordering and slope uniformity. *Journal of Educational Measurement 28*(4), 295–300.
- Lee, F. L. (1996). *Electronic homework: An intelligent tutoring system in logarithms*. Unpublished PhD dissertation, The Chinese University of Hong Kong.
- Lester, F. K. (1980). Problem solving: Is it a problem? In M. M. Lindquist (Ed.), *Selected issues in mathematics education* (pp. 29–45). Chicago: McCutchan.
- Linville, W. J. (1970). The effects of syntax and vocabulary upon the difficulty of verbal arithmetic problems with fourth grade students. *Dissertation Abstracts International, 30*, 4310A.
- Marzano, R. J., & Jesse, D. M. (1987). *A study of general cognitive operations in two achievement test batteries and their relationship to item difficulty*. Washington, D.C.: Office of Educational Research and Improvement, Department of Education. (ERIC Document Reproduction Service No. ED 299321)
- Mason, E., Zollman, A., Bramble, W. J., & O'Brien, J. (1992). Response time and item difficulty in a computer-based high school mathematics course. *Focus on Learning in Mathematics, 14*(3), 41–51.
- Newman, D. L., Kundert, D. K., Lane, D. S., & Bull, K. S. (1988). Effect of varying item order on multiple-choice test scores: Importance of statistical and cognitive difficulty. *Applied Measurement in education, 1*(1), 89–97.
- Silver, E. A., & Thompson, A. G. (1984). Research perspectives on problem solving in elementary school mathematics. *The Elementary School Journal, 84*, 529–545.
- Zweng, J. J., Turner, J., & Geraghty, J. (1979). *Children's strategies of solving verbal problems*. Columbus, OH. (ERIC Document Reproduction Service No. ED 178359)

## Appendix A

### Mathematics Test Items Used for the Measuring of Problem Difficulty

#### PAPER I

Simplify

1  $\log 5 + \log 2$

2  $\frac{\log 16}{\log 64}$

3  $\log 216$

4  $\log 8 + \log 125$

5  $\log 60 - \log 6$

6  $\log 3 - \log 30$

7  $\log \frac{81}{2}$

8  $\frac{\log 27}{\log 3}$

9  $\log \frac{100}{10} \times \frac{\log 100}{\log 10}$

10  $\log 75 + \log \frac{4}{3}$

11  $\log \frac{9}{4}$

12  $\log \frac{3}{5} - \log 12 + \log 2$

13  $\frac{1 + \log 2}{2\log 2 + \log 5}$

14  $\log \left( \frac{10 \times 10}{10 \times 10} \right) + \log 20$

15  $\log \frac{7}{9} + \log \frac{3}{28} - \log \frac{1}{2}$

16  $\log \left( \frac{\log 10}{10} \right)$

17  $\frac{\log \sqrt{3}}{\log 243}$

18  $\frac{\log \sqrt{3} + \log 2}{2\log 2 + \log 5}$

19  $\frac{\log 1/4}{\log(\log 100)}$

20  $\log 0.6$

#### PAPER IIA

Simplify

1)  $\log \frac{\sqrt{x}}{x}$

2)  $\frac{1/2 \log x}{\log \sqrt{x}}$

3)  $\log x^2 - 2 \log x$

4)  $\frac{\log x}{\log x - \log \sqrt{x}}$

5)  $\frac{3\log \sqrt{x} + 1/2 \log x}{2 \log \sqrt{x}}$

6)  $\frac{\log \sqrt{x} + \log x^3}{\log \sqrt{x} - \log x^4}$

#### PAPER IIB

Solve for x

1)  $\log(x + 6) + 1 = 0$

2)  $\log(9x - 26) = 2$

3)  $\log(2x - 5) = 1$

4)  $\log(x^2 + 1) = 1$

5)  $\log \sqrt{x} = \frac{1}{3} \log x + 1$

6)  $\log(x - 3) - \log(x^2 - 9) + 1 = 0$

## Appendix B

### Frequent Errors Found in the Mathematics Test

Code	Rule	Freq.	Example
AA1	$\log(A \pm B) \rightarrow \log A \pm \log B$	64	$\log(2 + 3) \rightarrow \log 2 + \log 3$
AA2	$\log A \pm \log B \rightarrow \log(A \pm B)$	23	$\log 5 + \log 5 \rightarrow \log(5 + 5)$
AA3	$\log(A \times x) = B \rightarrow \log x = B/A$	0	$\log 3x = 6 \rightarrow \log x = 2$
AA4	$\log(A \times x) \rightarrow A \times \log x$	4	$\log 2x \rightarrow 2 \times \log x$
AA5	$\log \text{Exp} = -\log A \rightarrow \text{Exp} = -A$	14	$\log x = -\log 10 \rightarrow x = -10$
AA6	$\log \text{Exp} = 0 \rightarrow \text{Exp} = 0$	5	$\log(x + 1) = 0 \rightarrow x + 1 = 0$
AA7	$\log(A \times B) \rightarrow A \times \log B$	23	$\log 2x \rightarrow 2 \times \log x$
AA8	$A \times \log B \rightarrow \log(A \times B)$	15	$2 \times \log 5 \rightarrow \log(2 \times 5)$
AA10	$\log A \rightarrow A \times \log;$ $\log A \rightarrow \log \times A$	5	$\log 2 \rightarrow 2 \times \log$ $\log 2 \rightarrow \log \times 2$
AA12	$A \times \log B = \log C \rightarrow A \times B = C$	5	$2 \log x = \log 4 \rightarrow 2x = 4$
AB2	$\log(A + B) \rightarrow \log A \times \log B$ or $\log(A - B) \rightarrow \frac{\log A}{\log B}$	19	$\log(2 + 3) \rightarrow \log 2 \times \log 3$ $\log(x - 2) \rightarrow \frac{\log x}{\log 2}$
AB3	$\log(A \times B) \rightarrow \log(A + B)$ or $\log A \rightarrow \log(A - B)$	7	$\log(2 \times 3) \rightarrow \log(2 + 3)$ $\log 9 \rightarrow \log(9 - 4)$
AB4	$\log A \times \log B \rightarrow \log(A + B)$ or $\frac{\log A}{\log B} \rightarrow \log(A - B)$	7	$\log x \times \log x \rightarrow \log(x + x)$ $\frac{\log 2x}{\log x} \rightarrow \log(2x - x)$
AB5	$\log A + \log B \rightarrow \log A \times \log B$ or $\log A - \log B \rightarrow \frac{\log A}{\log B}$	31	$\log 5 + \log 2 \rightarrow \log 5 \times \log 2$ $\log 3 - \log 30 \rightarrow \frac{\log 3}{\log 30}$
AB6	$\log A \times \log B \rightarrow \log A + \log B$ or $\frac{\log A}{\log B} \rightarrow \log A - \log B$	23	No example found due to no chance $\frac{\log 16}{\log 64} \rightarrow \log 16 - \log 64$
AB7	$\text{Log.Exp.1} - \text{Log.Exp.2} \rightarrow \frac{\log.\text{Exp.1}}{\log.\text{Exp.2}}$	27	$\log x - 1/2 \log x = \frac{\log x}{1/2 \log x}$
AB8	$\frac{\text{Log.Exp.1}}{\text{Log.Exp.2}} \rightarrow \text{Log.Exp.1} - \text{Log.Exp.2}$	34	$\frac{\log x}{\log x - 1/2 \log x} \rightarrow$ $\log x - (\log x - 1/2 \log x)$
AB9	$\text{Log.Exp.1} + \text{Log.Exp.2} \rightarrow$ $\text{Log.Exp.1} \times \text{Log.E}$	10	$\log 7/9 + \log 3/28 \rightarrow \log 7/9 \times$ $\log 3/28$ $1/2 \log x + 3 \log x \rightarrow$ $\log x (1/2 \times 3)$
AB10	$\text{Log.Exp.1} \times \text{Log.Exp.2} \rightarrow$ $\text{Log Exp.1} + \text{Log}$	10	$1/2 \log x \rightarrow 1/2 + \log x$

Code	Rule	Freq.	Example
AB11	$\log x \rightarrow$ when not in an equation of the form $\log x = 0$	21	$-\log 10 \rightarrow -10$ $\log 1/10 \rightarrow 0.1$
AB15	$\log A \rightarrow \log(\text{value of } \log A)$	17	$\log 1/10 \rightarrow \log(-1)$
AB16	$\log A^n \rightarrow (\log A)^n$	5	$\log \sqrt[3]{3} \rightarrow \sqrt[3]{0.4771}$
AB19	$\log A \rightarrow \log A + \log B$	8	$\log 3/28 \rightarrow \log 3 + \log 28$
AB21	$\log A^n \rightarrow 1/n \log A$	5	$\log x^2 \rightarrow 1/2 \log x$
AC1	$\log(A \times B) \rightarrow \log A \times \log B$ or $\log \frac{A}{B} \rightarrow \frac{\log A}{\log B}$	97	$\log(2 \times 3) \rightarrow \log 2 \times \log 3$ $\log \frac{2}{10} \rightarrow \frac{\log 2}{\log 10}$
AC2	$\log(A + B) \rightarrow \log A + B$ or $\log(A - B) \rightarrow \log A - B$	17	$\log(2 + 3) \rightarrow \log 2 + 3$ $\log(x - 3) \rightarrow \log x - 3$
AC4	$\log A + B \rightarrow \log(A + B)$ or $\log A - B \rightarrow \log(A - B)$	10	$\log(x + 6) + 1 \rightarrow \log(x + 7)$ $\log 1/3 - 1 \rightarrow \log(1/3 - 1)$
AC5	$\log A \times \log B \rightarrow \log(A \times B)$ or $\frac{\log A}{\log B} \rightarrow \log \frac{A}{B}$	36	$\log 2 \times \log 5 \rightarrow \log(2 \times 5)$ $\frac{\log 1/4}{\log 2} \rightarrow \log(1/4 - 2)$
BA1	unable to reject roots that cause $\log(-ve)$	11	unable to reject the root $-3$ in $\log(x - 3) - \log(x^2 - 9) + 1 = 0$