MAX JERMAN AND RAYMOND REES

PREDICTING THE RELATIVE DIFFICULTY OF VERBAL ARITHMETIC PROBLEMS

I. INTRODUCTION

There is, as yet, no adequate theory for learning mathematics that can be used as a basis to predict the rate of learning or the relative difficulty of a given set of verbal problems. However, linear regression models have been used to predict the relative difficulty of a variety of types of arithmetic exercises including verbal problems (Suppes *et al.*, 1968; Suppes *et al.*, 1969). One of the purposes of these regression studies has been to identify and quantify in a clear and explicit way a set of structural variables that account for a significant amount of the variance in the observed error rate.

A basic assumption of this approach is that the structure of the arithmetic problem itself, to a large measure determines its difficulty level. This is not to say that student aptitude-interaction factors do not come into play, but until clear evidence is available concerning the existence and nature of any such factors, we believe the structural type of analysis will prove to be a more fruitful avenue for research and curriculum development. What we hope for eventually is to be able to formulate a clear set of rules or a formula for generating sets of arithmetic problems of a specified difficulty level. Curriculum developers would then be in a better position to control difficulty level when preparing instructional materials.

The central problem, for purposes of this paper, is to report the attempts made to identify and define a meaningful set of variables that can account for a significant amount of the variance in the difficulty level of problems solved correctly. First, a brief history is given of the development of a set of variables that appear to account for much of the variation in the difficulty level of word problems solved in a CAI context. Second, the variables that account for a significant amount of the variance in a CAI context were compared with problems solved in a paper-and-pencil testing environment. Ultimately, we would like to derive a formula for preparing word problem items of a specific degree of difficulty. The notation we use for the regression model itself, follows Suppes *et al.* (1966). Denote the jth variable of problem i by v_{ij}. The weight assigned to the jth variable is denoted by α_j. Let p_i be the observed proportion of correct responses on problem i for a given group of students. The purpose of the model is to predict p_i for each problem. The linear regression model in terms of the variables v_{ij} and the weights α_j

Educational Studies in Mathematics **4** (1972) 306–323.

is then

$$p_i = \sum_j \alpha_j v_{ij} + \alpha_0 . \tag{1}$$

This model, as stated, may not preserve probability since the estimated weighting and values for the variables are combined to predict p_i. Therefore, it has been the practice, in order to insure that the predicted p_i's will always lie between 0 and 1, to make the following transformation and define a new variable z_i.

$$z_i = \log \frac{1 - p_i}{p_i} . \tag{2}$$

Then the regression model becomes

$$z_i = \sum_j \alpha_j v_{ij} + \alpha_0 . \tag{3}$$

To take care of the case when the observed p_i is either 1 or 0, the following transformation was used:

$$z = \begin{cases} \log(2n_i - 1) & \text{for} \quad p_i = 0 \\ \log \dfrac{1}{2n_i - 1} & \text{for} \quad p_i = 1 , \end{cases}$$

where n_i is the total number of students responding to problem i. The reason for putting $1 - p_i$ in the numerator of Equation (2) is to make the variables z_i increase monotonically in difficulty. It is desirable that the model reflect an increase in difficulty directly rather than inversely as the magnitude of the variables v_{ij} increases.

II. DEVELOPMENT OF VARIABLES IN A CAI CONTEXT

The variables considered were many. First, we examined in detail the variables defined and used in the studies mentioned earlier. The variables tested in the Suppes *et al.* (1969) paper were as follows.

X_1 Operations: the minimum nnmber of different operations required to reach the correct solution;

X_2 Steps: the minimum number of steps required to reach the correct solution;

X_3 Length: the number of words in the problem;

X_4 Sequential: assigned a value of 1 if the problem is of the same type (i.e., can be solved by the same operation(s)) as the problem that preceded it, and 0 otherwise;

X_5 Verbal-cue: assigned a value of 1 if the problem contains a verbal cue to the operation(s) required to solve the problem, and 0 otherwise;

X_6 Conversion: assigned a value of 1 if a conversion of units is required to solve the problem, and 0 otherwise.

The 1969 study reported the results of the goodness of fit for 68 word problems which 27 gifted fifth-grade students solved as part of their daily CAI program. The regression equation for all 68 items was $z_i = -7.36 + 0.87X_{i1}{}^1 + 0.18X_{i2} + 0.02X_{i3}{}^1 + 2.13X_{i4} + 0.26X_{i5} + 1.42X_{i6}{}^1$, The multiple R was 0.67 with standard error of 1.75, and an R^2 of 0.45. The χ^2 value for the 68 items was 555.76, which indicated a rather poor fit for the model.

Perhaps one of the variables in the above set which requires a description for clarification is X_5, verbal cue. The following words were considered cues.

Word	Cue for
and	addition
left	subtraction
each	multiplication
average / or / each	division

The conversion variable was defined for the case in which the student was to recall a fact from memory, such as 1 week = 7 days. This conversion or substitution of units was a 0, 1-variable.

Several people in the Institute for Mathematical Studies in the Social Sciences at Stanford University (IMSSS) became interested in formulating variables and testing their goodness of fit on the data from the Suppes *et al.* (1969) study. In particular, Dr. Barbara Searle formulated and tested several variables using the data set from the 1969 study. The set of variables she tested consisted of a mixture of some of the above variables and some new variables which she formulated. The variables were defined as follows:

Operations: the minimum number of operations required to reach a correct solution (values range 1–4).

Steps: the minimum number of binary operations, steps, needed to reach a solution (value range 1–7).

Length: the number of words in the problem (value range 7–51).

Conversion: this factor is present if a conversion is required and the equivalent units are not given in the problem (a 0, 1-variable).

[1] Indicates significance at the 0.05 level.

Verbal cue:	the cue for each operation is as follows.	
	Operation	Cue words
	Addition:	added, altogether, gained
	Subtraction:	how much less, lost, left
	Multiplication:	each
	Division:	average
	If a cue word was present, the value was 1, otherwise 0,	
Order:	If the steps to solution were in order as given in the problem statement, the value was 1, otherwise 0.	
Formula:	If knowledge of a formula was required, the value was 1. otherwise 0.	
Average:	If the problem statement contained the word 'average' the value was 1, otherwise 0.	
Addition:	If the problem required addition, the value was 1, otherwise 0.	
Subtraction:	If the problem required subtraction, the value was 1, otherwise 0.	
Multiplication:	If the problem required multiplication, the value was 1, otherwise 0.	
Division:	If the problem required division, the value was 1, otherwise 0.	
Sequence:	If the problem was in unusual order,[2] the value was 1, otherwise 0.	

Three of the problems in the original set of 68 were deleted due to their high χ^2 values. The above variables were tested on the data from the remaining 65 problems. Of the 16 variables in the expanded set, 12 were entered by the step-wise regression program BMD02R. The value of the multiple R was 0.820.

After studying the weights of the variables, their contribution to the total R^2 and their definitions, we formulated three additional variables. Two of these S_1 and S_2, were sequential variables; the third was a memory variable. The definitions for the three additional variables are as follows.

Memory (M) is defined as the sum of:

C the number of conversions + knowledge of formulas,
D the number of numerals in the problem statement,
0 the number of different operations.

S_1 is defined as the number of displacements of order of operations in successive problems.

[2] A clear definition of the term 'unusual order' was not obtainable.

Examples:

$$\left.\begin{matrix}3+4\\3-4\end{matrix}\right\}S_1=1 \qquad \left.\begin{matrix}3+5\\4+6\end{matrix}\right\}S_1=0$$

$$\left.\begin{matrix}(3+4)\times 2\\(3+4)+2\end{matrix}\right\}S_1=1 \qquad \left.\begin{matrix}(3+4)\times 2\\(3\times 4)+2\end{matrix}\right\}S_1=2$$

S_2 is defined as the number of displacements between order of operations required to solve the problem and the order operation given in the problem statement.

The result of the stepwise regression using all the variables which contributed at least 0.01 to the increase in R^2 on the data from the set of 65 original problems is presented in Table I.

TABLE I

Results of stepwise regression on 65 word problems

A. Three additional variables alone

Step	Variable	R	Increase in RSQ	Last reg coefficient
1	19 memor	0.491	0.241	0.106
2	17 S_1	0.501	0.010	0.073
3	18 S_2	0.513	0.012	0.087

B. Over-all variables

Step	Variable	R	Increase in RSQ	Last reg coefficient
1	4 opers	0.657	0.431	0.199
2	8 vblcu	0.697	0.055	0.262
3	15 divis[a]	0.729	0.046	0.299
4	6 lengt	0.761	0.047	0.018
5	10 formu[a]	0.785	0.038	0.879
6	17 S_1	0.805	0.032	0.130
7	7 convr	0.825	0.032	0.534
8	18 S_2	0.835	0.016	0.127
9	14 multi[a]	0.838	0.005	−0.090
10	19 memor	0.840	0.004	−0.048
11	12 add[a]	0.841	0.002	0.063
12	16 seque	0.842	0.001	0.053

[a] Variables suggested by Dr. Searle.

In Part A of Table I the last three variables are presented alone. The total R for the three was 0.51. Part B shows the results of the stepwise regression using all variables. Only those variables are shown that contributed at least

0.01 to an increase in R^2. Although the value of the multiple-R term for all 12 variables was fairly high, 0.842, the number of variables entered did not increase with the addition of the three new ones. Rather, the three new variables entered ahead of others with an R of 0.842 ($R^2 = 0.709$).

A total of 19 variables had now been formulated and tested on the original data. In addition to those described above, the following were formulated and tested.

Operations 2: The sum of the following.

(1) The number of different operations.

(2) Add 4 if one of the operations is division.
Add 2 if one of the operations is multiplication.
Add 1 if one of the operations is addition.

Order 2: The sum of the following.

(1) S_1.

(2) Verbal cue necessary to establish a new order. One point for each direct cue missing for each step.

Recall: The sum of the following.

(1) One count for a formula to be recalled and a count for each step in the formula, e.g., $A = 2l + 2w = 3$.

(2) One count for each conversion to be recalled and used.

(3) One count for each fact from a previous problem to be recalled and used.

Verbal cue 2:	The set of cues was expanded. In addition, one count was given for each cue present in the problem.
Addition:	added, altogether, gained, total
Substraction:	how much less, lost, left
	how much larger ... than
	how much smaller ... than
	how much greater ... than
	how much further ... than
Multiplication:	each, times
Division:	average
Distractors:	This variable was defined as 1 count for each verbal cue, which was not a cue for an operation, but a distractor; for example, if the word 'average' was used but multiplication rather than division was the required operation.

A complete list of variables, by number, follows.

Variable	Name	Variable	Name
1	P(correct-observed)	12	Multiplication
2	Operations	13	Division

3	Steps	14	Sequence
4	Length	15	S_1
5	Conversions	16	S_2
6	Verbal Cue	17	Memory
7	Order	18	Operations-2
8	Formula	19	Order-2
9	Average	20	Recall
10	Addition	21	Verbal Cue-2
11	Subtraction	22	Distractor Cue

The summary table for the step-wise regression with 22 variables and 65 items is presented in Table II. It appears from observing the values of R and

TABLE II

Summary table for the stepwise regression on 65 items

Step Number	Variable Entered	Multiple R	RSQ	Increase in RSQ	F value for del.	Last reg coefficients
1 Oper.	2	0.657	0.431	0.431	47.732	0.182
2 Convr.	5	0.702	0.493	0.062	7.559	0.552
3 Lengt	4	0.740	0.548	0.055	7.488	0.027
4 Order 2	19	0.780	0.609	0.061	9.321	0.143
5 Divis	13[a]	0.805	0.648	0.039	6.516	0.384
6 S_2	16	0.821	0.674	0.026	4.739	0.179
7 Order	7	0.829	0.687	0.012	2.232	−0.140
8 Memor	17	0.835	0.697	0.010	1.873	−0.084
9 Dist	22	0.841	0.707	0.010	1.904	−0.305
10 Recal	20	0.844	0.713	0.006	1.105	0.146
11 Vblcu	6	0.846	0.715	0.002	0.414	0.104
12 Vblcu 2	21	0.848	0.719	0.003	0.629	0.135
13 S_1	15	0.850	0.722	0.004	0.641	0.048
14 Aver	9[a]	0.850	0.723	0.001	0.202	−0.155
15 Oper 2	18	0.851	0.725	0.002	0.261	−0.033
16 Sub	11[a]	0.852	0.726	0.001	0.197	−0.061
17 Seq	14	0.853	0.727	0.001	0.210	0.086
18 Steps	3	0.853	0.727	0.000	0.062	0.030

[a] Variables suggested by Dr. Searle.

the increase in R^2 with each step that there was relatively little gain after the tenth step in either Table I or Table II. In fact had we considered important only those variables whose contribution to the increase in R^2 was 0.01 or greater, then the first eight variables in Table I and the first nine variables in Table II would comprise the set of variables of interest. It is most interesting to examine the order of entry of the variables in each case. The following list may be useful for ranking the variables under discussion.

From Table I			From Table II		
Step		R	Step		R
1	Operations	0.657	1	Operations	0.657
2	Verbal Cue	0.697	2	Conversions	0.702
3	Division	0.729	3	Length	0.740
4	Length	0.761	4	Order 2	0.780
5	Formula	0.785	5	Division	0.805
6	S_1	0.805	6	S_2	0.821
7	Conversions	0.825	7	Order	0.829
8	S_2	0.835	8	Memory	0.835
			9	Distractor Cues	0.841

Perhaps the most we can do at this point, on the basis of the analysis thus far, is to note which of the variables (operations, length, division, S_2 the internal sequence variable, and conversions) appears to be the most robust. Memory and distractor cues may or may not play important roles in subsequent analyses.

III. ANALYSES OF PROBLEM VARIABLES ON PAPER AND PENCIL TESTS

All previous analyses were performed on CAI curriculum where the students indicated the operations to be performed, but did not actually perform the computations. It was of interest to determine if these same variables were applicable to problems solved with pencil and paper. Eleven of the 22 variables described above were tested on word problems solved off-line to see if their order of entry in the stepwise regression was at all similar to that found in on-line CAI context. The variables selected for testing were the following.

Variable	Name	Variable	Name
2	Operations 2	8	Length
3	Order 2	9	Verbal Cue (as defined by Dr. Searle)
4	Recall		
5	S_1	10	Conversion
6	Memory	11	Formula
7	S_2	12	Division (as defined by Dr. Searle)

These variables were first tested on a collection of problems selected for analysis from a set used by average fifth-grade students in a typical paper-and-pencil classroom setting. Of these problems, 19 were solved by students

(N=20) using paper and pencil as part of the instructional treatment provided by the Modified Wanted-Given Program developed and tested in Jerman (1971). The remaining 10 problems came from the two test scales P501 and F502 (N=161). It should be noted that the number of students solving the problems was not the same for each problem of the set, i.e., some problems were solved by 20 students and others by 161 students. This should not cause any difficulty since the variable under study was the probability of a correct response, hence variation in responses for different students was averaged out, and only possible differences between groups could influence the analysis by changing the expected probability of a correct response. An examination of the two groups indicated they could both be classed as average fifth graders.

The test scales P501 and F502 were each sets of five arithmetic word problems. To obtain the problems for these scales, 150 word problems in three different forms were pilot tested in two classes at fourth-, fifth-, and sixth-grade levels. Thirty of the problems were also pilot tested in one class at fourth-, sixth-, and eighth- grade levels to determine the relative difficulty of the problems. Five problems were selected for use in the scale P501. The problems were chosen as being representative of their level of difficulty in terms of probability correct in the pilot test, and their complexity in terms of the number of steps required for solution. F502 is a parallel form of P501 in which only the names of the persons or objects in each problem were changed.

The results of the stepwise regression using these variables are summarized in Table III. As can be seen, Length, the count of the number of words in the statement of the problem, entered first followed by Memory, S_2, S_1, and Verbal Cue. The total R after 9 steps was 0.77, R^2 =0.595. This was somewhat surprising, but gratifying. The variables that accounted for most of the

TABLE III

Summary table for 29 word problems administered in paper and pencil format, 11 original variables

Step number	Variable entered	Multiple R	RSQ	Increase in RSQ	F value for del.	Last reg coefficients
1 Lengt	8	0.549	0.301	0.301	11.633	0.021
2 Memor	6	0.594	0.353	0.052	2.069	0.159
3 S_2	7	0.667	0.445	0.092	4.167	−0.305
4 S_1	5	0.698	0.487	0.042	1.949	−0.138
5 Vblcu	9	0.740	0.547	0.061	3.082	0.578
6 Oper 2	2	0.755	0.570	0.022	1.150	0.069
7 Recal	4	0.766	0.587	0.017	0.863	−0.279
8 Convr	10	0.771	0.594	0.007	0.356	−0.080
9 Divis	12	0.771	0.594	0.001	0.038	−0.086

variance on-line were also effective, though at a lower level, in accounting for much of the variance off-line where students were doing the required computation by hand.

The situation must be recognized as artificial, however. The students in fact did not solve all 29 problems at one sitting; they solved only the first 19. The last 10 problems were given seven weeks apart, 5 at a time, as posttest and follow-up test scales. The sequential variable is not correct, therefore, for problems 19 and 25. Otherwise the variable values are the same as if students had taken the entire set of problems at one time. This cannot be interpreted to mean, however, that students solved all 29 items at one sitting. Only the values of the variables are the same.

In an attempt to improve the fit, two new variables were defined. The first, Verbal Cue-1 (No. 13) was a redefinition of Dr. Searle's variable. It was essentially the definition used in Verbal Cue-2, except that it was a O-1 variable rather than a frequency variable as is Verbal Cue 2. The second new variable (No. 14) was a combination of Verbal Cue 1 and indirect cues, such as 'in all', for addition, 'short of ..., 'for subtraction and 'per ...,' for multiplication. The results of the regression analysis on the 29 problems using 14 variables are shown in the summary Table IV. The increase in the value of R due to the addition of the two additional variables was small, almost 0.03 from 0.771 to 0.799.

Again, two new variables were added for testing, Verbal Cue-2 (No. 15) and a distractor variable (No. 16). These two variables are the ones described earlier. The results of the regression using all 15 variables are shown in

TABLE IV

Summary table for 29 word problems administered in paper and pencil form, 13 original variables

Step number	Variable entered	Multiple R	RSQ	Increase in RSQ	F value for del.	Last reg coefficients
1 Lenth	8	0.549	0.301	0.301	11.633	0.027
2 Memor	6	0.594	0.353	0.052	2.069	0.211
3 S_2	7	0.667	0.445	0.092	4.167	−0.218
4 S_1	5	0.698	0.487	0.042	1.949	−0.184
5 Vblcu	9	0.740	0.547	0.061	3.082	0.586
6 TCU	14	0.769	0.591	0.044	2.350	−0.076
7 Oper 2	2	0.787	0.620	0.029	1.577	0.033
8 Convr	10	0.796	0.634	0.014	0.797	−0.116
9 Vbcu	13	0.797	0.636	0.002	0.080	−0.145
10 Divis	12	0.799	0.638	0.002	0.100	0.158
11 Recal	4	0.799	0.638	0.001	0.034	−0.126
12 Ordr 2	3	0.799	0.639	0.000	0.013	−0.008

Table V. The value of R increased from 0.799 with 13 variables to 0.834 with 15 variables.

Clearly, the fit of the variables selected and tried thus far was less than satisfactory. The fact that the off-line students did all their computation by hand led to the definition of four new computational variables. These definitions followed the work reported in Suppes and Morningstar (in press). The variables were:

17. EXMC. A count of 1 was assigned for each multiplication exercise required in the solution of the problem. If multiplication was not required, 0 was assigned.

18. NOMC2. A count of 1 was assigned each time a regrouping occurred in each multiplication exercise in the problem. For example

$$\begin{array}{r} {}^{14}38 \\ \times\ 5 \\ \hline 190 \end{array} \quad \text{NOMC} = 2$$

$$\begin{array}{r} {}^{\overset{1}{14}}38 \\ \times\ 25 \\ \hline 190 \\ 76 \\ \hline 950 \end{array} \quad \text{NOMC} = 3$$

TABLE V

Summary table for 29 word problems administered in paper and pencil form, 15 original variables

Step number	Variable entered	Multiple R	RSQ	Increase in RSQ	F value for del.	Last reg coefficients
1 Lenth	8	0.549	0.301	0.301	11.633	0.041
2 Memor	6	0.594	0.353	0.052	2.069	0.171
2 S_2	7	0.667	0.445	0.092	4.167	−0.320
4 Vbcu 2	15	0.718	0.516	0.070	3.488	−0.284
5 S_1	5	0.746	0.557	0.042	2.167	−0.240
6 Divis	12	0.767	0.588	0.030	1.629	−0.240
7 Convr	10	0.780	0.608	0.020	1.089	−0.161
8 Vblcu	9	0.791	0.626	0.018	0.954	0.725
9 S_2	3	0.808	0.652	0.026	1.445	0.095
10 S_1	4	0.817	0.667	0.015	0.807	0.454
11 TCU	14	0.830	0.689	0.022	1.180	−0.182
12 Vbcu	13	0.833	0.694	0.005	0.284	0.121
13 Divis	12[a]	0.833	0.694	0.000	0.004	
14 Distr	16	0.834	0.696	0.002	0.079	−0.052
15 Oper 2	2	0.834	0.696	0.000	0.014	0.012

[a] Variable going out of the equation at this step.

19. COLC2. For this variable a count of 1 was given for each column and a count of 1 was given for each regrouping in addition and subtraction exercises. This count applied to only the largest exercise in the problem. If no addition or substitution was required, 0 was given.

20. QUOT. A count of 1 was given for each digit in the quotient if division was required and 0 otherwise.

Before additional analyses were performed, another problem (P(correct) = 0.50) was added to bring the total number of problems up to 30. This set of 30 problems were coded on the 19 variables indicated above.

A regression was run on the 30 problems using all 19 variables. The summary of this analysis is presented in Table VI. The value of R after the first five steps was 0.93, $R^2 = 0.87$. This is a surprisingly good fit for just five variables. Perhaps even more surprising is the strength of the computational variables as indicated by their point of entry into the regression program. Of the first five variables which entered the regression, three were computational variables: NOMC, a multiplication variable; QUOT, a division variable; and COLC, an addition and subtraction variable. The variable LENTH, which accounted for the number of words in the problem statement, entered first and the distractor variable DIST entered on the fourth step of the regression. The cognitive variables, such as memory and order, did not enter as soon or in the same order as when students solved problems at a CAI terminal.

TABLE VI

Summary table for 30 selected word problems administered in paper and pencil form, 19 original variables

Step number	Variable entered	Multiple R	RSQ	Increase in RSQ	F value for del.	Last reg coefficients
1 Lenth	8	0.668	0.447	0.446	22.609	0.017
2 Nomc 2	18	0.840	0.705	0.258	23.586	0.241
3 Quo	20	0.913	0.833	0.128	20.066	0.232
4 Distr	16	0.925	0.856	0.022	4.013	0.133
5 Colc 2	19	0.932	0.868	0.011	2.054	0.068
6 S_1	5	0.935	0.874	0.006	1.188	0.042
7 S_2	7	0.940	0.884	0.009	1.761	−0.167
8 Recal	4	0.943	0.889	0.005	1.011	−0.372
9 Exmc 2	17	0.946	0.895	0.006	1.206	0.089
10 Memry	6	0.947	0.897	0.001	0.303	−0.090
11 Oper 2	2	0.948	0.898	0.001	0.220	0.059
12 Ordr 2	3	0.949	0.900	0.001	0.329	0.058
13 Vbcu 2	15	0.950	0.902	0.002	0.362	−0.143
14 TCU	14	0.953	0.908	0.006	1.051	0.087
15 Convr	10	0.954	0.910	0.001	0.259	0.056
16 Vblcu	9	0.954	0.911	0.000	0.037	−0.071
17 Vcl	13	0.954	0.911	0.000	0.010	0.029

Only the first five variables contributed to an increase in R^2 of 1 percent or more. This is good because it is not practical to take into account more than four or five variables when writing word problems, even if additional variables were able to account for a larger portion of the variance than that indicated in Table V. Thus it is important that the optimal set of variables be found. The regression equation after the fifth step was

$$z_i = -0.73 + 0.02X_8 + 0.19X_{16} + 0.22X_{18} + 0.03X_{19} + 0.23X_{20}\,. \tag{4}$$

Table VII presents the regression coefficients, standard errors of regression coefficients, and computed t values for each of the five variables. Although the variables (X_{19})COLC did account for 1 percent of the variance, they were not sufficient to reach significance at the 0.05 level. In order to get a clearer picture of the relative importance of each weight, we normalized the coefficients to β weights. The normalized equation was

$$\hat{z}_i = C + 0.32X_8 + 0.21X_{16} + 0.64X_{18} + 0.11X_{19} + 0.40X_{20}\,. \tag{5}$$

This standardized form of the equation, with the beta weights as shown, permits direct comparison of the influence or weight of each variable. Multiplying each weight by its variables' correlation with the predicted probability correct indicates the percentage of the total variance accounted for by each term. The expression that gives this account is as follows.

$$PV_d = PV_8 + PV_{16} + PV_{18} + PV_{19} + PV_{20}\,.$$

PV_i represents the percentage of variance contributed by the *ith* variable. The computed values are as follows:

$$0.88 = 0.21 + 0.11 + 0.32 + 0.01 + 0.23\,.$$

In a stepwise regression sense these variables entered in the order 8, 18, 20, 16, 19. The analysis program was stopped at five steps because the addition of variable 19 only increased the percentage of variance accounted for by 1 percent. The expression indicating the proportion of total variance accounted for by each variable also indicated only 1 percent was accounted for in the presence of the other four variables. This is not always the case as can be noted in Table VI. The column headed 'Increase in RSQ' indicates the percentage of increase accounted for when the new variable was added. Table VII also gives a tabulation of the proportion of variance accounted for by each variable in the equation using five variables as predictors. It also includes the order of entry (stepwise) of the variable, the percentage increase in variance accounted for, the raw regression coefficients and the actual order of importance within the equation. This comparison shows how the importance of a contribution by a variable is adjusted in the presence of other

TABLE VII

Regression coefficients, standard errors of regression coefficients and computed T values

	Regression coefficient	Standard error	Computed T value	Variance from regression	Variance added step by step	Order of entry	Order of import within
X_8 Length	0.02	0.01	3.11[b]	0.21	0.45	1	3
X_{16} Distractor	0.19	0.09	2.10[a]	0.11	0.03	4	4
X_{18} Multiplication	0.22	0.03	8.15[c]	0.32	0.26	2	1
X_{19} Addition-Subtraction	0.03	0.02	1.48	0.01	0.01	5	5
X_{20} Division	0.23	0.06	3.84[c]	0.23	0.13	3	2

[a] $p < 0.05$ [b] $p < 0.01$ [c] $p < 0.001$

variables. This table shows why one must resist the temptation to use the raw regression coefficients, the step at which variable was entered, or the increase in variance accounted for by each step in determining the importance of the contribution of each variable in the final equation. For example, order of entry suggests X_8 as the most important contributor, when in fact X_{18} and X_{20} both contributed more than X_8 in the final equation. The correlation matrix for the variables is shown in Table VIII.

Variable 1 is the observed percentage correct and variable 7 is the trans-generated dependent variable. The mean and standard deviations for each variable are given in Table IX. The summary for the five-step regression analysis is given in Figure 1.

TABLE VIII

Correlation matrix for five variables

Variable number	1	2	3	4	5	6	7
1 Lenth	1.000	−0.637	−0.533	−0.413	0.001	−0.454	−0.926
2 Distr		1.000	0.610	−0.007	−0.205	0.606	0.668
3 Nomc 2			1.000	−0.145	−0.147	0.605	0.543
4 Colc 2				1.000	−0.090	−0.226	0.503
5 Quo					1.000	−0.118	−0.092
						1.000	0.566
							1.000

STEP NUMBER	5
VARIABLE ENTERED	5
MULTIPLE R	0.9315
STD. ERROR OF EST.	0.2325

ANALYSIS OF VARIANCE:

	DF	SUM OF SQUARES	MEAN SQUARE	F-RATIO
REGRESSION	5	8.508	1.702	31.485
RESIDUAL	24	1.297	0.054	

VARIABLES IN EQUATION: (CONSTANT = −0.73442)

VARIABLE		COEFFICIENT	STD. ERROR	F TO REMOVE
LENTH	2	0.01955	0.00626	9.7476 (2)
DISTR	3	0.18769	0.08924	4.4233 (2)
NOMC	4	0.22076	0.02705	66.6153 (2)
COLC	5	0.02819	0.01966	2.0545 (2)
QUOT	6	0.22983	0.05979	14.7775 (2)

VARIABLES NOT IN EQUATION:

VARIABLE		PARTIAL CORR.	TOLERANCE	F TO ENTER
P(C)	1	−0.81452	0.3018	45.3397 (1)

Fig. 1. Summary table for stepwise regression with five variables.

Table X presents the observed proportion correct for each problem and the proportion correct predicted by the model at each step for the first five steps of the stepwise regression. Predicted values were computed by taking the antilog of the log transform given by Equation (2). By comparing the summary information of Table X with Table VI, we can see which variables entered at each step. On step 1, LENTH entered. On step 2, both LENTH and NIMC were in the equation, etc. The operations required to solve each problem are shown in the second column of Table X. It is interesting to see the change in prediction for problems of various different operations as each new variable is entered.

TABLE IX
Means and standard deviations

Variable	Mean	Standard deviation
P(C) 1	29.597	19.557
Lenth 2	35.733	9.634
Distr 3	0.333	0.661
Nomc 4	1.467	1.676
Colc 5	2.133	2.255
Quot 6	0.433	1.006
7	0.510	0.581

In summary, we have retraced the steps taken in the definition of a small set of structural variables that accounted for a significant amount of the observed variance in error rate. Much work remains to be done. A final linear regression model described in this paper gave a surprisingly good account of the difficulty level of a somewhat artificially arranged set of verbal problems for fifth-grade students. Five variables were found to account for almost 87 percent of the variance in the observed probability correct. The variable that accounted for most of the variance was NOMC (32 percent), the multiplication variable, followed by QUOT (23 percent), the division variable, then LENTH (21 percent), the number of words in the problem statement, DISTR (11 percent), the verbal distractor variable, and finally COLC (1 percent), the addition-subtraction variable.

An Equation (4) has been derived for predicting the relative difficulty of verbal problems for fifth graders although it is not simple to implement. Before rushing off to apply these findings to a real-life situation, however, we need a good deal of replication to either confirm or deny findings reported here. We have made a start and are currently working with some 500 students in grades 4, 5, 6, 7, 8, and 9. Of course, much remains to be done. For example, we would like to know which variables enter first for students at

TABLE X

Observed proportion of students who correctly solved each item and the proportion predicted by the model at each step for five steps

Problem	Operations	Observed	Predicted				
			1 Variable	2 Variables	3 Variables	4 Variables	5 Variables
1	$a+b$	0.777	0.366	0.512	0.544	0.553	0.548
2	$a+b+c$	0.620	0.285	0.420	0.490	0.511	0.503
3	$(a\times b)-c$	0.560	0.433	0.481	0.462	0.460	0.455
4	$a\times b$	0.538	0.304	0.261	0.276	0.286	0.317
5	$a-(b+c+d)$	0.517	0.304	0.442	0.504	0.521	0.482
6	a/b	0.500	0.479	0.626	0.455	0.474	0.520
7	$a-(b+c+d)$	0.454	0.304	0.442	0.504	0.521	0.482
8	$(a\times b)\geqslant(c\times d)$	0.438	0.410	0.559	0.571	0.574	0.617
9	$((a-b)+(c-d))\times e$	0.387	0.266	0.226	0.255	0.269	0.285
10	$a\times b$	0.377	0.266	0.226	0.255	0.269	0.298
11	$(a+b)-a$	0.372	0.266	0.397	0.477	0.500	0.540
12	$((a-b)+(c-d))\times e$	0.355	0.285	0.243	0.266	0.278	0.294
13	$(a\times b)\geqslant(c\times d)$	0.341	0.388	0.536	0.557	0.564	0.606
14	$a\times b\times c$	0.310	0.285	0.176	0.182	0.189	0.211
15	$a\times(b+c)$	0.294	0.232	0.267	0.334	0.358	0.347
16	$(a\times b)+c$	0.266	0.388	0.254	0.225	0.224	0.216
17	$a\times b$	0.244	0.324	0.205	0.198	0.202	0.226
18	$a-(b\times c)$	0.241	0.616	0.566	0.448	0.421	0.387
19	$a-(b\times c)$	0.230	0.304	0.261	0.276	0.286	0.240
20	$a+(b\times c)$	0.174	0.285	0.176	0.182	0.189	0.145
21	a/b	0.166	0.098	0.164	0.195	0.128	0.140
22	$(a+b+c)\times d$	0.161	0.266	0.226	0.255	0.195	0.166
23	$a\times b\times c$	0.115	0.058	0.047	0.110	0.141	0.153
24	$(a+b)/c$	0.105	0.266	0.226	0.255	0.195	0.175
25	$(a-b)/c$	0.100	0.147	0.238	0.145	0.140	0.125
26	a/b	0.087	0.070	0.119	0.094	0.068	0.074
27	$a\times(b+c)$	0.071	0.045	0.036	0.095	0.087	0.083
28	$(a+b+c)\times d$	0.059	0.054	0.092	0.045	0.081	0.070
29	$a\times b\times c$	0.010	0.200	0.017	0.014	0.015	0.016
30	$(a\times b)/c$	0.010	0.054	0.043	0.010	0.008	0.009

different grade levels. We would also like to be able to compare variables for students solving problems on a CAI system that eliminates the need for computational facility with the variables that were found to account for most of the variance in this study where students were required to do their own computation. It appears that the variables for each situation may be different. It may be that regression techniques will provide a more suitable vehicle for determining whether aptitude-interaction factors do exist. In any case, the results from studies in progress should soon confirm or deny our hopes of being able to predict successfully the average difficulty of word problems for students using paper and pencil and working on-line with CAI systems.

The Pennsylvania State University
Stanford University

BIBLIOGRAPHY

Jerman, M., 'Problem Solving in Arithmetic as Transfer from a Productive Thinking Program,' Unpublished doctoral dissertation, Stanford University, 1971.

Suppes, P., Hyman, L., and Jerman, M., 'Linear Structural Models for Response and Latency Performance in Arithmetic on Computer-controlled Terminals', in J. P. Hill (ed.), *Minnesota Symposium on Child Psychology* **1** (1967), Univ. Minn. Press. Minneapolis, p. 160–200.

Suppes, P., Jerman, M., and Brian, D., *Computer-Assisted Instruction: the* 1965–66 *Stanford Arithmetic Program*, Academic Press, New York, 1968.

Suppes, P., Loftus, E., and Jerman, M., 'Problem Solving on a Computer-Based Teletype', *Educational Studies in Mathematics* **2** (1969), 1–15.