DOCUMENT RESUME

ED 070 678	24	SE 015 512

AUTHOR	Jerman, Max E.
TITLE	Predicting the Relative Difficulty of Problem-Solving Exercises in Arithmetic. Final Report.
INSTITUTION	Pennsylvania State Univ., University Park.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO	BR-2-C-047
PUB DATE	Dec 72
GRANT	OEG-3-72-0036
NOTE	65p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Curriculum; *Elementary School Mathematics; *Mathematics Education; Problem Sets; *Problem Solving; *Research; Secondary School Mathematics

ABSTRACT

The possibility of preparing a set of word problems of a predicted level of difficulty based on six variables (for multiplication, division, recall, conversions, operations, and number of words in the problem statement) and on regression equations developed in previous work was investigated. Four problem sets were used in grades four through six, and four different sets were given in grades seven through nine; a total of 340 students participated. The data indicated that the relative difficulty of the exercises was nearly the same over grade levels. Results showed that the general equation used in previous studies did not yield accurate predictions for the problems, based on a chi-square test. New equations computed for each grade level gave more accurate, though not significant, predictions. (Appendix B, pages 54-65, may be illegible.) (Author/DT)

Final Report

Project No. 2-C-047
Grant No. OEG-3-72-0036

Max E. Jerman
The Pennsylvania State University
University Park, PA 16802

PREDICTING THE RELATIVE DIFFICULTY OF PROBLEM-SOLVING EXERCISES IN ARITHMETIC

December 14, 1972

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

National Center for Educational Research and Development*
(Regional Research Program)

Author's Abstract

Several studies have been concerned with the definition of structural variables in word problem (problem solving) exercises in arithmetic with the view of identifying those variables which account for the most variance in P(correct) in a linear regression analysis. This study attempted to determine whether it was practical and possible to prepare a set of word problems of a predicted level of difficulty based on the variables and regression equations developed in previous work. Four sets of problems were given to students in each of Grades 4-6 and 7-9, eight sets in all.

The general equation from the previous did not yield accurate predictions for the problems in the present study using a chi-square test. However, when new equations, based on the old data, were computed for each grade level, the predictions were more accurate, though still not significant. Although the predictions were not as close as hoped, the means of the residuals was only 11 per cent, range 4-15 per cent. This is quite close for a first attempt. The data indicated that the relative difficulty of the exercises was nearly the same over grade levels indicating that the model should be capable of predicting more accurately, within 5 per cent, with further refinement.

Final Report

Project No. 2-C-047
Grant No. OEG-3-72-0036

PREDICTING THE RELATIVE DIFFICULTY OF PROBLEM-SOLVING EXERCISES IN ARITHMETIC

Max E. Jerman, Principal Investigator
Sanford Mirman, Associate Investigator

The Pennsylvania State University

University Park, Pennsylvania

December 14, 1972

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development*

Table of Contents

List of Tables

List of Figures

Introduction

Problem and Objectives

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 problem-solving exercises. However, the results of a series of studies, using linear regression models, seem to indicate that a small but manageable set of variables, which strongly influence the relative difficulty of verbal problems, have been identified and defined in such a way as to pursue a test of their adequacy by using them in the preparation of exercises of a specified level of difficulty in advance of their being solved by students.

The objective of this study was to test the set of above-mentioned variables by preparing sets of verbal problem-solving exercises having a predicted level of difficulty, in terms of a predicted probability correct, and to compare the predicted level of difficulty for each set of items and each item individually with the actual performance of students in public schools classrooms who attempt to solve the exercises.

The use of linear regression models to predict the relative difficulty of a variety of types of exercises including verbal problem solving exercises in well documented (Suppes, Jerman, and Brian 1968; Suppes, Loftus and Jerman, 1969; Jerman, 1971; Suppes and Morningstar, 1972). 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, the structural type of analysis may prove to be a more fruitful avenue for research and curriculum development. What is hoped 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.

Review of the literature

The purpose of this section is to review some of 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.

The regression model: The notation used here for the regression model itself, follows Suppes, Hyman and Jerman (1966) and denote the j^{th} variable of problem i by v_{ij}. The weight assigned to the j^{th} 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 is then

$$\hat{p}_i = \sum_j \alpha_j v_{ij} + \alpha_0 \qquad (1)$$

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.

$$z_i = \log \frac{1 - p_i}{p_i} \qquad (2)$$

Then the regression model becomes

$$z_i = \sum_j \alpha_j v_{ij} + \alpha_0 \qquad (3)$$

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

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

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.

<u>Development of Variables in a CAI Context</u>. The variables considered were many; those tested on a set of 68 word problems and reported in the Suppes, Loftus and Jerman (1969) paper were subsequently modified and retested on the same data base.

The modified set of 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).

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,* 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 .820.

After studying the weights of the variables, their contribution to the total R^2 and their definitions, three additional variables were formulated. 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,
O the number of different operations.

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

Examples:

$$\left.\begin{array}{ll}\text{First Problem} & 3 + 4 \\ \text{Second Problem} & 3 - 4\end{array}\right\} S_1 = 1 \qquad \left.\begin{array}{ll}\text{First Problem} & 3 + 5 \\ \text{Second Problem} & 4 + 6\end{array}\right\} S_1 = 0$$

$$\left.\begin{array}{ll}\text{First Problem} & (3 + 4) \times 2 \\ \text{Second Problem} & (3 + 4) \div 2\end{array}\right\} S_1 = 1 \qquad \left.\begin{array}{ll}\text{First Problem} & (3 + 4) \times 2 \\ \text{Second Problem} & (3 \times 4) + 2\end{array}\right\} S_1 = 2$$

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

The total R for the last three variables, Memory, S_1 and S_2, when used alone, for the set of 65 original problems was .51. Using all 19 variables in the stepwise regression produced a multiple-R of .842 for the 12 variables that contributed at least .001 to an increase in R^2. Using the last three variables did not increase the number of variables entered. Rather, the three new variables entered ahead of the others with an R of .842.

In addition to the 19 variables 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 = 21 + 2w (count = 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

Subtraction: 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

When the stepwise regression was applied using these 22 variables and the previous 19 variables, there was relatively little gain in R and R^2 after the tenth step. In fact, had one considered only those variables whose contribution to the increase in R^2 was .01 or greater, the first 8 variables from the set of 19 and the first nine variables from the set of 22 would comprise the set of variables of interest. It is most interesting to compare the order of entry of the variables in each case. The following list may be useful for this purpose.

19 Variables			22 Variables		
Step		R	Step		R
1	Operations	.657	1	Operations	.657
2	Verbal Cue	.697	2	Conversions	.702
3	Division	.729	3	Length	.740
4	Length	.761	4	Order 2	.780
5	Formula	.785	5	Division	.805
6	S_1	.805	6	S_2	.821
7	Conversions	.825	7	Order	.829
8	S_2	.835	8	Memory	.835
			9	Distractor Cues	.841

Perhaps the most that can be done 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.

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
4	Recall		
5	S_1	10	Conversion
6	Memory	11	Formula
7	S_2	12	Division

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.

When the stepwise regression, using the set of 11 variables, was applied, 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 .77 with $R^2 = .594$. The variables that accounted for most of the 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.

In an attempt to improve the fit, two new variables were defined. The first, Verbal Cue 1 (No. 13) was redefined. It was essentially the definition used in Verbal Cue 2, except that it was a 0-1 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 increase in the value of R, upon applying the stepwise regression, due to the addition of the two additional variables was small, almost .03 from .771 to .799.

Two additional 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 value of R obtained as the result of using 15 variables was .834 as compared to an R of .799 when 13 variables were used.

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 (1972). 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 regouping occurred in each multiplication exercise in the problem. For example:

*Variable 1 was the observed p(correct).

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

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

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 subtraction 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) = .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 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.

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 thus far. Thus is is important that the optimal set of variables be found. The regression equation after the fifth step was:

$$z_i = -.73 + .02X_8 + .19X_{16} + .22X_{18} + .03X_{19} + .23X_{20}. \qquad (4)$$

The raw regression coefficients for the variables X_8-Length, X_{16}-Distractor, X_{18}-NOMC2, X_{19}-COLC, and X_{20}-Division, were .02, .19, .22, .03, and .23 respectively. The percent of the total variance that was accounted for by each of the variables in the presence of the other four variables, was 21, 11, 32, 1, and 23 for variables X_8, X_{16}, X_{18}, X_{19}, and X_{20}, respectively. The percent of increase in variance accounted for by each step was 45, 3, 26, 1, and 13 for the variables X_8, X_{16}, X_{18}, X_{19}, and X_{20}, respectively. The order of entry, by steps, for the variables X_8, X_{16}, X_{18}, X_{19}, and X_{20}, was 1, 4, 2, 5, and 3, respectively. Comparing these results shows how the importance of a contribution by a variable is adjusted in the presence of other variables. These results show why one must resist the temptation to use the raw regression coefficients, the step at which the 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.

The final linear regression model described in this section gave a surprisingly good account of the difficulty level of a 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.

The first follow-up study was intended to replicate the previous study using as subjects students in Grades 4-9 with different achievement levels from different schools located in different socioeconomic-level communities. It was hoped that any differences among students due to either grade level, achievement level, or economic background might be evidenced by a different order of entry of the variables in the regression analysis for each group. The 19 variables and the same set of 30 problems used in the follow-up study were the same as those used in the previous study.

Two of the variables which accounted for a large portion of the variance in probability correct in the earlier study also accounted for a significant amount of the variance in the present study. In addition, the number of words in the problem statement, the ability to recall needed facts and perform needed conversion of units, and the number of different operations were variables that entered as one of the first six steps in a stepwise linear regression for the total population.

The importance of three of the variables, length ($p < .01$), multiplication ($p < .01$) and division ($p < .01$) is indicated by their level of significance in the regression equation for the total group. In terms of the results of the follow-up study, it appeared that the variables for multiplication, division, length, recall, conversion, and operations were the most important determinants of word-problem difficulty over all for students in Grades 4-9 when problems were solved using paper and pencil.

A second follow-up study examined the influence of the number of words in problem statements on error rate. Three forms of the 30-problem set used in the other studies were prepared in which the number of words in the problem statements were systematically varied, were administered to classes of students in Grades 4-8.

Three variables which accounted for a large proportion of the variance in the first follow-up study also accounted for a significant amount of the variance in the second follow-up study. The three variables were those for multiplication, division, and recall. A second set of three variables which also entered consistently among the first six in the linear regression in the second follow-up study were those for the number of words in the problem statement, conversions, and operations. Of particular interest was the variable for the number of words in the problem statement. The failure of the length variable to enter the regression consistently over all forms of the test sets leads one to conclude that it is not simply the number of words in the problem statement that influences difficulty, but the number of words in relation to other factors. This was indicated in the second follow-up study by the level of significance of the length variable on only one form of the test set, Form 2.

In summary, the six variables listed above, when used in a natural statement of a verbal problem statement have been found to account for a significant amount of variance in error rate in a series of studies with students at different grade levels from different socioeconomic backgrounds. The six variables used in this study were those for multiplication, division, recall, conversions, operations, and the number of words in the problem statements. The variables were defined above and were labeled, NOMC2, QUOT, Recall, Conversion, Operations, and Length, respectively.

The purpose of the present study was to determine whether it was possible to prepare a set of word problems in arithmetic in terms of the above-mentioned variables and using the regression equation specified earlier, for which the predicted level of difficulty was a close approximation of the actual level of difficulty as indicated by the proportion of students who were able to solve each exercise.

Methods

Subjects

Three hundred forty students in grades four through nine participated in the study; one hundred sixty-one were in elementary school in Pleasant Gap, Pennsylvania, and 179 were in junior high school in Bellefonte, Pennsylvania. Bellefonte and Pleasant Gap are small towns in low-middle class areas in central Pennsylvania. Two classes each of fourth-, fifth-, and sixth-grade students participated in the elementary school and two classes each of seventh-, eighth-, and ninth-grade students participated in the junior high school. The elementary school classes were of average ability. One class of seventh grade students was considered a high ability group. The other class was an average ability group. All students in each class took part in the study. The distribution of students by grades and classes is shown in Table 1.

Insert Table 1 About Here

Table 1

Distribution of Students by Grades

Class	Grade					
	4	5	6	7	8	9
1	29	23	32	36	31	30
2	26	27	24	33	30	19
Total	55	50	56	69	61	49

Construction of Problem Sets

Eight sets of arithmetic problem solving exercises, each containing twenty word problems were prepared during the month of April, 1972. Problem sets I-IV and sets V-VIII were prepared for user in grades four through six and seven through nine, respectively. The primary purpose for preparing two groups of exercises was to accommodate the higher computational abilities of students in the upper grades. The basic differences in the two groups of exercises were that division exercises in sets I-IV had no remainders and the use of fractions was not required; whereas, division exercises in sets V-VIII had remainders and computation with fractions was required. Problem sets I and IV consisted of problems requiring one or two operations for solution. Divisors were all single digits with one, two, or three digit quotients. Problem sets II and V contained problems requiring one or two operations for solutions. Divisors were one or two digits and quotients were one, two, or three digits. Problem sets III and VII consisted of problems requiring from one to three operations for solution. Divisors were multiples of 10 or single digits and quotients were one, two, or three digits. Sets IV and VIII contained problems requiring from one to four operations for solution. Divisors were one or two digits and quotients were one, two, or three digits. One problem sets I, II, V, and VI problems were randomly arranged with respect to operations necessary for solution. On Problem sets III, IV, VII, and VIII, the problems were arranged in order, beginning with those which required one operation for solution followed by those which required a greater number of operations. The operations required for the solution of each exercise in each problem set are given in Table 2.

Insert Table 2 About Here

Identical problem sets were used for each of the three elementary grades (4,5, and 6) and for each of the three grades in the junior high school (7,8, and 9). Each set of problems was mimeographed. The first page of each test booklet gave directions to the students for taking the problem set and the remaining five pages presented the problems, four per page, with work space provided.

The number of students to whom each problem set was administered is shown in Table 3.

Insert Table 3 About Here

The sets of problems were administered according to the time schedule given in Figure 1.

Insert Figure 1 About Here

Table 2

Operations Required in the Exercises

in Each Set

No. of Exercises	Operations Sets I,V	Operations Sets II,VI	No. of Exercises	Operations Sets III,VII	Operations Sets IV,VIII
3	+	+	1	+	+
3	-	-	1	-	-
2	x	x	1	x	x
2	÷	÷	1	÷	÷
1	+,-	+,+	1	+,+	+,+
1	-,+	-,-	1	-,+	x,+
1	x,+	÷,-	1	x,+	+,-
1	÷,+	+,x	1	÷,+	-,-
1	x,-	-,x	1	+,-	+,x
1	÷,-	x,x	1	-,-	÷,x
1	+,x	÷,x	1	x,-	-,÷
1	-,x	+,÷	1	x,x	÷,÷
1	+,÷	x.÷	1	÷,x	+,+,-
1	-,÷	÷,÷	1	-,÷	-,-,+
			1	x,÷	+,+,-
			1	÷,÷	x,-,÷
			1	+,+,+	-,-,+,x
			1	+,+,-	÷,x,-,+
			1	+,+,x	+,+,+,÷
			1	+,x,÷	+,+,÷,-

Table 3

The Number of Students Taking Each Problem Set

Grade	Problem Set							
	I	II	III	IV	V	VI	VII	VIII
4	52	53	55	53				
5	47	49	46	49				
6	54	56	49	46				
7					65	64	66	69
8					53	61	57	60
9					43	44	50	49

I. Wednesday, May 3, 1972
Grades 4,5, and 6

II. Thursday, May 11, 1972
Grades 4,5, and 6

III. Thursday, May 18, 1972
Grades 4,5, and 6

IV. Thursday, May 25, 1972
Grades 4,5, and 6

V. Friday, May 5, 1972 - Grades 7 and 9
Tuesday, May 9, 1972 - Grade 8

VI. Tuesday, May 9, 1972 - Grade 7 and 9
Tuesday, May 23, 1972 - Grade 8

VII. Tuesday, May 16, 1972
Grades 7,8, and 9

VIII. Tuesday, May 23, 1972 - Grades 7 and 9
Thursday, June 1, 1972 - Grade 8

Figure 1. Time Schedule for Administering Problem Sets

The reason the eighth grade classes did not always take each problem set on the same day as the seventh and ninth grade classes was that the eighth grade classes were on a field trip the day the first problem set, Set V, was initially given.

In the junior high school, the problem sets were administered during the regular 47-minute mathematics class period. In the elementary schools, the sets were administered at a time that was convenient to the teacher and which allowed the students ample time to finish the problems; this time was approximately 35-40 minutes.

A script was followed by the experimenters while introducing the sets of exercises to each class to attempt to standardize the testing situation. The text of the script is included as Appendix A. The experimenters answered all questions relevant to procedural difficulties and helped to clarify meanings of certain words or phrases but gave no assistance in solving any of the problems.

Problem sets I and V are included as Appendix B.

The eight sets of problems were coded for each of the six variables mentioned earlier. A small computer program was written to apply the following regression equation to each of the problems in each problem set to determine the predicted probability correct for the problem.

$$\hat{P}_i = -.96105-0.559X_2+0.0227X_3+0.218X_4-0.028X_5+0.199X_6+0.254X_7 \text{ where}$$

$\hat{P}_i$ is the predicted probability correct for the i^{th} problem in the problem set and X_2, X_3, X_4, X_5, X_6, and X_7 are the variables RECAL, LENTH, CONVR, OPER3, NOMC2, and QUO, respectively. An antilog transform was applied to the value of P obtained using the regression equation so that the predicted probability correct would be between zero and one. The regression equation above was derived from a study by Jerman (1971) in which the best over-all fit to the data for Grades 4-9 was given by this equation with the variables listed above.

Each of the problem sets was corrected by a project staff member. Problems which were omitted were not included in the analysis; i.e., the percent correct for a given problem was the percent of those students who attempted the problem who got the correct answer. A stepwise linear regression computer program, BMDO2R, (UCLA), was modified to include the aforementioned log and antilog transforms.

Results

The predicted and observed probability correct for each exercise in Problem Sets I-IV pooled over student in Grades 4-6 are shown in Table 4 along with the chi-square value of each exercise. The predicted probability correct was derived from the regression equation which gave the best over-

---

Insert Table 4 about here

---

all fit to the data for students in Grades 4-9 in a previous study, as mentioned in the previous section. The data from Set I in Table 4 are shown graphically in Figure 2. It is clear from inspection of both Table 3 and Figure 2 that the predictions are only fair, at best, in terms of chi-square.

Insert Figure 2 about here

The predicted and observed probability correct for each exercise in Problem Sets V-VIII and chi-square is shown in Table 5. The data from Set V

Insert Table 5 about here

---

in Table 5 are shown in Figure 3. It is obvious from inspection of the data

---

Insert Figure 3 about here

---

presented in Table 5 and Figure 3 that the prediction equation is in need of further refinement. The values for R presented in Table 6 for each test at

---

Insert Table 6 about here

---

each grade level and the total (T) are not low, but neither are they as high

Table 4

Predicted, Observed, and Chi-Square Values for Each Item

Pooled Over Problem Sets I-IV

Problem	Set I Pred.	Set I Obs.	Set I Chi-Square	Set II Pred.	Set II Obs.	Set II Chi-Square	Set III Pred.	Set III Obs.	Set III Chi-Square	Set IV Pred.	Set IV Obs.	Set IV Chi-Square
1	0.593	0.651	2.035	0.730	0.829	7.558	0.691	0.824	12.261	0.817	0.850	1.071
2	0.265	0.434	21.262	0.502	0.688	19.928	0.790	0.792	0.003	0.781	0.828	1.873
3	0.757	0.557	30.443	0.783	0.909	14.389	0.410	0.471	2.154	0.246	0.433	25.263
4	0.221	0.271	2.091	0.563	0.664	5.680	0.589	0.521	2.712	0.484	0.764	43.949
5	0.589	0.628	0.911	0.781	0.908	14.334	0.776	0.794	0.263	0.622	0.838	29.369
6	0.260	0.179	4.945	0.450	0.556	6.537	0.737	0.612	11.205	0.610	0.679	2.802
7	0.726	0.503	36.249	0.691	0.691	0.000	0.597	0.742	10.836	0.710	0.237	146.690
8	0.256	0.403	15.770	0.719	0.755	0.917	0.711	0.427	48.673	0.698	0.615	3.987
9	0.421	0.134	45.280	0.289	0.804	191.033	0.514	0.470	1.039	0.311	0.531	29.364
10	0.822	0.290	266.938	0.781	0.753	0.688	0.534	0.388	11.050	0.420	0.353	2.451
11	0.764	0.809	1.584	0.659	0.541	9.170	0.814	0.582	47.637	0.762	0.056	346.296
12	0.553	0.603	1.426	0.495	0.485	0.054	0.713	0.427	49.566	0.798	0.327	147.255
13	0.461	0.396	2.363	0.333	0.280	1.808	0.389	0.311	3.123	0.625	0.541	4.004
14	0.743	0.600	14.993	0.098	0.432	166.584	0.516	0.302	23.105	0.561	0.400	13.156
15	0.745	0.877	12.657	0.502	0.229	42.929	0.602	0.320	40.493	0.431	0.400	0.451
16	0.325	0.357	0.602	0.199	0.481	64.358	0.285	0.248	0.786	0.344	0.376	0.567
17	0.762	0.694	3.417	0.526	0.858	62.335	0.718	0.713	0.016	0.571	0.270	42.534
18	0.577	0.616	0.779	0.250	0.171	4.094	0.600	0.630	0.476	0.539	0.452	3.503
19	0.713	0.831	8.846	0.237	0.262	0.435	0.809	0.275	221.453	0.727	0.319	97.293
20	0.349	0.336	0.091	0.142	0.420	83.097	0.263	0.130	10.495	0.318	0.333	0.106

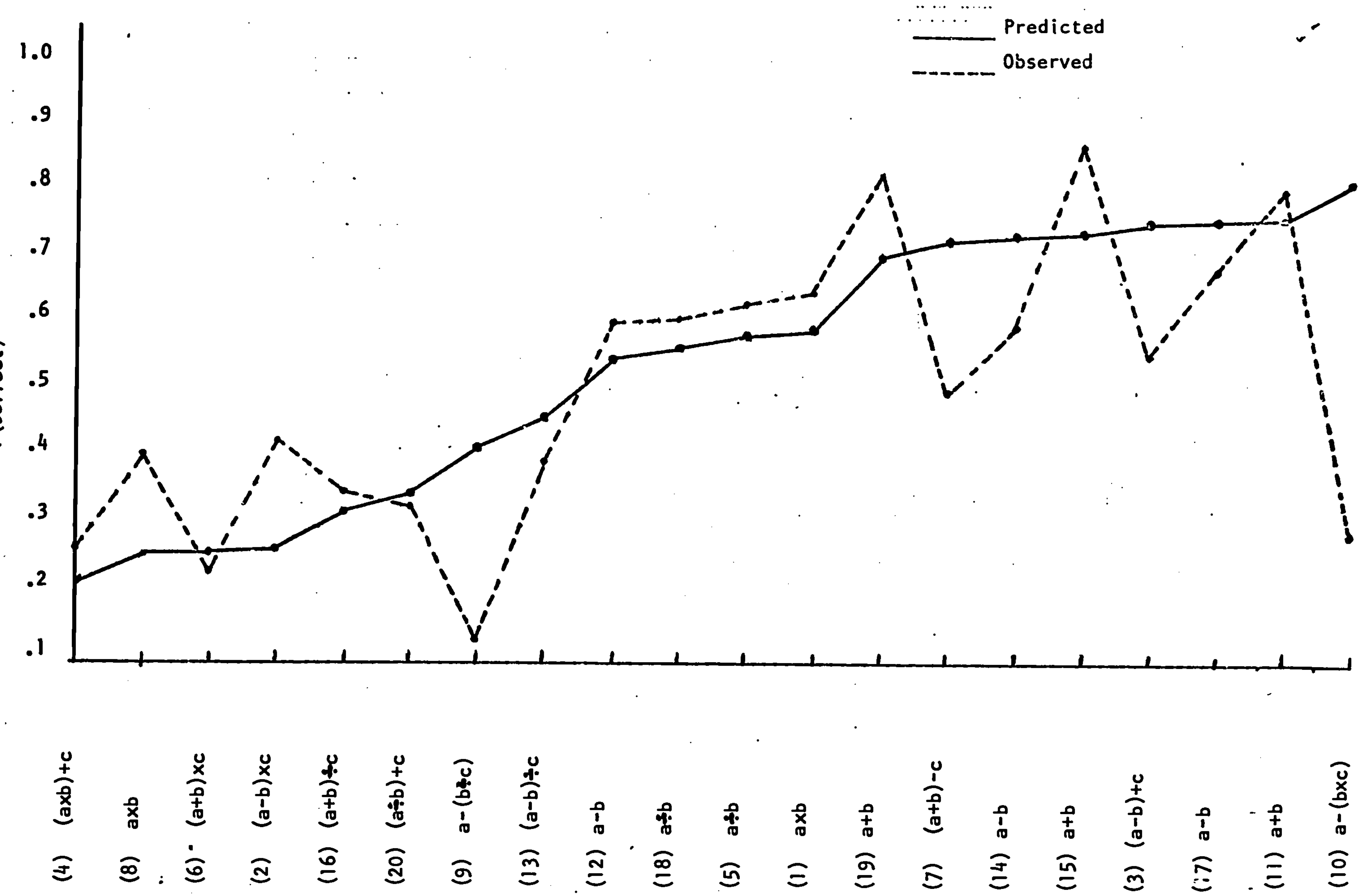

Figure 2. Observed and predicted probability correct for each exercise in Problem Set I with the required operations.

Table 5

Predicted, Observed, and Chi-Square Values for Each Item

Pooled Over Problem Sets V-VII

	Set V			Set VI			Set VII			Set VIII		
Problem	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square
1	0.246	0.706	182.528	0.750	0.470	68.574	0.691	0.873	26.838	0.702	0.904	34.719
2	0.743	0.870	13.599	0.594	0.790	26.602	0.732	0.919	30.838	0.654	0.702	1.812
3	0.829	0.211	382.573	0.410	0.702	59.216	0.270	0.826	269.769	0.211	0.677	217.836
4	0.203	0.742	285.509	0.741	0.761	0.340	0.355	0.690	83.810	0.431	0.673	38.686
5	0.888	0.615	116.901	0.444	0.934	161.452	0.138	0.659	387.916	0.148	0.598	279.430
6	0.028	0.090	21.892	0.497	0.372	9.750	0.694	0.765	3.940	0.401	0.865	152.375
7	0.357	0.377	0.277	0.075	0.640	754.636	0.440	0.764	68.592	0.260	0.202	2.850
8	0.513	0.846	69.242	0.884	0.794	13.033	0.588	0.601	0.110	0.847	0.825	0.575
9	0.080	0.321	125.474	0.594	0.879	55.573	0.237	0.473	52.052	0.210	0.413	41.482
10	0.671	0.836	19.609	0.442	0.428	0.126	0.281	0.297	0.209	0.603	0.253	82.898
11	0.380	0.580	26.655	0.202	0.702	249.696	0.583	0.753	20.209	0.499	0.473	0.446
12	0.309	0.318	0.057	0.094	0.595	495.140	0.316	0.552	42.002	0.284	0.715	144.338
13	0.237	0.226	0.104	0.421	0.876	136.738	0.153	0.315	28.959	0.116	0.154	2.197
14	0.713	0.917	31.929	0.076	0.255	67.984	0.701	0.543	19.533	0.346	0.404	2.170
15	0.597	0.673	3.673	0.262	0.558	74.767	0.129	0.179	3.471	0.324	0.474	15.615
16	0.867	0.797	6.502	0.267	0.185	5.016	0.416	0.265	15.204	0.344	0.789	145.668
17	0.755	0.940	27.939	0.516	0.752	35.905	0.650	0.763	9.486	0.431	0.659	36.035
18	0.514	0.314	21.937	0.732	0.700	0.783	0.082	0.381	184.084	0.175	0.423	66.456
19	0.620	0.827	25.280	0.709	0.681	0.619	0.844	0.461	183.829	0.985	0.582	1813.705
20	0.508	0.458	1.310	0.252	0.529	56.174	0.058	0.276	132.214	0.959	0.588	535.595

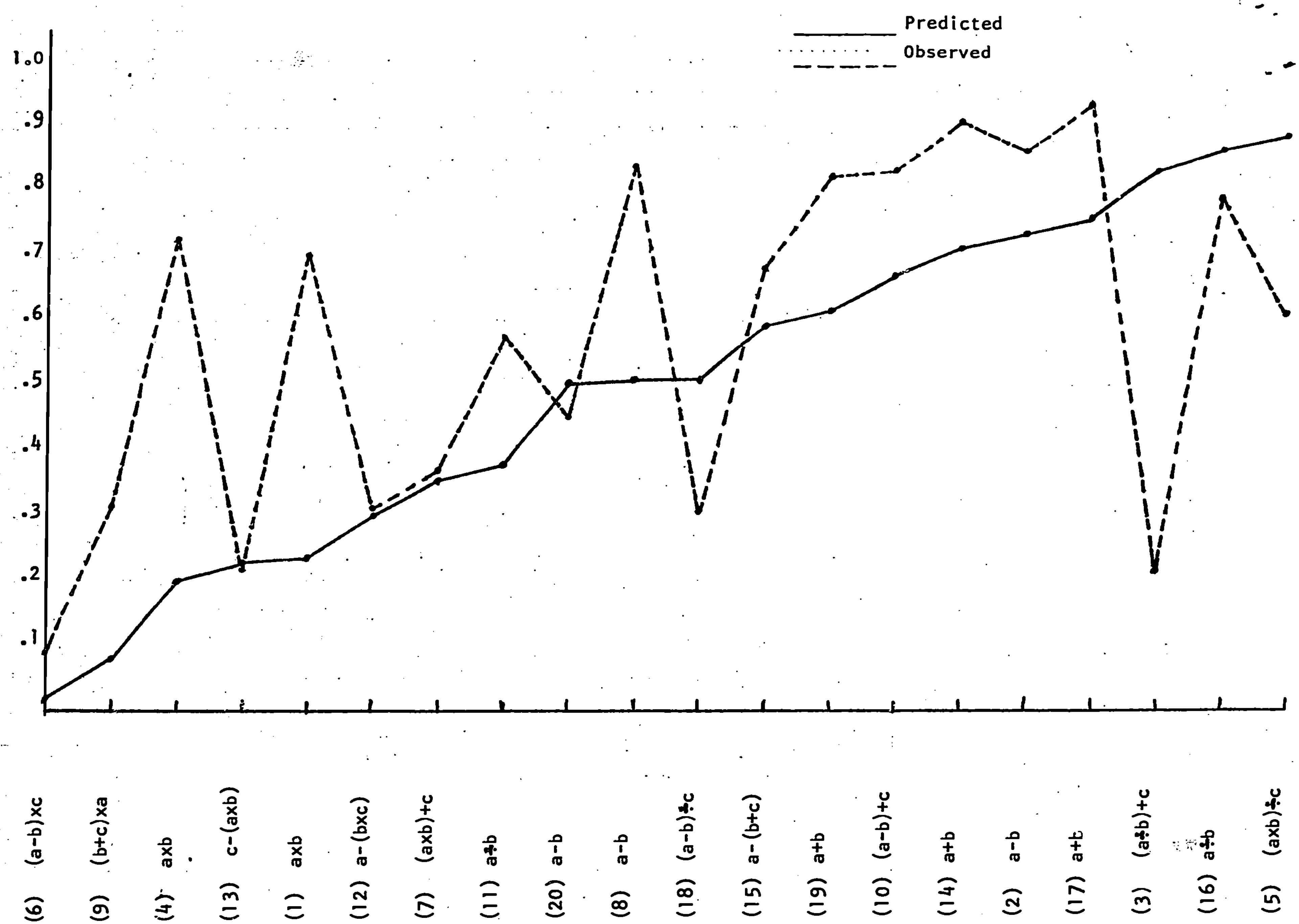

Figure 3. Observed and predicted probability correct for each exercise in Problem Set V, with the required operation.

Table 6

The Order of Entry of Each Variable For Each

Problem Set at Each Grade Level, R and R^2

	GRADE 4					GRADE 5					GRADE 6				
Step	1	2	3	4	T	1	2	3	4	T	1	2	3	4	T
1	5	5	5	5	5*	5	5	5	5	5	5	5	5	5	5
2	6	6	6	6	6	6	6	6	7	6	6	6	6	7	6
3	7	7	2	7	3	2	3	7	6	3	2	3	3	6	3
4	3	3	7	2	2	3	7	2	3	4	7	7	7	3	7
5	2		3	4	4	7			4	2	3			4	2
6				3	7				2	7					
R_6	.85	.89	.88	.82	.82	.81	.81	.83	.79	.73	.76	.78	.87	.78	.73
R^2_6	.72	.78	.77	.68	.67	.66	.66	.69	.63	.53	.57	.61	.76	.60	.54

	GRADE 7					GRADE 8					GRADE 9				
Step	5	6	7	8	T	5	6	7	8	T	5	6	7	8	T
1	5	7	7	6	7	5	2	7	3	7	5	7	7	6	7
2	3	6	5	3	6	3	3	2	6	3	3	3	2	3	3
3	7	2	6	4	3	6	7	5	4	6	7	6	5	7	6
4	6	5	2	2	5	7	6	6	7	5	6	2	3	2	5
5	2		3	7	2	2	5	3	2	4	2	5	6	4	4
6										2				5	
R_6	.83	.61	.74	.56	.58	.82	.42	.72	.65	.54	.80	.58	.77	.67	.61
R^2_6	.68	.38	.54	.32	.34	.67	.18	.52	.43	.29	.64	.34	.59	.44	.37

*2 - Recall
3 - Length
4 - Conversion
5 - Operations 3
6 - Nomc 2
7 - Quo

as the corresponding values in previous studies. The order of entry of each variable is also shown in Table 6. These data do indicate, however, that the model is still accounting for a significant amount of the observed variance.

The data from the previous study were reexamined and new regression equations for each grade level were computed since the over-all predictions were less accurate than hoped. As one might expect, the variables which gave the best general account of the observed variance for all grades 4-9 in the previous study were not necessarily the same set of variables which gave the best account of the observed variance at any particular grade leve. Therefore, the Problem Sets I-VIII were recoded in terms of the six variables which gave the best account of the observed variance is probability correct for each grade level in the previous study. The new regression equations were used to generate predictions for all four tests administered at each grade level. These predicted and observed probabilities and chi-square values for each exercise on each test, by grade level, are shown in Tables 7-12. It is

---

Insert Tables 7-12 about here

---

clear from the magnitude of the chi-square values in each table that the regression equations gave a generally better predictions in the lower grade levels than in the higher grade levels. Figures 4, 5, and 6 present graphs of the observed and predicted probability correct for each exercise in Set III and Set VII for Grades 4, 6, and 8 respectively. In this case, the best prediction, of the three problem sets graphed appears to be at the sixth-grade level. However, it is not yet a satisfactory fit for curriculum development purposes.

Insert Figures 4, 5, 6 about here

Table 7

Predicted, Observed, and Chi-Square Values

For Each Item - Problem Sets I-IV For Grade 4

Problem	Set I			Set II			Set III			Set IV		
	Pred.	Obs.	Chi Square	Pred.	Obs.	Chi Square	Pred.	Obs.	Chi Square	Pred.	Obs.	Chi Square
1	0.458	0.298	4.847	0.324	0.735	37.791	0.497	0.792	18.450	0.669	0.906	13.444
2	0.136	0.152	0.100	0.364	0.452	1.405	0.777	0.647	4.974	0.767	0.667	2.854
3	0.518	0.386	3.071	0.619	0.843	10.851	0.197	0.196	0.000	0.161	0.180	0.134
4	0.091	0.087	0.009	0.233	0.487	14.079	0.198	0.226	0.262	0.179	0.520	39.562
5	0.176	0.383	13.887	0.767	0.800	0.305	0.545	0.722	6.822	0.356	0.774	40.392
6	0.110	0.021	3.803	0.257	0.609	29.848	0.604	0.404	8.696	0.454	0.469	0.044
7	0.591	0.292	17.753	0.610	0.583	0.147	0.331	0.531	8.851	0.332	0.100	12.135
8	0.113	0.109	0.007	0.593	0.568	0.114	0.585	0.178	30.704	0.739	0.364	32.080
9	0.114	0.023	3.607	0.255	0.700	52.119	0.181	0.163	0.107	0.073	0.133	2.394
10	0.816	0.188	126.082	0.767	0.580	9.784	0.281	0.188	2.055	0.202	0.106	2.687
11	0.698	0.674	0.126	0.395	0.383	0.028	0.808	0.313	75.812	0.381	0.023	23.911
12	0.522	0.354	5.430	0.066	0.178	9.157	0.346	0.156	7.179	0.218	0.046	7.636
13	0.150	0.152	0.001	0.098	0.200	5.296	0.072	0.068	0.011	0.407	0.255	4.499
14	0.725	0.413	22.459	0.009	0.095	34.828	0.129	0.087	0.722	0.247	0.191	0.708
15	0.565	0.833	14.027	0.069	0.021	1.686	0.098	0.043	1.574	0.164	0.081	1.859
16	0.053	0.091	1.266	0.022	0.209	69.886	0.017	0.043	1.901	0.056	0.116	2.928
17	0.650	0.522	3.313	0.494	0.745	11.846	0.402	0.489	1.480	0.176	0.105	1.321
18	0.271	0.262	0.017	0.105	0.051	1.210	0.381	0.311	0.935	0.073	0.175	6.150
19	0.524	0.682	4.404	0.044	0.095	2.597	0.709	0.045	94.026	0.265	0.051	9.170
20	0.077	0.146	2.747	0.068	0.182	9.023	0.023	0.025	0.007	0.020	0.147	27.979

Table 8

Grade 5

Predicted, Observed, and Chi-Square Values
for Each Item - Problem Sets I-IV

Problem	Set I			Set II			Set III			Set IV		
	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square
1	0.659	0.822	5.320	0.671	0.816	4.667	0.714	0.761	0.498	0.801	0.771	0.271
2	0.302	0.467	5.812	0.563	0.729	5.376	0.789	0.795	0.010	0.783	0.875	2.391
3	0.657	0.545	2.449	0.750	0.918	7.376	0.462	0.535	0.922	0.356	0.488	3.268
4	0.227	0.267	0.410	0.460	0.660	7.568	0.530	0.561	0.158	0.443	0.898	41.111
5	0.442	0.727	14.491	0.783	0.959	8.933	0.640	0.762	2.713	0.515	0.878	25.850
6	0.226	0.222	0.004	0.591	0.378	8.446	0.736	0.667	1.029	0.580	0.745	5.253
7	0.666	0.622	0.392	0.683	0.809	3.446	0.607	0.933	13.365	0.597	0.227	25.037
8	0.298	0.465	5.733	0.687	0.761	1.171	0.736	0.444	15.797	0.730	0.692	0.286
9	0.289	0.122	5.565	0.444	0.804	24.149	0.466	0.600	2.886	0.253	0.707	44.715
10	0.811	0.238	89.965	0.830	0.783	0.720	0.451	0.405	0.316	0.402	0.442	0.286
11	0.736	0.860	3.403	0.542	0.596	0.552	0.805	0.600	10.709	0.655	0.048	66.850
12	0.598	0.636	0.264	0.235	0.488	14.598	0.628	0.410	7.934	0.801	0.546	13.462
13	0.429	0.386	0.332	0.278	0.133	4.714	0.359	0.314	0.308	0.431	0.614	6.008
14	0.784	0.659	4.060	0.046	0.463	162.461	0.304	0.235	0.765	0.352	0.558	8.000
15	0.721	0.902	6.677	0.274	0.386	2.775	0.484	0.412	0.706	0.304	0.513	8.051
16	0.216	0.421	9.430	0.122	0.538	63.008	0.181	0.242	0.828	0.244	0.386	4.810
17	0.682	0.784	1.775	0.506	0.881	23.628	0.467	0.886	24.686	0.240	0.237	0.002
18	0.521	0.778	9.528	0.275	0.143	3.059	0.562	0.806	8.707	0.225	0.538	21.911
19	0.699	0.946	10.729	0.176	0.282	3.022	0.812	0.235	74.151	0.409	0.425	0.042
20	0.278	0.206	0.878	0.196	0.550	31.809	0.143	0.032	3.117	0.120	0.412	27.452

Table 9

Predicted, Observed, and Chi-Square Values

For Each Item - Problem Sets I-IV For Grade 6

Problem	Set I			Set II			Set III			Set IV		
	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square
1	0.773	0.815	0.543	0.861	0.926	1.906	0.791	0.918	4.781	0.886	0.870	0.117
2	0.443	0.648	9.197	0.726	0.833	3.108	0.874	0.939	1.880	0.868	0.957	3.180
3	0.851	0.712	7.924	0.861	0.963	4.694	0.579	0.717	3.594	0.485	0.683	6.435
4	0.384	0.434	0.560	0.804	0.804	0.000	0.736	0.813	1.465	0.747	0.902	5.212
5	0.840	0.759	2.636	0.868	0.962	4.087	0.829	0.911	2.134	0.691	0.870	6.903
6	0.438	0.283	5.173	0.569	0.660	1.790	0.773	0.800	0.187	0.649	0.841	7.120
7	0.763	0.596	8.020	0.832	0.685	8.348	0.667	0.844	6.347	0.720	0.415	18.919
8	0.499	0.620	2.928	0.868	0.906	0.668	0.796	0.674	3.941	0.812	0.821	0.021
9	0.679	0.245	42.345	0.167	0.904	203.038	0.591	0.689	1.788	0.529	0.773	10.514
10	0.871	0.438	80.096	0.868	0.889	0.208	0.638	0.591	0.421	0.574	0.535	0.267
11	0.848	0.885	0.552	0.698	0.630	1.185	0.848	0.848	0.000	0.863	0.097	203.474
12	0.685	0.816	3.897	0.721	0.760	0.378	0.787	0.750	0.327	0.889	0.500	46.004
13	0.537	0.633	1.816	0.478	0.472	0.008	0.394	0.558	4.844	0.617	0.786	5.076
14	0.840	0.720	5.357	0.382	0.694	20.205	0.663	0.565	1.977	0.517	0.450	0.719
15	0.833	0.898	1.488	0.723	0.283	51.235	0.764	0.548	10.868	0.484	0.590	1.755
16	0.585	0.553	0.198	0.416	0.681	13.586	0.503	0.514	0.018	0.476	0.658	5.046
17	0.855	0.784	2.074	0.629	0.942	21.831	0.742	0.809	1.102	0.512	0.462	0.390
18	0.725	0.809	1.663	0.458	0.286	5.840	0.591	0.804	8.634	0.531	0.667	2.674
19	0.808	0.878	1.548	0.422	0.400	0.089	0.826	0.548	22.584	0.710	0.486	9.017
20	0.368	0.596	10.505	0.275	0.532	15.570	0.340	0.295	0.397	0.443	0.441	0.001

Table 10

Predicted, Observed, and Chi-Square Values

For Each Item - Problem Sets V-VIII For Grade 7

Problem	Set V			Set VI			Set VII			Set VIII		
	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square
1	0.262	0.656	51.382	0.920	0.516	141.927	0.864	0.879	0.126	0.872	0.913	1.039
2	0.908	0.831	4.613	0.779	0.857	2.226	0.990	0.924	29.040	0.845	0.638	22.574
3	0.560	0.309	14.063	0.518	0.781	17.730	0.291	0.831	91.867	0.657	0.719	1.092
4	0.546	0.778	13.679	0.872	0.774	5.335	0.411	0.719	25.080	0.458	0.770	23.921
5	0.655	0.656	0.000	0.658	0.905	17.080	0.033	0.754	1058.872	0.061	0.657	415.500
6	0.001	0.097	571.963	0.899	0.361	194.452	0.855	0.859	0.008	0.918	0.851	3.995
7	0.870	0.469	90.992	0.286	0.639	37.223	0.574	0.836	17.124	0.248	0.262	0.068
8	0.962	0.857	19.000	0.943	0.825	16.320	0.960	0.695	107.898	0.973	0.871	24.554
9	0.318	0.323	0.007	0.779	0.919	7.059	0.379	0.585	11.720	0.836	0.477	61.101
10	0.836	0.831	0.012	0.631	0.484	5.754	0.799	0.387	65.530	0.933	0.254	464.647
11	0.386	0.556	7.682	0.078	0.656	297.310	0.723	0.754	0.312	0.290	0.476	10.585
12	0.845	0.367	104.669	0.241	0.625	51.592	0.073	0.623	272.679	0.458	0.774	24.940
13	0.076	0.197	12.718	0.938	0.825	13.833	0.189	0.352	9.360	0.093	0.226	13.002
14	0.880	0.937	1.938	0.090	0.237	15.567	0.546	0.700	5.740	0.738	0.466	22.193
15	0.766	0.717	0.804	0.820	0.641	13.893	0.223	0.224	0.000	0.184	0.509	37.285
16	0.930	0.783	19.916	0.789	0.203	121.699	0.596	0.290	24.111	0.109	0.875	386.663
17	0.908	0.982	3.671	0.442	0.794	31.650	0.658	0.810	6.468	0.482	0.754	19.261
18	0.668	0.327	27.264	0.902	0.767	12.370	0.095	0.467	96.575	0.144	0.542	75.820
19	0.803	0.796	0.015	0.839	0.790	1.102	0.718	0.565	7.168	0.967	0.631	229.960
20	0.895	0.467	87.718	0.470	0.411	0.783	0.020	0.464	563.245	0.902	0.644	44.428

Table 11

Predicted, Observed, and Chi-Square Values

For Each Item - Problem Sets V-VIII For Grade 8

	Set V			Set VI			Set VII			Set VIII		
Problem	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square
1	0.563	0.736	6.447	0.988	0.508	1146.554	0.960	0.842	20.668	0.963	0.867	15.519
2	0.975	0.887	16.838	0.976	0.700	195.123	0.932	0.912	0.360	0.956	0.700	93.480
3	0.524	0.189	23.847	0.743	0.683	1.131	0.758	0.877	4.400	0.748	0.621	4.963
4	0.407	0.642	12.127	0.937	0.746	36.462	0.963	0.754	69.878	0.947	0.630	108.115
5	0.556	0.635	1.315	0.950	0.983	1.376	0.113	0.561	114.137	0.642	0.517	4.079
6	0.016	0.115	32.371	0.394	0.375	0.085	0.870	0.782	3.766	0.303	0.897	96.900
7	0.590	0.385	9.034	0.080	0.633	249.301	0.908	0.768	13.139	0.879	0.222	219.154
8	0.916	0.868	1.587	0.913	0.864	1.783	0.694	0.561	4.748	0.991	0.804	199.957
9	0.326	0.305	0.104	0.880	0.883	0.005	0.760	0.518	17.980	0.952	0.298	533.521
10	0.852	0.830	0.203	0.934	0.382	271.863	0.130	0.246	6.782	0.856	0.232	176.897
11	0.929	0.660	58.144	0.330	0.821	61.061	0.794	0.807	0.059	0.906	0.552	85.345
12	0.910	0.314	221.197	0.223	0.567	40.977	0.653	0.536	3.383	0.955	0.759	48.271
13	0.879	0.308	159.405	0.394	0.909	61.095	0.558	0.333	10.468	0.354	0.077	17.447
14	0.965	0.885	9.853	0.060	0.346	75.415	0.797	0.429	46.874	0.130	0.377	28.589
15	0.996	0.673	1361.719	0.825	0.475	50.060	0.854	0.094	245.523	0.642	0.429	11.054
16	0.824	0.863	0.535	0.597	0.184	34.739	0.924	0.200	410.539	0.285	0.772	66.341
17	0.974	0.923	5.341	0.723	0.649	1.559	0.814	0.754	1.355	0.898	0.610	53.427
18	0.996	0.429	3954.041	0.996	0.607	2126.993	0.422	0.400	0.109	0.681	0.386	22.834
19	0.938	0.885	2.512	0.908	0.627	55.769	0.977	0.393	849.944	0.465	0.586	3.413
20	0.742	0.500	15.296	0.366	0.640	16.177	0.070	0.167	7.805	0.974	0.561	383.921

Table 12

Predicted, Observed, and Chi-Square Values

For Each Item - Problem Sets V-VIII For Grade 9

Problem	Set V			Set VI			Set VII			Set VIII		
	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square	Pred.	Obs.	Chi-Square
1	0.197	0.744	81.332	0.869	0.341	100.406	0.798	0.900	3.227	0.808	0.939	5.420
2	0.853	0.907	1.000	0.702	0.818	2.830	0.999	0.920	312.366	0.608	0.796	7.266
3	0.999	0.088	8245.880	0.524	0.614	1.429	0.242	0.760	73.138	0.977	0.689	166.102
4	0.990	0.814	134.542	0.456	0.762	15.854	0.174	0.580	57.344	0.249	0.596	30.263
5	0.999	0.525	8996.137	0.261	0.907	93.035	0.006	0.646	3296.576	0.011	0.617	1586.549
6	0.004	0.049	20.840	0.995	0.385	2916.961	0.787	0.617	8.103	0.998	0.844	534.686
7	0.997	0.233	8391.406	0.968	0.651	139.496	0.169	0.659	75.224	0.340	0.091	12.157
8	0.999	0.800	1585.645	0.999	0.651	5212.746	0.998	0.524	4727.699	0.999	0.780	1968.394
9	0.921	0.262	250.687	0.702	0.814	2.578	0.042	0.271	62.560	0.975	0.467	476.426
10	0.765	0.854	1.806	0.350	0.405	0.558	0.997	0.239	8836.422	0.995	0.279	4430.988
11	0.364	0.512	3.879	0.008	0.610	1872.297	0.516	0.688	5.686	0.896	0.364	133.639
12	0.980	0.250	1087.552	0.979	0.591	322.190	0.763	0.478	20.662	0.180	0.571	43.503
13	0.170	0.167	0.003	0.996	0.907	85.492	0.046	0.237	31.590	0.096	0.143	1.069
14	0.818	0.929	3.476	0.677	0.158	46.809	0.909	0.479	107.293	0.984	0.343	913.417
15	0.688	0.610	1.162	0.986	0.548	583.702	0.078	0.222	12.975	0.052	0.488	165.817
16	0.999	0.738	2863.983	0.983	0.158	1547.703	0.321	0.311	0.021	0.607	0.689	1.268
17	0.853	0.907	1.000	0.295	0.829	56.215	0.545	0.714	5.644	0.302	0.587	17.725
18	0.521	0.139	21.050	0.521	0.735	6.239	0.489	0.225	11.157	0.110	0.300	14.750
19	0.728	0.789	0.714	0.743	0.595	4.818	0.573	0.404	5.486	0.999	0.500	10468.620
20	0.994	0.389	2209.409	0.971	0.563	189.170	0.006	0.167	182.542	0.999	0.541	7769.113

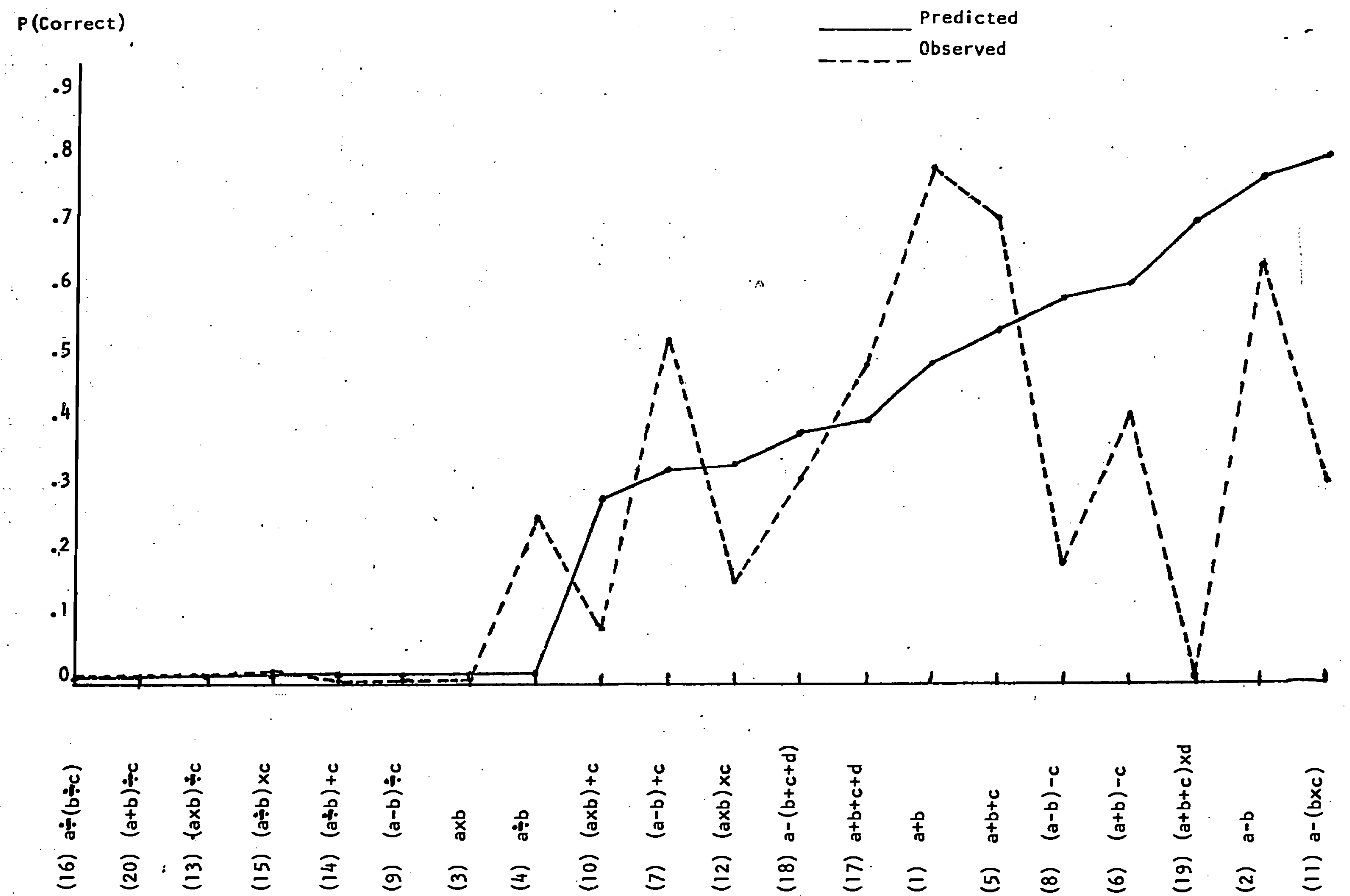

Figure 4. Observed and predicted probability correct for each exercise in Problem Set III for Grade 4, with required operations.

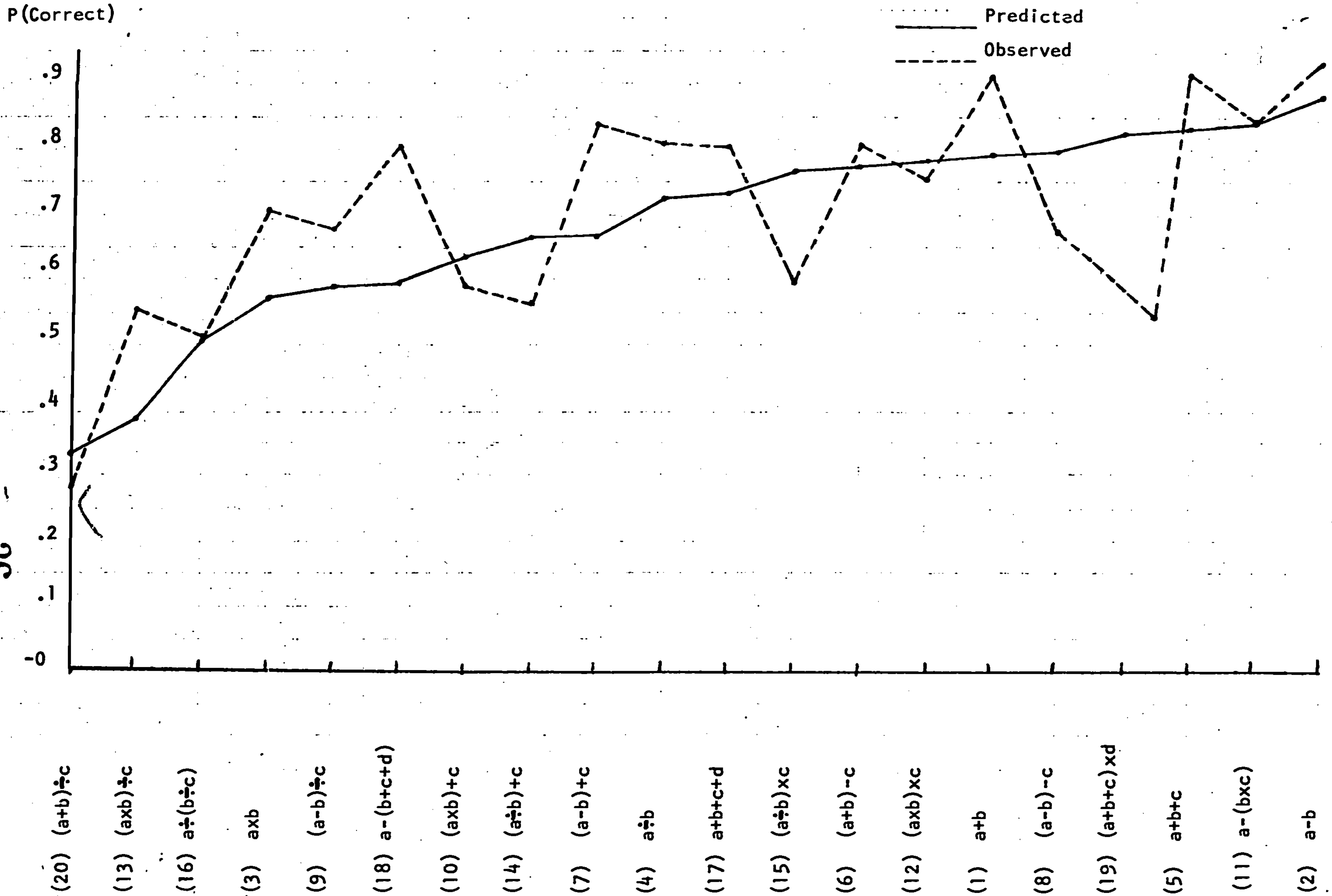

Figure 5. Observed and Predicted probability correct for each exercise in Problem Set III for Grade 6, with required operations.

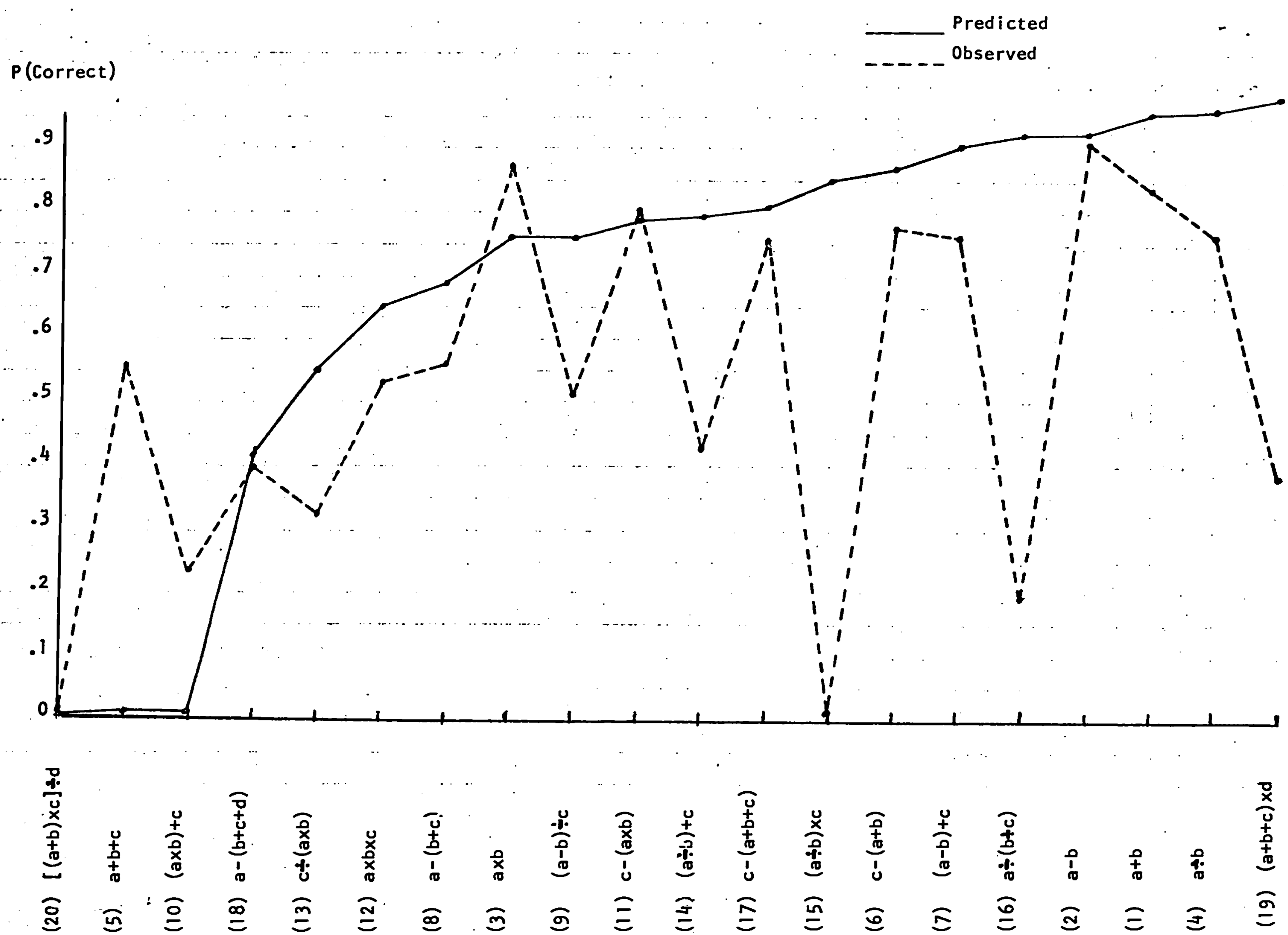

Figure 6. Observed and predicted probability correct for each exercise in Problem Set 7 for Grade 8, with required operations.

Two sets of exercises, Problem Sets I and IV, were administered to students in Grades 4, 5, and 6. The probability correct for each exercise on each problem set is shown in Figure 7. Perhaps the most important finding from this set of data is the degree of parallelness of the curves for each grade. What seems to be indicated is that the exercises have the same relative difficulty for students at each grade level. A similar pattern can

---

Insert Figures 7, 8 about here

---

be seen in the graphs presented in Figure 8. In Figure 8, the graphs are much closer, physically, indicating that the difficulty level for each exercise for each of the grades, 7-9, is much the same. Apparently the relative difficulty of the exercises was almost exactly the same for students at the several different grade levels.

The types of operations and the number of steps required to solve each exercise was presented in Table 2. As can be seen in Table 2, Problem Sets I and V had the same number of each type of exercise. In fact, the problem sets were identical in terms of the number of operations required for solution, the operation itself, and differed only in vocabulary level and the difficulty of the combination used. Problem Sets I and V were the simplest in terms of the variables just mentioned.

Problem Sets IV and VIII were the most complex in terms of the number of steps required for solution, as can be seen in Table 2. The graphs of the respective probabilities for each exercise are shown in Figure 9 and 10.

Insert Figures 9, 10 about here

Again, there is a strong indication that the relative difficulty of the

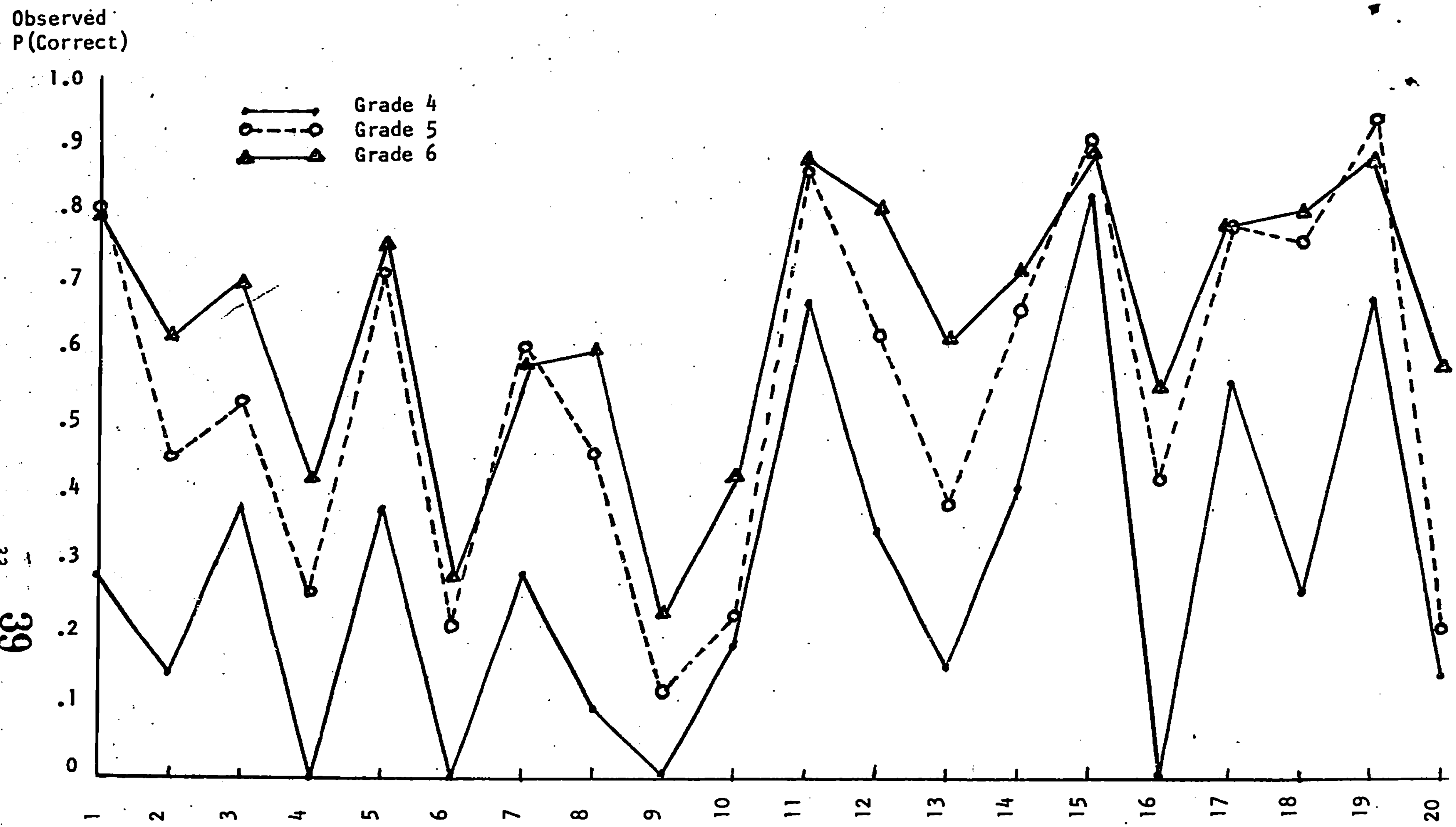

Figure 7. Probability correct for each exercise on Problem Set I by grade level.

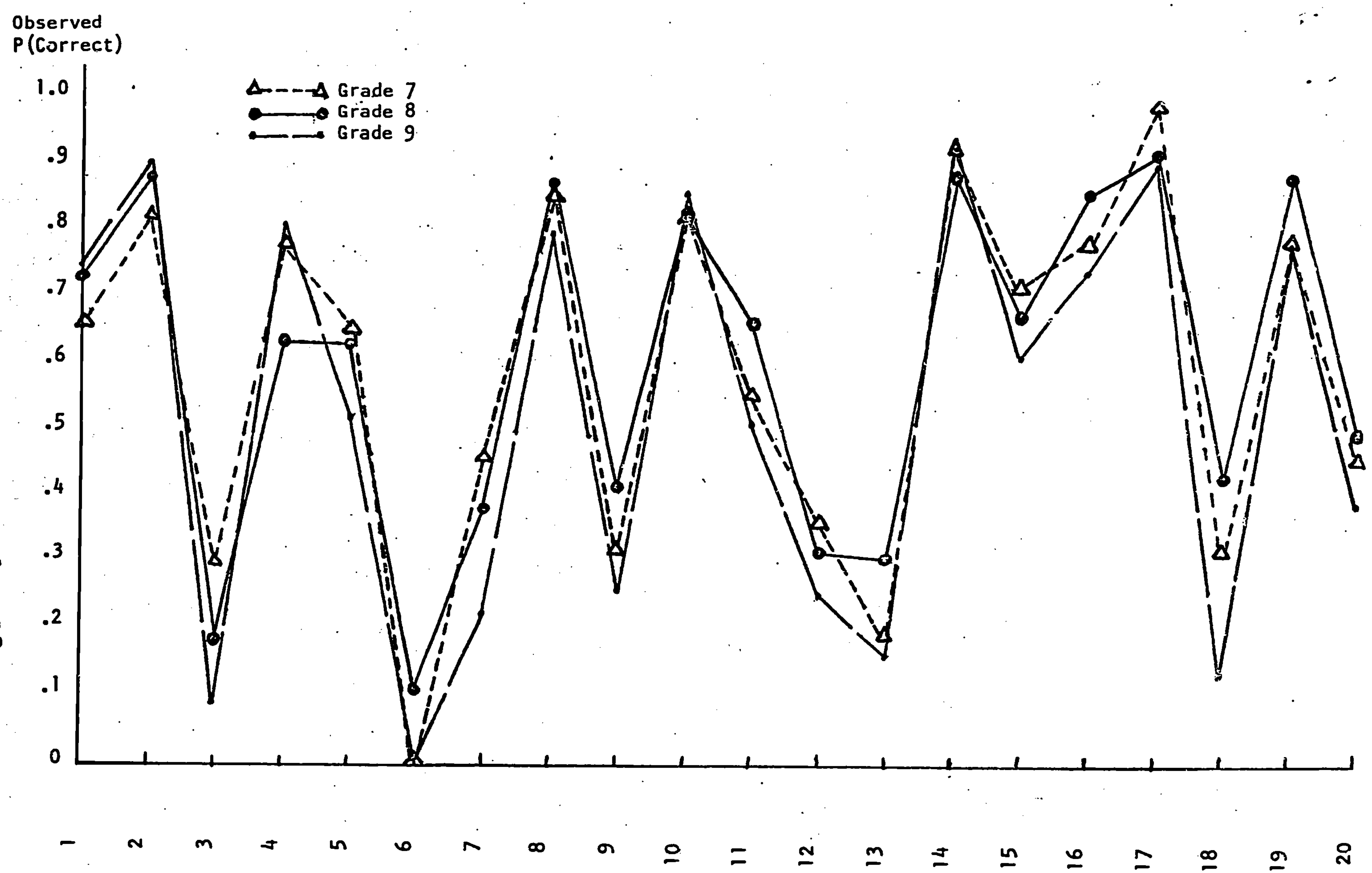

Figure 8. Probability correct for each exercise on Problem Set V by grade level.

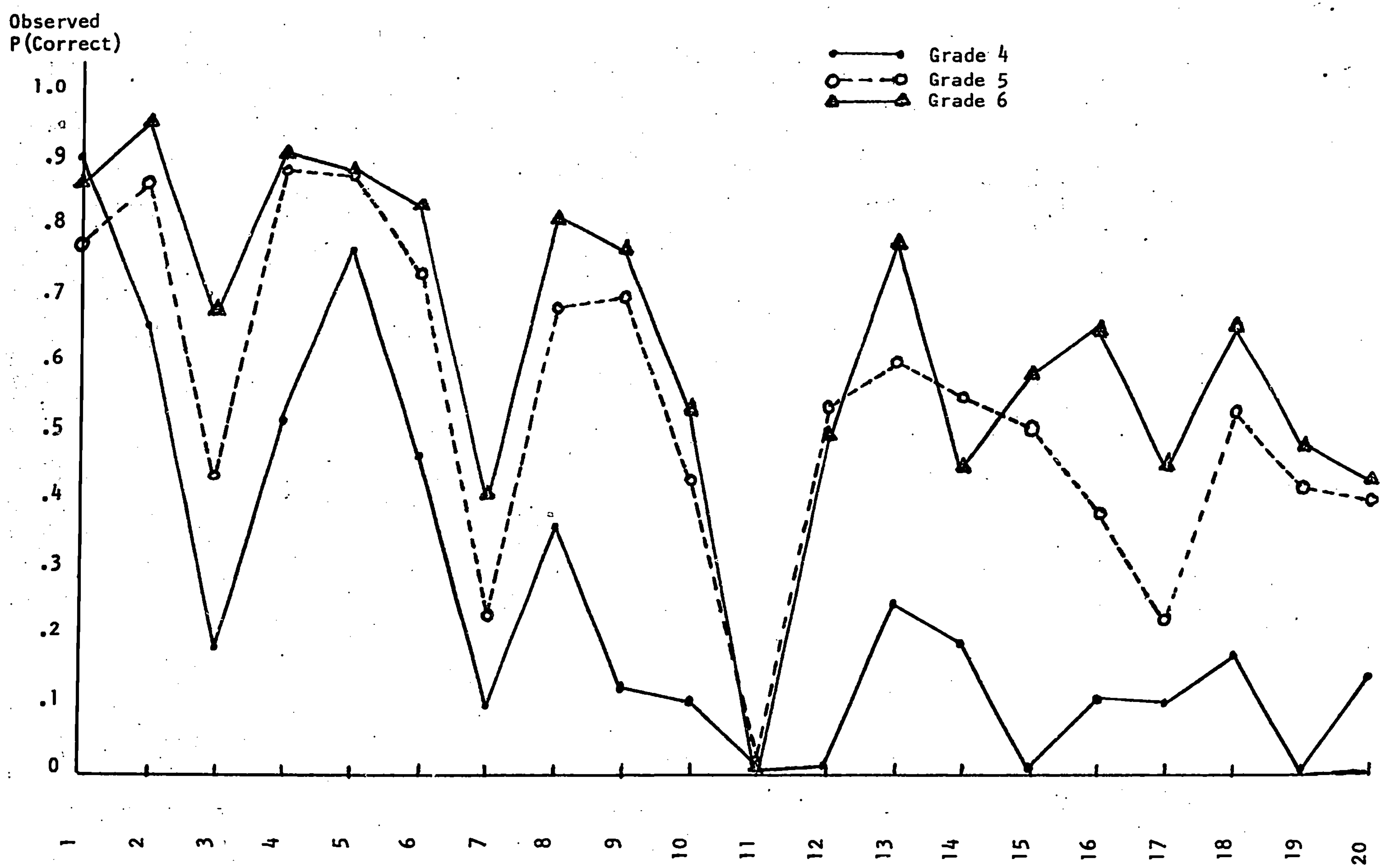

Figure 9. Probability correct for each exercise on Problem Set IV by grade level.

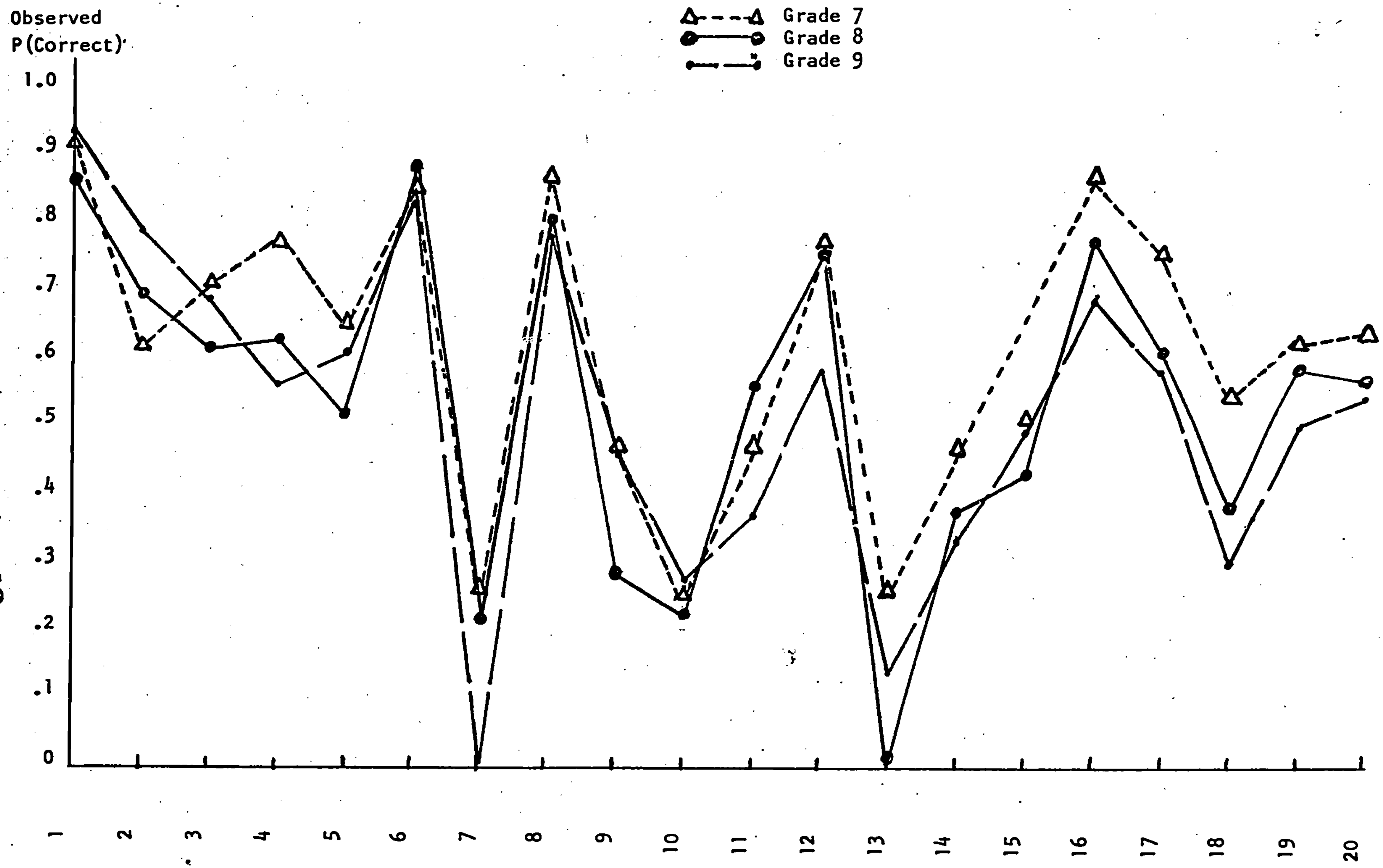

Figure 10. Probability correct for each exercise on Problem Set VIII by grade level.

exercise remained constant for students at different grade levels.

The residuals tend to give a somewhat clearer picture of the degree of accuracy of the prediction. The residuals are presented in Tables 13-18. The mean of the residuals is shown at the bottom of each column. As can be

---

Insert Tables 13-18 about here

---

seen, the mean residual per cent for each problem set is fairly small. The range is from 4 per cent, Table 15, to 15 per cent, Table 18. The finding that the mean residuals are as low, overall 11 per cent, is quite encouraging for a first effort. What one world like to be able to do is to reduce the residual to less than 5 per cent. Perhaps, in future efforts, this will be possible.

Table 13

Residual Per Cent Correct For Each Problem Set, Grade 4

Problem	Set I	Set II	Set III	Set IV
1	.002*	.270	.268	.285
2	.061	.100	.026	-.035
3	-.126	.137	-.034	.049
4	.032	.124	-.004	.034
5	.168	.053	.249	.343
6	.042	.063	-.058	.042
7	-.190	.112	.101	-.164
8	.018	.054	-.422	-.195
9	-.094	.001	-.010	-.190
10	-.000	.167	.032	.009
11	.104	.261	.148	-.045
12	.082	.034	-.085	-.000
13	-.000	.042	-.027	-.074
14	-.184	.002	-.014	-.119
15	.283	.065	-.015	-.002
16	.028	.116	.007	.056
17	-.094	.021	.103	-.021
18	.023	.040	-.050	.131
19	.162	.124	-.053	-.019
20	.052	.078	.008	-.050
$\overline{X}$	.09**	.09	.08	.09

* .2 per cent
** 9 per cent

Table 14

Residual Per Cent Correct For Each
Problem Set, Grade 5

Problem	Set I	Set II	Set III	Set IV
1	.100*	.110	-.030	.080
2	.124	.081	-.019	.091
3	-.239	.054	.103	.011
4	-.023	.064	-.011	.119
5	.072	.049	-.004	.141
6	-.122	-.296	-.099	.002
7	-.129	-.007	.167	-.294
8	.012	-.010	-.370	-.130
9	-.280	.141	.064	-.083
10	-.000	-.126	.043	.121
11	.030	-.116	.160	-.085
12	-.029	.101	-.131	.000
13	.027	-.186	-.149	-.046
14	-.184	-.068	-.117	-.136
15	.091	-.007	.227	.163
16	.100	.254	.061	-.027
17	.075	.074	.146	-.156
18	.226	-.132	.067	.271
19	.166	-.239	-.135	.173
20	.003	.286	-.043	-.222
$\overline{X}$	.10	.12	.10	.12

* 10 per cent

Table 15

Residual Per Cent Correct For Each
Problem Set, Grade 6

Problem	Set I	Set II	Set III	Set IV
1	.109*	.104	.049	.059
2	.099	.076	.033	.083
3	-.081	.048	.028	.034
4	-.020	.023	.058	.064
5	.098	.023	.051	.054
6	-.184	-.163	-.051	.021
7	-.191	-.206	.018	.219
8	.063	.056	-.220	-.091
9	-.332	.083	-.005	-.054
10	-.000	-.050	-.037	.061
11	.062	-.208	.084	-.186
12	.002	.171	.014	-.009
13	.032	-.008	-.003	.036
14	-.124	.010	-.084	-.320
15	.078	-.256	-.013	.120
16	.062	.207	.040	.114
17	-.064	.052	-.010	-.046
18	.138	-.121	.010	.299
19	.065	-.299	-.129	.100
20	.070	.129	.002	-.236
$\overline{X}$	.09	.11	.04	.11

* 10.9 per cent

Table 16

Residual Per Cent Correct For Each Problem Set, Grade 7

Problem	Set V	Set VI	Set VII	Set VIII
1	-.006*	-.272	.059	.149
2	-.058	.058	.105	-.110
3	-.306	.022	.151	.130
4	.114	-.041	.182	.040
5	.141	.276	.116	.050
6	.063	-.328	.030	.235
7	-.240	.037	.048	-.136
8	.136	.000	-.111	.015
9	-.095	.120	.069	-.182
10	.000	-.062	-.335	-.444
11	.019	.296	-.043	-.300
12	-.057	.039	.085	.036
13	-.143	.031	-.141	-.170
14	.071	-.133	.136	-.115
15	-.058	.084	-.053	.050
16	.074	-.318	-.321	.200
17	.091	.091	-.020	.204
18	-.343	-.026	.013	.073
19	.004	.065	-.204	-.037
20	.092	-.218	.027	.039
$\overline{X}$	.10	.12	.11	.14

* -.6 per cent

Table 17

Residual Per Cent Correct For Each Problem Set, Grade 8

Problem	Set V	Set VI	Set VII	Set VIII
1	.099*	-.164	.057	.119
2	.015	-.081	.117	-.022
3	-.374	-.058	.197	.104
4	.013	.014	.324	-.046
5	.108	.337	.060	.047
6	.067	-.297	-.012	.405
7	-.281	.063	.088	-.110
8	.133	.000	-.186	-.036
9	-.036	.102	.153	-.245
10	.020	-.171	-.445	-.402
11	.046	.324	.021	-.195
12	-.149	.060	.163	.082
13	-.198	.069	-.005	-.109
14	.038	-.230	.064	-.080
15	-.089	-.064	-.116	.049
16	.173	-.449	-.317	.247
17	.053	-.036	-.030	.134
18	-.249	-.090	.118	.050
19	.094	.056	-.228	-.044
20	.011	-.018	-.085	.046
$\bar{X}$	.11	.13	.14	.13

* 9.9 per cent

Table 18

Residual Per Cent Correct For Each Problem Set, Grade 9

Problem	Set V	Set VI	Set VII	Set VIII
1	.100*	-.346	.157	.148
2	.044	.037	.151	.030
3	-.268	-.132	.130	.238
4	.134	-.006	.157	-.116
5.	.167	.371	.200	.135
6	.036	-.278	-.134	.276
7	-.419	.058	.112	-.138
8	.194	.000	-.170	-.030
9	-.127	.032	-.045	-.113
10	-.096	-.006	-.371	-.422
11	.049	.286	-.066	-.221
12	-.098	.064	.141	-.126
13	-.185	.087	-.036	-.110
14	.107	-.244	.071	-.156
15	-.058	.091	-.024	.174
16	.158	-.283	-.201	.261
17	.046	.140	-.010	.134
18	-.413	.022	-.039	-.010
19	.071	.002	-.217	-.084
20	-.037	-.073	.004	.084
$\overline{X}$	.14	.13	.12	.15

* 10.0 per cent

Discussion - Conclusions

Although several studies have attempted to define structural variables that are capable of accounting for a significant amount of the observed variance in the probability correct of word problem exercises in arithmetic, to the writer's knowledge this study has been the first to attempt to predict difficulty probability correct, in advance.

Eight sets of word problem exercises were prepared in which such things as the number of words used, the number of steps required for solution, the operations required, and the difficulty of the computations involved were carefully controlled. The problem exercises were coded in terms of the variables found to account for the most variance in observed probability correct in previous a study. The regression analysis did verify that the six variables which accounted for the most variance in the pooled data for Grades 4-9 in the previous study also accounted for a significant amount of the variance in the present study. The predictions, however, were not as accurate as it was anticipated they would be in terms of a chi-square test. Even when separate regression equations were used to predict the probability correct for each individual grade level, the predictions were not as accurate as anticipated although they were somewhat better than for the general case. The actual magnitude of the residuals does not present such a dark picture. In fact, it was encouraging to observe that the over all mean residual was 11 per cent, range 4-15 per cent.

It was also encouraging to find that the relative difficulty of the word problem exercises was constant over grade levels. This lends support to the assumption that there must be a set of variables which can be found to account for the observed consistent levels of difficulty over grades. That is, because those graphs are as they are, lending support to earlier findings, we believe we are correct in following the approach in this study.

What remains to be done is to improve the set of variables and thereby improve the predictive power of the regression equations. A small set of independent structural variables will be defined. The scope will be broadened to include syntactic variables which should account for some of the complexity in the language of the word problem itself. Together, the computational and syntactic structural variables may provide the key to unlocking a complex situation.

It is recognized that the set of variable used in this study is in need of further refinement. However, for the purposes of this project, the objectives have been met. That is, it does seem possible to use structural variables to predict the relative difficulty of word problems in arithmetic with a reasonable degree of accuracy. We are encouraged, but much work remains to be done.

References

Jerman, M. Individualized Instruction in Problem Solving in Elementary School Mathematics. Journal For Research in Mathematics Education (In Press).

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. Minneapolis: University of Minnesota Press, 1967, 1, 160-200.

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

Suppes, P.; Loftus, E.; and Jerman, M. Problem solving on a computer-based teletype. Educational Studies in Mathematics, 1969, 2, 1-15.

Suppes, P.; Morningstar, M.; Jerman, M.; Oksenberg, L.; and Searle, B.; and Morningstar, M.; Regression Models and Response Data for 1966-67 in Suppes, P., Computer-assisted Instruction at Stanford, 1966-68: Data, Models, and Evaluation of the Arithmetic Programs. New York: Academic Press, 1972, 53-132.

Appendix A

Administering The Problem Sets

1. Be pleasant - smile
2. Introduce yourself in something like the following manner:

Good morning, I am ________ from Penn State University at Penn State. One of the things we are trying to do is to find out what makes word problems so difficult for students to learn to solve. We have already found several clues and hope to find more, with your help.

I am going to give each of you a set of word problems to solve. We think we know which ones are easy and which are difficult because of our research. Now I want you to work these for us so that we will know whether or not we have found the right clues.

All you need is your pencil. It is important that you try every problem. Do your best on each problem. If a problem is too difficult, go on to another and come back to it later if there is enough time.

We want you to put your name on the first page so that if you are able to solve a problem no one else can solve, we would like to know how you did it. Also, if you are not able to do a problem everyone else can do, we may want to get you to tell us what you thought was difficult about the problem. Do all your work on the paper I give you. If you finish ahead of time, turn your paper over. I will collect them when everyone is finished.

Any questions?

Alright, as soon as I hand out the papers you may begin.

APPENDIX B

[illegible]

[illegible]

[illegible]

[illegible] did the [illegible]

3. [illegible] packs of gum [illegible] John decided to sell the remainder of [illegible] each pack for five cents, how much [illegible] will he [illegible]

[illegible]

4. In an old store, an antique dealer [illegible] found [illegible]

5. Jim has to mail some letters. He has eighty-four cents. [illegible] [illegible] stamps can he buy?

6. A rabbit keeper fed each of his 467 rabbits [illegible] carrots each morning and [illegible] carrots each night in addition to their other [illegible]. [illegible] carrots did all the rabbits eat each day?

7. If her brother took away 152 postcards, how many postcards [illegible] have left? Mary had 686 postcards. Her mother gave her [illegible] more.

8. A [illegible] drove his car [illegible] miles one day. On the trip he used [illegible] of gasoline. If gasoline costs [illegible] per gallon, how much [illegible] gas [illegible]?

[illegible] drove three hundred sixty miles in [illegible] hours [illegible] average forty-four miles an hour [illegible]
[illegible]

[illegible] bought 2[illegible] feet of rope for a project. The project used only [illegible] of rope. How many feet of rope were left?

11. How many gallons does the tank have now? The tank had [illegible] gallons of water in it then, 375 gallons were added.

12. How many more bicycles were produced the first week [illegible] of all the wheels [illegible] the number of [illegible] in the United States was [illegible] the first week [illegible] second week.

13. How many 7 foot lengths could they get from the remainder of the roll? The sixth graders were decorating a float for the parade [illegible] they bought a 48 foot roll of crepe paper ribbon, they lost 11 feet of it.

14. John had [illegible] marbles. He decided to give [illegible] of his marbles to his friend Bill. How many marbles did John have left?

15. There are 475 cars parked in lot A and 326 cars parked in lot B. How many cars are parked in the two lots?

16. Eight neighbors went together to buy a garden tractor for [illegible] and a snow blower attachment for $405. If they share the cost equally, what is each person's share?

17. [illegible] many shells were not placed in the showcase? Tom collected 79[illegible] seashells. He placed 6[illegible] of them in a showcase.

18. [illegible] made 56 candles for a customer of [illegible]. If the candle [illegible] he had each held 8 candles, how many boxes were needed in order to package the candles?

19. There are [illegible] people seated in section [illegible] of the stands [illegible] seated in section C. How many people are seated [illegible]

20. [illegible] gave 8 boys a total of $[illegible] to [illegible] boys shared the money equally. The man was [illegible] that he gave an additional $[illegible].50 to each boy. How [illegible] boy receive for his work?

Problem Solving

Name ______________________

Grade __________ Date __________

This set of problem solving exercises is intended to identify the problem solving skills you now have. It is important that you do your best work for the few minutes you spend working these exercises.

Do not use scratch paper. Do all your work on these pages in the space provided.

If you do not know how to do a problem go on to the next problem and come back later and try it again. <u>It is important that you try every problem.</u>

1. Mr. Jones, a salesman, [illegible] miles last week. His car used 36 gallons of gasoline [illegible] \$.33 per gallon. [illegible] did he pay for gasoline [illegible]

2. Fred bought a [illegible] \$[illegible].85 and gave the salesman a [illegible] bill. How much [illegible] receive?

3. Connie rode her bicycle [illegible] to town B in 8 hours. On her return trip she [illegible] average 2.5 miles per hour faster. At [illegible] speed, how [illegible] make the return trip? The distance between the two [illegible] miles.

4. [illegible] machines which made candy cones. If [illegible] cones per hour, how many were made [illegible]

5. One book had [illegible] pages, [illegible] other had 212 pages. [illegible] an average of [illegible] day? David [illegible] days.

6. [illegible] and [illegible] the market. Their [illegible] steak, [illegible] medium rare. They [illegible] of steak which [illegible] of 7 3/4 pounds. [illegible] didn't have [illegible] for all of the meat, [illegible] the smaller [illegible] 3-3/16 pounds. If [illegible] \$1.50 a pound, [illegible]?

7. A ship sailed [illegible]. During the next [illegible] days it sailed [illegible] miles each day. [illegible] the ship sailed in the first [illegible] days of its [illegible].

[illegible] thousand nine hundred twenty seven [illegible] greater is the diameter of the earth [illegible]
At [illegible] of mercury is three thousand one hundred [illegible].

9. A farmer weighed [illegible] watermelons while they were still on the vine. They weighed [illegible] lbs. each. Later when they each had gained [illegible] lbs. more weight, they were sold. What was the weight of the watermelons when the farmer sold them?

9. ________

10. Although John had $[illegible] yesterday, he doesn't today because he spent $8.12. If John's father gives him $4.30 tomorrow how much money will John have then?

10. ________

11. Mrs. Smith needed 68 pieces of chalk for her class. She found that she must get chalk by the box. If there are 9 pieces of chalk per box, how many boxes must she get?

11. ________

12. A 36 foot long piece of ribbon was cut up for decorations. If each decoration took 2 1/3 feet of ribbon, how much was left after 8 decorations had been made?

13. The salesman sold 4 suits for a total of [illegible]. How much did he make on each suit? Each suit cost the salesman [illegible].

13. ______

14. A company manufactured 1,500 cases of chewing gum one day and 1,750 cases the following day. How many cases of gum were manufactured during the two days?

14. ______

15. How many did he deliver on York Road? Paul delivered nine hundred thirty six papers. Of these he delivered two hundred eighty eight on Poplar Street, three hundred sixty on Garfield Avenue, and the rest on York Road.

15. ______

16. The new car was driven 500 miles in 8 hours. What was its average speed per hour? The car was a new red Cadillac with a vinyl top.

17. How many fish did he catch that day? ... 465 fish in the morning and 224 in the afternoon.

17. ______

18. Together Jeff and Betty sold ninety ... If Betty sold twice as many as George, how many boxes did George sell? Jeff sold ... seven boxes of candy in the yearly candy sale his school held.

18. ______

19. A rocket traveled 67.132 miles during the first minute after lift off and 121.352 miles during the second minute after lift off. How many miles did it travel in the first two minutes?

19. ______

20. Jane bought 1 1/8 yards of curtain material. She used 6/10 of a yard for a kitchen window. What part of a yard is left?

20. ______