PATRICK SUPPES, ELIZABETH F. LOFTUS, AND MAX JERMAN

PROBLEM-SOLVING ON A COMPUTER-BASED TELETYPE*

I. INTRODUCTION

At Stanford University over the last five years, the Institute has been developing a working computer-assisted instruction (CAI) system for classroom use that follows two distinct approaches: tutorial and supplementary drill and practice. The tutorial approach to CAI uses the computer as a "teacher" to present new concepts as well as to determine subsequent student work with the concepts. In contrast, drill-and-practice systems supplement classroom instruction by improving the skills and concepts introduced by the classroom teacher.

In the spring of 1965, a CAI drill-and-practice program was initiated in an elementary school. To implement this program a computer at Stanford was used; telephone lines connected the computer to the teletypes located at the school. Fourth-, fifth-, and sixth-grade students received daily drills in arithmetic (Suppes *et al.*, 1966). Beginning in the fall of 1965, this operation was expanded (Suppes *et al.*, 1968); by the fall of 1966, computer-controlled drills were given to approximately 900 students in six different schools. During the past academic year, 1967–68, these drills reached over 2000 students per day. Elementary schools in Kentucky and Mississippi, where children received daily drills in arithmetic, were also linked to the central computer at Stanford.

The research reported here is a small part of an investigation of the potential use and value of CAI drill-and-practice systems. This study, in particular, reports a new use for such systems. The students who participated were first taught the mechanics of how to use a computer-based teletype to solve arithmetic word problems. Following this, a series of word problems was presented to them. A central characteristic of these problems was the requirement of a quantitative answer, but the arithmetical operations were not explicitly indicated. An example of a problem in arithmetic providing the pupil with an opportunity to use his knowledge of subtraction is the following:

> Tom collected 500 seashells and placed 43 of them in a showcase. How many shells were not placed in the showcase?

We attempted to determine the factors related to problem difficulty by

Educational Studies in Mathematics 2 (1969) 1–15; © *D. Reidel, Dordrecht-Holland*

analyzing the solutions of the problem series. An example of what we mean by a factor related to problem difficulty is the length of the problem. A natural assumption is that the larger the number of words in a problem, the harder the problem is to solve.

II. THE THEORY

The discussion of the regression model follows Suppes *et al.* (1968). What is desired is an analysis of factors that lead to varying difficulty. We would like to attach weights to the various factors that may be objectively identified in each item, and then to use estimates of a few such weights to predict the relative difficulty of each of a large number of items. To this end, the aim of the present paper is to formulate and test some linear structural models that lead to parametric predictions of relative difficulty.

For the word problems analyzed in this paper, the central difficulty was to identify the factors that contributed to the complexity of the problem. As a matter of notation, the jth factor of problem i in the set of problems is denoted by X_{ij}. The statistical parameters estimated from the data are the weights attached to the factors. The weight assigned to the jth factor is denoted by α_j. It should be emphasized that the factors identified and used in the model presented in this paper are not abstract constructions from the data. Rather, they are always objective factors identifiable by the experimenter in the problems themselves, independent of any data analysis. Which factors turn out to be important is a matter of the estimated weights α_j. All the factors used in the analyses presented here have an intuitive and direct relevance to commonsense ideas of difficulty, and their definitions are straightforward.

Consider the analysis of the response data. Let p_i be the observed proportion of correct responses for a group of students on problem i. The central task of a model is to predict the observed proportion p_i. The natural linear regression model in terms of the factors X_{ij} and the weights α_j is simply

$$p_i = \sum_j \alpha_j X_{ij} + \alpha_0 .$$

A difficulty with this model, however, is that probability will not necessarily be preserved as the estimated weightings and the identifiable factors are combined to predict new observed proportions of correct responses.

In order to guarantee preservation of probability, that is, to insure that predicted p_i's will always lie between 0 and 1, it is natural to make the following transformation and to define a new variable z_i[1]:

$$(1) \qquad z_i = \log [(1 - p_i)/p_i] .$$

We then use as the regression model

$$(2) \qquad z_i = \sum_j a_j X_{ij} + a_0 .$$

It should be noted that the reason for putting $1-p_i$ rather than p_i in the numerator of Equation (1) is that it is desirable to make the variables z_i increase monotonically in difficulty. For example, if the length of a problem increases with the difficulty of the problem, it is desirable that the model reflect this increase directly rather than inversely.

The variables we consider are of two types. The first type are 0, 1-variables. Such variables would be appropriate, for example, in dealing with a problem that requires a conversion of units. If a problem requires a conversion of units, such as from months to weeks, the *conversion* variable for that problem receives a value of 1, and 0 otherwise. The second kind of variable is one that assumes a finite set of values, but the set is greater than 2. Such a variable would be appropriate, for example, in dealing with the length of the problem; the *length* variable receives a value which is equal to the number of words in the problem.

Two other variables of the second type are the *operations* variable and the *steps* variable. The *operations* variable refers to the minimum number of different operations required to reach the correct solution. For a given problem, this variable could take on a value of 1, 2, 3, or 4. The *steps* variable refers to the minimum number of steps required to reach the correct solution.[3] These two variables may be distinguished more clearly if we consider a problem which requires two or more computational processes before the answer can be found. This type of problem is called a "multiple-step" problem, and "multiple-step", as used here, refers not to the details of the processes, but to the number of binary operations – addition, subtraction, multiplication, or division – required to obtain the answer. A problem that asks the student to find the average of 11 numbers would give a value of 11 to the *steps* variable and a value of 2 to the *operations* variable.

A few words must be said about the *length* variable. Sentence length is frequently proposed as the most obvious and plausible variable in determining sentence difficulty. This factor is generally determined by total count of the number of words in the sentence. Studies in language acquisition (Ervin, 1964; Miller and Ervin, 1963) give evidence of the gradual progression of children's language development from one-word sentences, holophrases, to two-word pivot sentences, to sentences consisting of greater numbers of words. In imitation of adult sentences, children tend to use a "telegraphic code", a sentence form which is a shortening of adult sentences that retains only content words. Braun-Lamesch (1962) found that younger children cannot recall whole sentences easily. Because this evidence indicates that young children in early language development lack the ability to process long sentences, it seems safe to say that long sentences are more difficult for children to comprehend than shorter sentences. For the present, we shall generalize these results and assume they imply that longer word problems will be more difficult than shorter ones. In subsequent studies, however, we hope to look at the actual syntactic structure of the sentences, which should be a more meaningful index of difficulty than mere word count alone.

The *sequential* variable is the first 0, 1-variable. Post (1958) completed a carefully designed study which investigated the effects of several factors on problem-solving in arithmetic. The factors studied were: (a) size of numbers; (b) superfluous numerical data; (c) number of steps; (d) familiarity with setting; (e) type of operation; and (f) symbolic terms. Each factor was investigated on two levels that were studied in conjunction with each level of the remaining factors, giving sixty-four (2^6) treatment combinations in all. The findings indicated that the type of operation was the most important factor, although familiarity of setting and superfluous numerical data were significant also. These results suggest a new factor which we chose to call the *sequential* variable. If a problem may be solved by the same operation(s), in the same order, as the problem that preceded it, the *sequential* variable for that problem is assigned the value of 1, and 0 otherwise. The possible importance of this factor in the present context was suggested to us by Professor Leon Henkin. Successful use of it in the analysis of fractions is found in Suppes *et al.* (1968, Chap. 7).

The *verbal-clue* variable is the second 0, 1-variable. Brownell and Stretch (1931) felt that a problem could be analyzed into several elements or factors, one of which was a verbal clue to the operations. This factor was not varied systematically, and so no conclusions could be drawn about it. Brownell and Stretch did suggest that there were other factors, yet unknown, which influence problem-solving. We are again indebted to Leon Henkin for suggesting this variable in the present context. He chose to define it as follows:

(1) The verbal clue for problems requiring a single addition is the word "and"; if the problem contains this word, the *verbal-clue* variable for that problem is to be assigned a value of 1, and 0 otherwise.

(2) The corresponding verbal clues for the other operations are:

(a) "left" or a comparative for subtraction;

(b) "each" for multiplication;

(c) "average" or "each" appearing in the question sentence of the problem for division.

(3) Problems requiring multiple operations must contain all of the verbal clues pertaining to the required operations in order that the *verbal-clue* variable be assigned a value of 1.

The *conversion* variable is the last 0, 1-variable. If a problem requires a conversion of units, such as from months to weeks, the conversion variable for that problem is assigned a value of 1, and 0 otherwise. The importance of this variable was suggested by the results of an informal pilot study described below.

In summary, the variables we investigated are:

X_1 = the *operations* variable, that is, the minimum number of different operations required to reach the correct solution;

X_2 = the *steps* variable, that is, the minimum number of steps required to reach the correct solution;

X_3 = the *length* variable, that is, the number of words in the problem;

X_4 = the *sequential* variable, 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 = the *verbal-clue* variable, assigned a value of 1 if the problem contains a verbal clue to the operation(s) required to solve the problem, and 0 otherwise;

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

III. DESIGN AND EXPERIMENTAL PROCEDURE

A. *Subjects*

The 27 subjects used in this study were taken from an accelerated mathematics group composed of bright fifth-grade students from four different elementary schools near Stanford University. The children all came from middle-class, suburban communities. The students had received teletype instruction in logic and mathematics drill and practice, so familiarizing them with the machine was not a problem.

B. *Equipment*

The student terminals used in this project were commercially available teletype machines, connected by private telephone lines to a computer at the Institute for Mathematical Studies in the Social Sciences at Stanford. There were 10 teletypes, all operating in a single classroom at one of the elementary schools. The children from the other three schools were bussed to that school for one hour every day. When not operating the teletypes, the children in the special mathematics group received classroom instruction in elementary mathematics from Mr. James Newland, a teacher associated with the project.

The control functions for the entire system were handled by the PDP-1, a medium-sized computer with a 32000-word core and a 4000-word core interchangeable with any of 32 bands of a magnetic drum, together with two large IBM-1301 disk files. All input-output devices were processed through a time-sharing system. Two high-speed data channels permitted simultaneous computation and servicing of peripheral devices.

C. *Instructional Program*

To initiate a lesson, a student typed "*P*" (for problem-solving) followed by his assigned number and his name. When this was correctly done, the program began. If the student made an error or gave a fictitious name, such as *Superman*, he was asked to try again.

The computer consulted the student's file and began with the item following the last one completed. The items were divided into two parts, with the set of instructions presented before the set of problems.

TABLE I

Operation abbreviations taught in the instruction set

Code	Comments
X THE ANSWER KEY G (1) 21 *1X*	The line number followed by X indicates what line the answer is on.
A ADD G (1) 36 G (2) 41 *1.2A* (3) 77	
S SUBTRACT G (1) 500 G (2) 48 *1.2S* (3) 452	
M MULTIPLY G (1) 59 G (2) 4 *1.2M* (3) 236	
Q DIVIDE G (1) 77 G (2) 7 *1.2Q* (3) 11	Q rather than D was used for divide because D was used for something else in the system.
E ENTER G (1) 41 *E* (2) *7*	E is used to enter a number that is not entered by the computer program. For example, in a problem that asks the student to find the number of days in 12 weeks, the student would be required to enter the number 7, the number of days in one week. The number 12 would be entered by the computer as a "given number".

Note: Student entries are italicized.

The set of instructions. Instructions were presented, via computer, to teach the students how to command the computer to perform operations on given numbers. Table I lists and gives an example of each of the abbreviated operation names that the student learned in the instruction set.

The following sequence of interactions between the student and the computer illustrates how a problem is solved in this context. Student entries are underlined. The computer first types out the problem, and then types out the numbers in that problem. The student sees on the printout sheet before him:

> Tom collected 500 seashells and placed 43 of them in a showcase. How many shells were not placed in the showcase...
>
> G (1) 500
> G (2) 43

"G" stands for "given number".[3]

The student then responds by telling the computer the operation he wants the computer to perform, and the line numbers to which the operation should apply. In the present case, the student ordinarily types out "1.2S" meaning "from the number shown on line 1 subtract the number shown on line 2". The computer responds by typing the result of applying the operation, or by typing an error message if the operation could not be applied validly.

The student also learned to indicate the answer by typing the line number followed by an X. The complete protocol for a correct response in the above example, then, might be:

> Tom collected 500 seashells and placed 43 of them in a showcase. How many shells were not placed in the showcase...
>
> G (1) 500
> G (2) 43
> *1.2S* (3) 457
> *3X*
> Correct

(Again, student entries are italicized.) If the answer is incorrect, "answer is wrong" appears in place of "correct". The protocol for a response which elicits an error message might be:

> Tom collected 500 seashells and placed 43 of them in a showcase. How many shells were not placed in the showcase...
>
> G (1) 500
> G (2) 43
> *1.2AD* There is no rule name "AD".

There are often many ways to solve a given problem. Which rule to use and the details of use are matters of strategy determined by the experience and ingenuity of the student. The computer allows any valid step, regardless of whether it helps reach the solution. Any combination of steps reaching a solution, valid within the rules, is entirely acceptable, however idiosyncratic.

In the instruction set, easier examples preceded more difficult ones. On several of the problems, the student was invited to ask for help after a certain time lapse by the message, "Type H and a space if you want a hint". No hints were available on multiple-choice problems; the student had to guess until he got the problem correct.

The computer did four things while the student was trying to reach a solution.

(1) It examined each instruction by the student to see if the syntax was correct and was a valid step. If the instruction was incorrect, the computer printed out an error message.

(2) It performed whatever valid step the student commanded, regardless of whether the step contributed to the correct solution.

(3) It compared the solution indicated by "X" with the desired solution. If they were identical, it terminated the problem after typing "correct". If the solution was incorrect, it typed "answer is wrong".

(4) On certain problems it offered a hint after a fixed-time lapse. The hints programmed were usually starting hints. If the student had already completed steps, the hint might no longer be appropriate. Hints were available only for certain problems in the instruction set, not for those in the problem set.

At the end of the six-minute session, the student was signed off automatically

```
COMMITTEE MEMBERS BOUGHT 3 JARS OF CANDY
WITH 14 OUNCES IN EACH JAR, AND 2 BOXES OF CANDY
WITH 27 OUNCES IN EACH BOX. THEY PUT THE CANDY
INTO BAGS THAT CONTAINED 4 OUNCES EACH.
HOW MANY BAGS OF CANDY DID THEY FILL?

G          (1)  3
G          (2)  14
G          (3)  2
G          (4)  27
G          (5)  4
1.2M       (6)  42
3.4M       (7)  54
6.7A       (8)  96
8.5Q       (9)  24
9X

CORRECT
```

Fig. 1. Sample solution of a word problem.

as soon as he completed an unfinished problem, or if he had a two-minute interval with no response.

The word-problem set. Because these fifth-grade students were from an accelerated group, the 68 word problems used in this study were designed to be of appropriate difficulty for sixth-grade students. The students used the rules they learned in the instruction set to solve these problems. As was done in the examples in the instruction set, the computer, after typing out each problem, typed out all the numbers given in the problem as "given numbers". The student then told the computer what to do with these numbers. Figure 1 illustrates how a student went about solving a word problem in this way. The type wheel of the teletype was positioned at the left-hand side of the paper. After the student made his response, the computer positioned the type wheel at the center of the page, typed the line number, and the result of the operation the student had commanded the computer to perform. If the final answer was correct, the computer typed the message "correct" and went on to the next problem. If the final answer was incorrect, the computer typed "answer is wrong" and went to the next problem.

The students were not allowed to use pencil or paper when working on the teletype. Each exercise was worked on the machine, so that all responses could be recorded.

The student was signed off, as during the instruction set, with a "goodbye" message, and "please tear off on the dotted line".

D. *Informal Pilot Study*

Prior to running the experiment, three students at the elementary school served as subjects in an informal pilot study. Each subject was asked to read each problem aloud, to point out any difficult words in the problem, and to indicate how to solve the problem. The student did not actually perform the operations. The entire interview was recorded, later studied, and the following information was extracted from the recordings, some of which resulted in the following modifications of the program or problems.

(1) Four- and five-digit numbers did not have commas separating the hundreds from the thousands digit. For all three students, this led to difficulty in reading the numbers for the five-, but not the four-digit numbers. The five-digit numbers were changed to include commas.

(2) All three students had difficulty with the following problem:

> "Jerry counted 444 names listed on a page in the telephone directory, and there were 55 pages in the book. How many telephone subscribers were listed in his directory..."

"Directory" was changed to "book". "Subscribers" was changed to "names".

(3) The word "equatorial" in the phrase "the equatorial diameter" was dropped; all three students found it difficult to read the word.

(4) Reading difficulty resulted when a phrase was split so that half of it occurred on one line and half on the line just below.

> "Each of the 27 children in Miss
> Brown's room planned to bring 250
> pounds of newspaper..."

was changed to:

> "Each of the 27 children in Miss Brown's room
> planned to bring 250 pounds of newspaper..."

(5) Final sentences such as "How much did they both have", or "how much did they have together" offered real cues as to what operation the problem calls for. However, sentences such as "What was their net gain in yardage" left the students without the slightest idea of what operation to use. This suggested the possibility that the presence or absence of a key word might be a powerful index of item difficulty.

(6) A problem such as the following could be solved in several ways:

> "Paul delivered 140 papers. Of these he delivered 61 on Poplar Street, 58 on Garfield Ave., and the rest on York Road. How many did he deliver on York Road..."

It could be solved:

$$(140 - 61) - 58 \quad \text{or} \quad 140 - (61 + 58).$$

The interesting finding was that all three students used the latter approach in solving this type of problem.

(7) Two out of three students could not solve the following problem:

> "Steve has 13 toy soldiers, Tom has 18 and Richard has 41. What is the average number of toy soldiers..."

Their understanding of the concept of average was unsatisfactory. Some brief instructions were included in the instruction set of the program to teach the students how to do such problems, since most of the division problems required understanding of averaging.

IV. RESULTS

In this section, the main task is to report the predictive worth of the six variables described earlier. The objective is successfully to predict the probability of a correct response for each item. The first step in analysis was to obtain regression coefficients for each of the factors. A multiple linear regression analysis program, adapted for the PDP-1 computer at Stanford, was used to obtain regression coefficients, multiple correlation R and R^2. The regression equation was

$$z_i = -7.36 + .87X_{i1}^* + .18X_{i2} + .02X_{i3} + 2.13X_{i4}^* + .26X_{i5} + 1.42X_{i6}^*.$$

(* indicates significance)

with a multiple R of .67, a standard error of estimate of 1.75, and an R^2 of .45. The results obtained from this model were reasonably successful, considering the complexity of the problems.

From scanning the coefficients, we see that X_4, the sequential variable, is the most important of the six variables. The other weightings indicate that the *conversion* variable, X_6, and the *operations* variable, X_1, are valuable

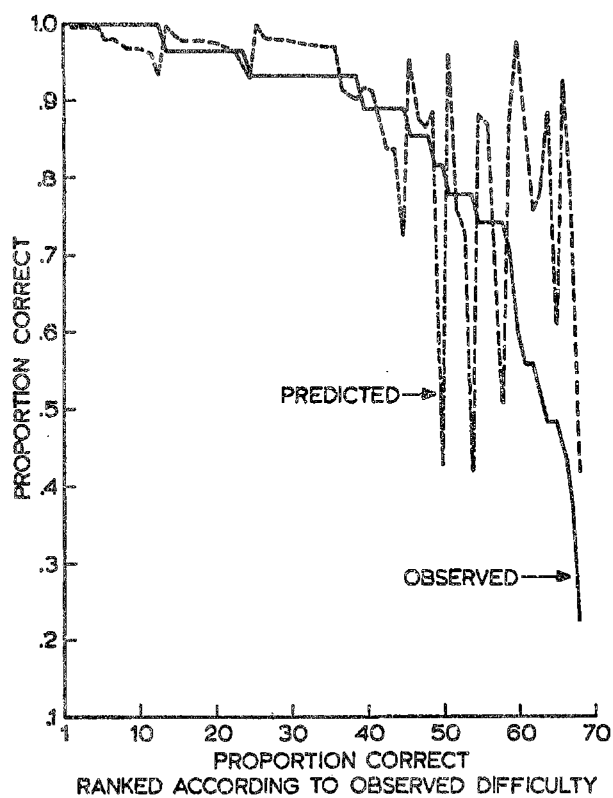

Fig. 2. Problem rank order according to proportion correct.

predictors of the probability of a correct response for each item. A rough indication of the goodness of fit of the regression line is given by the multiple correlation coefficient R and its square, R^2, which is an estimate of the amount of variance accounted for by the regression model. In this case, 45% of the variance in probability of a correct response is accounted for by the model.

Figure 2 presents a graph of the predicted and observed proportions of correct responses for each of the 68 items. The probabilities are plotted as a function of the rank of observed proportion of correct responses.

Consequently, the curve of the observed probabilities is monotonically decreasing and smoother than the predicted curve. An inspection of the two curves shows a reasonable fit for the regression model, especially in view of the heterogeneity of problem types. For an analysis of goodness of fit of the probability of a correct response predicted from the regression model and the observed probability of a correct response, a computer program was written to calculate the predicted probability, p_i, of a correct response for problem i, and to give as a measure of fit χ^2, where

$$\chi^2 = \sum_i (f_i - p_i N)^2 / [p_i (1 - p_i) N]$$

and f_i = observed frequency of correct response, N = number of students. For the above model, $\chi^2 = 555.76$.

This rather high value for χ^2 is an indication of a poor fit, but a closer look at the components of χ^2 shows that a few problems made extremely large contributions to the total χ^2. The following problem, for example, contributed 26% to the total χ^2 obtained:

> "A school playground is rectangular, 273 feet long and 21 feet wide. What is the total length of the fence around the playground..."

The observed proportion of correct responses for this item was .59, while the predicted proportion was .97; clearly, this is a very poor fit. As a second example, the following problem contributed 16% to the total χ^2 obtained.

> "Mary is twice as old as Betty was 2 years ago. Mary is 40 years old. How old is Betty..."

A reduction in χ^2, obtained by deleting those few extreme problems, is still insufficient to yield a value of χ^2 such that the model would normally be accepted. An analysis on a reduced set of data is suggestive, however, and useful. This reduced set excludes seven of the problems that have extreme individual χ^2 contributions. Since calculation of the regression coefficients included the extreme problems, a recalculation of the regression coefficients omitting these problems from the data yields better fits of the model to data than those previously obtained.

We emphasize that this procedure of dropping individual problems with large χ^2 values is certainly not admissible as an inference procedure. The χ^2 values reported here provide a useful descriptive statistic for summarizing the order of magnitude of deviations between the observed and predicted results for the bulk of the problems, and for identifying types of problems, such as those two just mentioned, that require a more elaborate theory.

The regression equation for the reduced set of 63 problems,

$$z_i = -7.85 + .78X_{i1}^* + .29X_{i2} + .02X_{i3} + 2.35X_{i4}^* + .27X_{i5} + 1.33X_{i6}^*$$

(* indicates significance)

has a multiple R of .73, with a standard error of estimate of 1.59, and R^2 of .53. For this reduced set, $\chi^2 = 168.51$.

Consideration of the partial correlation coefficients indicates that most of the variance can be accounted for by X_1, X_2, X_4, and X_6. If we reduce the number of variables in the regression equation to include only these, the reduction in multiple R and R^2 is very slight. Considering only these four variables, the regression equation (for the 63 problems) becomes

$$z_i = -7.55 + .90X_{i1}^* + .30X_{i2} + 2.42X_{i4}^* + 1.34X_{i6}^*$$

with a multiple R of .72, a standard error of estimate of 1.58, and R^2 of .52. For this model, $\chi^2 = 178.33$.

V. DISCUSSION

Although the predictive results of our first relatively crude analysis of the data are far from what we ultimately hope to be able to offer, they are somewhat promising. There is considerable difficulty in intuitively rank-ordering the expected proportions of correct responses obtained in word problems. We believe that our results give a sense of the real possibility of analyzing and predicting in terms of meaningful variables, the response performance of children who are solving arithmetical word problems. At first glance, the problem set appears to be quite complex. Yet, with a few variables we have brought a considerable amount of order to it. The most suggestive single finding is probably the importance of the sequential variable in all the analyses. It is significant beyond the .001 level, indicating that it is clearly an important variable contributing to problem difficulty.

The relatively subtle results obtained in this first study give a clear indication of the difficulty in building a processing model, or, to put it another way, in constructing an explanatory theory that is adequate to account for all the difficulties students encounter in solving word problems. From a theoretical standpoint, it is apparent that nothing short of a full syntactic and semantic analysis will suffice to predict all the details that must be accounted for in the behavior of students. Even then, it is not simply a matter of an abstract syntax and semantics for some significant portion of English or another natural language; it is a matter of having a behaviorally sensitive syntax and semantics. Many mathematicians concerned with mathematics education perhaps do not appreciate sufficiently that until better fundamental theories are available, certain directions of deeper progress in mathematics education are hardly possible. The present study was meant to be a very modest step in a direction of much needed additional research.

Inst. for Mathematical Studies in the Social Sciences,
Stanford University

BIBLIOGRAPHY

Braun-Lamesch, M.: 1962, 'Le rôle du contexte dans la compréhension du langage chez l'enfant', *Psychol. Franc.* 7, 180–189.

Brownell, W. and Stretch, L. B.: 1931, 'The Effect of Unfamiliar Settings on Problem-Solving', *Duke University Research Studies in Education*, No. 1, Duke University Press, Durham, North Carolina.

Ervin, S.: 1964, 'Imitation and Structural Change in Childrens' Language', in E. Lenneberg (ed.), *New Directions in the Study of Language*, M.I.T. Press, Cambridge, Mass., pp. 163–189.

Miller, W. and Ervin, S.: 1963, 'Language Development', in *Child Psychology*, 62nd Yearbook for the National Society for the Study of Education, pp. 108–143.

Post, R.: 1958, 'A Study of Certain Factors Affecting the Understanding of Verbal Problems in Arithmetic', Unpublished doctoral dissertation, Teachers College, Columbia University, New York.

Suppes, P., Jerman, M., and Groen, G.: 1966, 'Arithmetic Drills and Review on a Computer-Based Teletype', *Arithmetic Teacher*, pp. 303–309.

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

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

REFERENCES

* This research was supported by the National Science Foundation and the U.S. Office of Education. We are indebted to Professor Leon Henkin for several specific suggestions that are acknowledged in appropriate places in the text, and also to Mr. Roulette Smith for programming assistance.

[1] To take care of the case when the observed p_i is either 0 or 1, we use the following transformation

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

where n_i = the total number of subjects responding to item i. The exact form of this transformation is not important.

[2] To avoid any ambiguity, we always first minimize the number of steps and then the number of operations.

[3] The reason for designing the program in this way was to reduce the time required for students to input large numbers themselves.