DOCUMENT RESUME

ED 113 212 95 SE 019 830

AUTHOR Roman, Richard A.

TITLE The Word Problem Program: Summative Evaluation.

INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.; National Science Foundation, Washington, D.C.
REPORT NO LRDC-1975-23
PUB DATE Jul 75
NOTE 14p.

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS *Computer Assisted Instruction; *Curriculum; Curriculum Development; Elementary Education; *Elementary School Mathematics; Evaluation; Individualized Instruction; Instruction; *Problem Solving; *Summative Evaluation; Testing

ABSTRACT

A programed sequence for teaching students to solve word problems was developed using a combination of the information processing and structural variables approaches. Students using the sequence proceeded individually through mastery of a sequence of objectives. In order to evaluate the program, fourth and fifth graders were randomly selected from classes; the remaining students in these classes served as controls. All students were given the appropriate level of the Stanford Acheivement Test as a pretest. During the 11 weeks that experimental subjects completed the Word Problem Program, control subjects received regular mathematics instruction. The computation and applications sections of the Stanford Acheivement Test served as posttests. Both fourth- and fifth-grade experimental groups scored higher on their respective applications posttests than the comparable control groups. (SD)

THE WORD PROBLEM PROGRAM: SUMMATIVE EVALUATION

Richard A. Roman

Learning Research and Development Center

University of Pittsburgh

July, 1975

This research was financed by a grant from the National Science Foundation (NSF-GJ-540X) and by the Learning Research and Development Center, supported in part by funds from the National Institute of Education, United States Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of the sponsoring agencies and no official endorsement should be inferred.

2/3

Abstract

A computer assisted instruction program to teach arithmetic word problems was designed, developed and tested in a school setting. Experimental groups of fourth and fifth graders each gained significantly when compared to control groups on national standardized tests of arithmetic applications.

THE WORD PROBLEM PROGRAM: SUMMATIVE EVALUATION

Richard A. Roman

University of Pittsburgh

Teaching elementary school students to solve arithmetic word problems has been a difficult task and has generated much research (Gorman, 1967; Suydam & Riedesel, 1969). Many mathematicians regard arithmetic word problems as crucial in mathematical development, so the problem must be confronted (Polya, 1962). Earlier reports described the design, rationale and formative evaluation of the Word Problem Program, a computer assisted instruction program that teaches students to solve word problems (Roman & Laudato, 1974; Laudato & Roman, 1975; Laudato, 1975). This report describes a summative evaluation study of that program in a school setting.

Method

Design

The study was a pretest, treatment, posttest design, with one experimental group and one control group. The experimental treatment consisted of work on the Word Problem Program. The control group received only that instruction in word problems provided in the normal mathematics curriculum. The study was replicated in fourth and fifth grade. Since the fourth and fifth graders received different tests, the results were analyzed separately for the two grades.

Materials

The Word Problem Program utilizes a combination of the information processing approach of Bobrow (1968) and Paige and Simon (1966) with the structural variable approach of Loftus and Suppes (Loftus, 1972, Suppes, Fletcher, Zanotti, Lorton, & Searle, 1973) to create and sequence instructional objectives, each of which consists of problems of homogenous difficulty. Each problem in an objective requires the same number and same kind of information processing steps to reach a solution.

Once in the program, students receive a sequence of objectives based on their individual performance. The program evaluates performance continuously and tells the students when they finish each objective. Work within an objective consists of a set of problems selected from the target group and practice groups that have already been mastered. The students cannot distinguish between target and practice problems, but only their work on target problems contributes to the evaluation of objective performance.

The literature supports the importance of students solving many problems when learning word problems. No method has consistently been found superior to a carefully sequenced set of problems (Laudato, 1975). There are indications that the use of analytic steps can improve performance (Suydam & Riedesel, 1969). The Word Problem Program, therefore, offers three analytic hints to each problem. Students may request those hints at any time.

All calculations are carried out by the computer under the student's direction. Typical computational commands utilize letters as variable names, for example, "A + B," or "(C + D) / A," or "D - C." The decision to use the computational power of the computer allows the program to use large numbers, larger than those the student could calculate. Numbers are chosen to be particularly deceptive with regard to the correct operation. For example, in single step problems, one number is always a multiple of the other. While this situation provides a cue for division, in the Word Problem Program division is correct only one-fourth of the time. In other curricula, divisibility of the numbers reliably cues the division operation.

Details about the program, including the methods of problem generation, problem selection, problem analyses, objective sequencing, and hint preparation can all be found in Roman and Laudato (1974), and Laudato (1975).

Subjects

A suburban, lower middle class, elementary school housed the computer upon which the Word Problem Program was implemented. The fourth and fifth grade teachers agreed to allow random sampling of their classrooms for this study, and accordingly, one-third of the students from each of four classrooms (two fourth grades, two fifth grades) were randomly selected for participation. In a conference before the study, the teachers rejected seven of the 28 randomly selected students. Reasons for rejection included: never finishes work, so can't afford time (three cases); knows how to do word problems already (two cases); and behaves badly and is not allowed to use computer (two cases). The rejected students were replaced by seven other students randomly selected from the same classrooms.

Procedure

All students in fourth and fifth grades took appropriate levels (Intermediate I and II, respectively) of the Stanford Achievement Tests (Madden, Gardner, Rudman, Karlsen, & Merwin, 1973) in January as part of the district's normal testing procedure. The subtests on computational skills and mathematical applications served as pretests for the study. The computational subtest was selected since success on word problems requires correct computation, and students with differential ability to compute should perform differently. Since the Word Problem Program requires no computation, no gains were expected from the training. The mathematical applications subtest consists of approximately 30 word problems and 10 problems related to graphs and charts. Gains were expected on the subtest due to the training.

In February, each experimental student was instructed on how and when to use the program and completed two introductory sessions. Each experimental subject then had a number of sessions on the program during time that was devoted to mathematics study by the control subjects. The number of sessions varied from 4 to 47 (mean 18.4, S. D. 10.81). Sessions stopped when the teacher or student felt that learning had ceased. Feedback from the program on which objectives were attempted and which were mastered was provided after each session to help the teacher and student make their decisions.

The study lasted eleven weeks, including one week of school vacation. The termination of the study coincided with previously scheduled schoolwide testing. When the study ended, 13 of the 28 students were still actively working. Indeed, several continued for five weeks after the study ended and quit only because school closed for the summer. The other fifteen students worked from three to nine weeks.

The control students worked on their regular mathematics curriculum when the experimental students used the computer. No special attention was given to word problem study for the control group.

The computation and applications subtests of the Stanford Achievement Test (Kelley, Madden, Gardner, & Rudman, 1964) were included in a special testing battery in late April. These tests were given as part of the Learning Research and Development Center's continuous monitoring program in the school. These subtests served as the posttests for the study.

Results

The raw scores for all pretests, computation, and applications were analyzed separately to determine if the experimental and control groups differed. The fifth grade experimental group scored significantly higher than the controls on the applications pretest, $F(1,43) = 4.593$; $p < .05$. No other differences were detected. Because of this difference, the analysis of the posttest was done as an analysis of covariance.

The raw scores from the tests were analyzed by a one-way analysis of covariance for each grade level separately. Covariates were the two pretest scores. The fourth grade experimental group scored significantly higher on the applications posttest, $F(1,29) = 4.398$, $p < .05$, as did the fifth grade experimental group, $F(1,42) = 4.364$, $p < .05$. No differences were observed for the computation posttest. The mean scores for all groups are given in Table 1.

Table 1

Applications Posttest Raw Score Means and Means Adjusted for Pretest Scores

Group	*N*	Test Data: Raw Means	Adjusted Means	Adjusted Standard Error
Fourth[a]				
Control	21	14.86	15.39	1.00
Experimental	12	19.83	18.90	1.33
Fifth[b]				
Control	30	17.30	19.00	0.74
Experimental	16	24.88	21.69	1.00

Note. Means are not comparable across grades.

[a]Fourth grade subjects took Stanford Achievement Test, Intermediate I (1964).

[b]Fifth grade subjects took Stanford Achievement Test, Intermediate II (1964).

Looking only at experimental students, some measures of within-treatment performance were examined. On most criteria, no differences between the grades were observed. However, the fifth graders mastered significantly more objectives that the fourth graders, $F(1,22) = 9.7853$, $p < .01$. The respective means were 6.92 and 4.92. Other program variables considered included. total time in treatment, number of problems seen, time per session, and number of days worked. Means and standard deviations for these measures are presented in Table 2. Since the groups did not differ in these measures, Table 2 presents data for the combined groups.

Table 2

Means and Standard Deviations for Internal Treatment Measures
Fourth and Fifth Grades Combined

Measures	N	Mean	SD
Total time	28	4 625 (hours)	83
Total problems	28	177 79	80 10
Time/problem	28	1.57 (minutes)	31
Days worked	28	15 57	8 89
Time/day worked	28	18 72 (minutes)	3 60
Objectives mastered	28	8 04	5 70

Discussion

Students spending approximately four and a half hours over a period of two and a half months working on the Word Problem Program increased their knowledge of arithmetic applications, as measured by a national standardized test, approximately five months more than their control group classmates. The gain is both significant and important since it occurs in an area that has traditionally been difficult to teach and was achieved with a small time investment.

The fact that no gain occurred on computational skills was expected since the Word Problem Program provides no computational practice. The gain in the application posttest reflects greater problem solving skills, as opposed to greater computational skills.

The applications subtest of the Stanford Achievement Tests contains many problems that require an understanding of money, time, fractions, percentage, and units of measurement such as ounces, degrees, and gallons. None of these concepts are treated in the Word Problem Program. The finding of significant gains then takes on added significance since the results demonstrate a transfer effect from a carefully selected subdomain to a larger group of problems. This implies that the fundamental problem solving skills taught have generality beyond their initially limited scope.

The fact that fifth graders master more objectives than fourth-graders is not surprising, but in view of the unorthodox manner in which the objectives are defined and sequenced, this finding serves to confirm at least the gross features of the design of objectives. Further analysis of the data should reveal information regarding the hierarchical assumptions made in the program.

This study will be carried further, the data generated during each student session will be analyzed to determine if specific features designed into the Word Problem Program accomplish their purpose. Among the aspects to be investigated are: the effectiveness of hints, the effect of sequencing decisions, the validity of the evaluation rule, and the homogeneity of problems within objectives.

The next goals for the program are to design a procedure to place students at their initial competence level more quickly, probably based on initial performance measures within the program, and to determine internally when the student has "stopped learning," a determination that currently requires teacher or student decision making. With these improvements, the program will operate more efficiently; and the time required of each student may be further reduced.

References

Bobrow, D. G. Natural language input for a computer problem-solving system. In M. Minsky (Ed.), *Semantic information processing*. Cambridge: Massachusetts Institute of Technology, 1968.

Gorman, C. J. *A critical analysis of research on written problems in elementary school mathematics*. Unpublished doctoral dissertation, University of Pittsburgh, 1967.

Kelley, T. L., Madden, R., Gardner, E. F., & Rudman, H. C. *Stanford achievement test*. New York: Harcourt, Brace, & World, Inc., 1964.

Laudato, N. C. *The design and development of a computer-assisted instructional program in word problem solving for the elementary school*. Unpublished doctoral dissertation, University of Pittsburgh, 1975.

Laudato, N. C., & Roman, R. A. *Computer-assisted instruction in word problems: Design and assessment*. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1975. (LRDC Publication 1975/12)

Loftus, E. J. F. *An analysis of the structural variables that determine problem solving difficulty on a computer-based teletype*. Stanford: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1970. (Technical Report No. 162)

Madden, R., Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, J. C. *Stanford achievement test*. New York: Harcourt Brace Jovanovich, Inc., 1973.

Paige, J. M., & Simon, H. A. Cognitive processes in solving algebra word problems. In B. Kleinmetz (Ed.), *Problem solving*. New York: Wiley, 1966.

Polya, G. *Mathematical discovery*. New York: Wiley, 1962.

Roman, R. A., & Laudato, N. C. *Computer-assisted instruction in word problems: Rationale and design*. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1974. (LRDC Publication 1974/19)

Suppes, P.; Fletcher, J. D., Zanotti, M., Lorton, P., & Searle, B. W. Evaluation of computer-assisted instruction in elementary mathematics for hearing-impaired students. Stanford: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1973. (Report No. 200)

Suydam, M., & Riedesel, C. A. Interpretative study of research and development in elementary school mathematics (3 vols.). Final Report. College Park: Pennsylvania State University, 1969.