EVALUATION OF COMPUTER-ASSISTED INSTRUCTION IN ELEMENTARY MATHEMATICS FOR HEARING-IMPAIRED STUDENTS

BY

P. SUPPES, J. D. FLETCHER, M. ZANOTTI,
P. V. LORTON, JR., and B. W. SEARLE

TECHNICAL REPORT NO. 200

MARCH 17, 1973

PSYCHOLOGY AND EDUCATION SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

EVALUATION OF COMPUTER-ASSISTED INSTRUCTION IN ELEMENTARY MATHEMATICS FOR HEARING-IMPAIRED STUDENTS

by

P. Suppes, J. D. Fletcher, M. Zanotti,

P. V. Lorton, Jr., and B. W. Searle

TECHNICAL REPORT NO. 200

March 17, 1973

PSYCHOLOGY AND EDUCATION SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

EVALUATION OF COMPUTER-ASSISTED INSTRUCTION IN ELEMENTARY MATHEMATICS FOR HEARING-IMPAIRED STUDENTS*

P. Suppes, J. D. Fletcher, M. Zanotti,
P. V. Lorton, Jr., and B. W. Searle

INTRODUCTION

The Institute for Mathematical Studies in the Social Sciences at Stanford University (IMSSS) has been developing curriculums and techniques for computer-assisted instruction (CAI) since 1963. In 1970 the Office of Education funded the Institute for CAI research in schools for the deaf. During the 1970-71 school year approximately 1,000 students in schools for the deaf participated in the project, and during 1971-72 this number increased to more than 2,000 students taking CAI lessons at 15 schools for the deaf located in four states and the District of Columbia.

The students who participated in the experiment reported in this article were chosen from among the entire population of deaf students receiving CAI lessons in mathematics and language arts through the IMSSS system in 1971-72. The degree of hearing loss among the students was essentially that adopted for admission standards by the schools. Generally, this loss is at least 60 decibels in the better ear. Students selected by the schools for CAI represent a cross section of

---

*This research was supported by Office of Education Grant OEG-0-70-4797(607), OE Project No. 14-2280.

their elementary and secondary school population. Some of these students may be significantly handicapped in ways other than that of hearing loss, but these additional handicaps do not prohibit the students from meeting primary school advancement requirements.

This experiment evaluated the effectiveness of the CAI mathematics program on the acquisition of computational skills. We first describe the models of evaluation tested, and then the curriculum, before turning to the analysis of data. In the models tested, we have emphasized variation in intensity of experimental treatment, as opposed to simply comparing experimental with control groups.

THE MODELS

To investigate the relationship of posttreatment scores to pretreatment scores and the number of CAI mathematics strands sessions given, five models were tested. In all five models, T_{i1} is the pretreatment score of student i, T_{i2} is the posttreatment score of student i, and N_i is the number of CAI mathematics strands sessions taken by student i. Following standard notation, $E(T_{i2})$ is the expected posttreatment score of student i.

Model I: Linear.

$$E(T_{i2}) = a_0 + a_1 T_{i1} + a_2 N_i \ .$$

In this model, the expected effect of pretreatment score and number of sessions on posttreatment performance is assumed to be linear.

Model II: Linear with interaction.

$$E(T_{i2}) = a_0 + a_1 T_{i1} + a_2 N_i + a_3 T_{i1} N_i \ .$$

In Model II, a linear effect of pretreatment score and number of sessions is assumed, but a linear effect from the interaction of pretreatment score and sessions is also postulated.

Model III: Cobb-Douglas.

$$E(\ln T_{i2}) = a_0 + a_1 \ln T_{i1} + a_2 \ln N_i \ .$$

Model III is based on a formulation of the Cobb-Douglas type (from econometrics), namely,

$$T_{i2} = a_0 T_{i1}^{a_1} N_i^{a_2} \ .$$

This model is multiplicative and assumes a 'weighted interaction' between pretreatment score and number of sessions in accounting for change in posttreatment scores.

Model IV: Log quadratic.

$$E(T_{i2}) = a_0 + a_1 T_{i1} + a_2 \ln N_i + a_3(\ln N_i)^2 + a_4(\ln N_i)^3 \ .$$

In Model IV, the effect of the pretreatment score is assumed to be linear, but the effect of number of sessions is assumed to be logarithmic, rather than linear. In order to explore this logarithmic assumption fully, we included second- and third-order terms in $\ln N_i$.

Model V: Exponential.

$$E(\ln T_{i2}) = a_0 + a_1 N_i T_{i1} \ .$$

Model V is based on an exponential formulation, namely,

$$T_{i2} = a_0 e^{a_1 N_i T_{i1}} \ .$$

In this model, the effect of number of sessions and pretreatment score may be strictly increasing or strictly decreasing, depending on the sign of a_1. Pretreatment score and number of sessions are assumed to interact.

In each of these models we treat pretreatment and posttreatment scores separately, i.e., we do not, for instance, regress the difference in the two scores on the number of CAI sessions. Even though reasons for avoiding difference scores have been discussed from several standpoints in the literature of evaluation (e.g., Cronbach & Furby, 1970; Lord, 1963), we give here a direct but elementary analysis from first principles. So far as we know, this argument has not appeared in this form in the literature.

Let X, Y, and N be random variables with $E(X) = E(Y) = E(N) = 0$, and with finite, nonzero variance. The pairwise correlation coefficients of these random variables are then well defined. Consider now the correlation

$$r(Y - X, N) = \frac{E((Y - X)N)}{\sigma(Y - X)\sigma(N)} . \tag{1}$$

From the linearity of the expectation operator,

$$E((Y - X)N) = E(YN) - E(XN) . \tag{2}$$

Further, since by hypothesis $E(X) = E(Y) = 0$,

$$\begin{aligned} \sigma^2(Y - X) = E((Y - X)^2) &= E(X^2) + E(Y^2) - 2E(XY) \\ &= E(X^2) + E(Y^2) - 2r(X,Y)\sigma(X)\sigma(Y) \\ &= \sigma^2(X) + \sigma^2(Y) - 2r(X,Y)\sigma(X)\sigma(Y) . \end{aligned} \tag{3}$$

Substituting (2) and (3) into equation (1) we obtain:

$$r(Y - X, N) = \frac{E(YN) - E(XN)}{\sigma(N)\{\sigma^2(X) + \sigma^2(Y) - 2\sigma(X)\sigma(Y)r(X,Y)\}^{1/2}} .$$

Let us now assume that the correlation of X and N is zero, i.e.,

$$r(X,N) = 0 \qquad \text{(I)}$$

and also that X and Y are highly correlated, but $r(X,Y) < 1$, and that approximately

$$\sigma(X) \approx r(X,Y)\sigma(Y) \, , \qquad \text{(II)}$$

then, on the basis of (I) and (II)

$$r(Y - X,N) \approx \frac{E(YN)}{\sigma(N)\{\sigma^2(Y) - \sigma^2(Y)r^2(X,Y)\}^{1/2}}$$

$$\approx \frac{r(Y,N)\sigma(Y)\sigma(N)}{\sigma(N)\sigma(Y)(1 - r^2(X,Y))^{1/2}} \, .$$

Therefore,

$$r(Y - X,N) \approx \frac{r(Y,N)}{\sqrt{1 - r^2(X,Y)}} \, . \qquad (4)$$

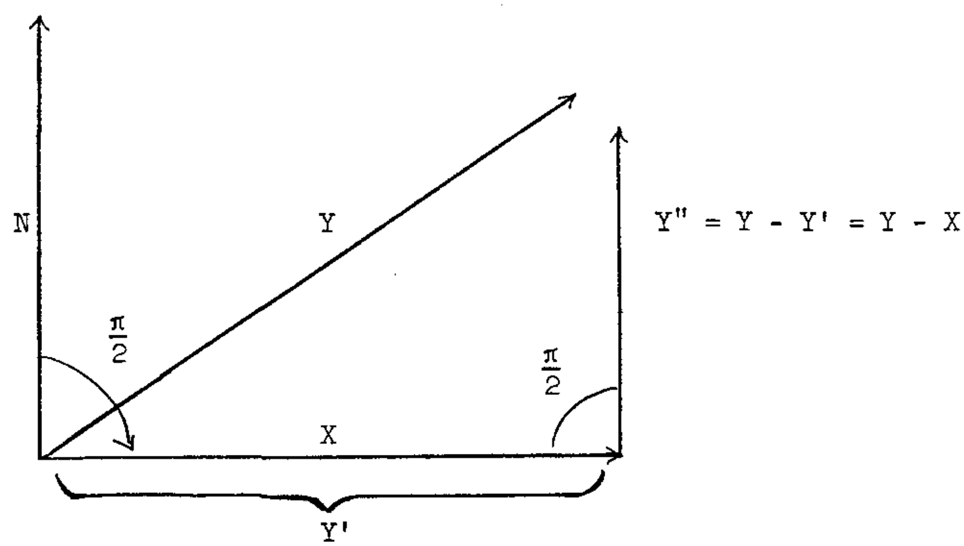
Equation (4) shows that with appropriate choice of a random variable X, $r(X,Y)$ large, and X orthogonal to N, we can obtain a correlation coefficient, $r(Y - X,N)$, as close to 1 as we please so long as $r(Y,N)$ is not zero. The correlation of N with the Y - X difference score may depend solely on the correlation of Y with X and may have nothing to do with the effect of N. For that matter,

$$r^2(Y,N) + r^2(X,Y) = 1$$

implies that

$$r(Y - X,N) = 1 \, .$$

The geometrical interpretation is clear:



Assumption (II), $\sigma(X) \approx r(X,Y)\sigma(Y)$, is the requirement that the norm of the projection Y' of Y on X approximately equals the norm of X.

Assumptions (I) and (II) hold approximately for many experiments using Y as the posttreatment measure, X as the pretreatment measure, and N as some measure of the amount or intensity of treatment.

From these and other results in the literature about difference scores, it is clear that studies of evaluation that use models built around difference scores must be approached with considerable caution. We numerically illustrate later in the analysis of data pertinent to the five models these remarks about difference scores, even though for practical purposes we include some general remarks about gains in the analysis of data.

THE MATHEMATICS STRANDS CURRICULUM

The research reported here attempts to assess the pedagogical effect of the Institute's elementary mathematics curriculum on achievement among the hearing-impaired students. CAI in elementary mathematics provided by the Institute has had a history of success with hearing students (Suppes & Morningstar, 1969, 1970a), and comparable success was anticipated for students in schools for the deaf.

The objectives of the strands program are (a) to provide supplementary individualized instruction in elementary mathematics at a level of difficulty appropriate to each student's level of achievement, (b) to allow acceleration in any concept area in which a student demonstrates proficiency, and to allow repeated drill and practice in areas of deficiency, and (c) to report a daily profile of each student's progress through the curriculum.

A strand is a series of exercises of the same logical type (e.g., horizontal addition, vertical subtraction, multiplication of fractions) arranged sequentially in equivalence classes according to their relative difficulty. The 14 strands in the program and the grade levels spanned by each strand are shown in Table 1. Each strand contains either five

Insert Table 1 about here

or ten equivalence classes per half year, with each class labeled in terms of a grade-placement (GP) equivalent. The GP of equivalence classes was determined by analysis of three major elementary-school mathematics texts (Clark, Beatty, Payne, & Spooner, 1966; Eicholz &

TABLE 1

Grade Level Spanned by Each Strand in
the Elementary Mathematics Program

Strand	Content	Grade level
NUM	Number concepts	1.0-7.9
HAD	Horizontal addition	1.0-3.9
HSU	Horizontal subtraction	1.0-3.4
VAD	Vertical addition	1.0-5.9
VSU	Vertical subtraction	1.5-5.9
EQN	Equations	1.5-7.9
MEA	Measurement	1.5-7.9
HMU	Horizontal multiplication	2.5-5.4
LAW	Laws of arithmetic	3.0-7.9
VMU	Vertical multiplication	3.5-7.9
DIV	Division	3.5-7.9
FRA	Fractions	3.5-7.9
DEC	Decimals	4.0-7.9
NEG	Negative numbers	6.0-7.9

O'Daffer, 1968; Suppes, 1966). Data collected during several years of the earlier drill-and-practice mathematics program at Stanford were used to arrange the equivalence classes in an increasing order of difficulty and to insure that new skills (e.g., regrouping in subtraction) were introduced at the appropriate point.

In addition to ordering the equivalence classes within a strand, it was necessary to determine how much emphasis to give each strand at a given grade level. To determine this emphasis, we divided the curriculum into 14 parts, each corresponding to a half year. A probability distribution was defined for the proportion of problems on each strand for each half year. Both the problem count from the three textbook series mentioned above and the average latency for problem types based on past data were used to define the curriculum distribution. The final proportions in terms of time and problems for each half year for each strand are shown in Table 2.

Insert Table 2 about here

The analysis contained in Table 2, which embodies not only empirical analysis, but also some normative decisions about relative emphasis in curriculum, is one of the few explicit quantitative analyses of curriculum distribution in elementary-school mathematics to be found anywhere in the mathematics education literature.

A student's progress through the strands structure is purely a function of his own performance and is independent of the performance of other students; in fact, his progress on a given strand is independent of his own performance on other strands. A scheme defining movement

TABLE 2

Proportion of Time and Proportion of Problems for Each Strand for Each Half Year

Strand		Half year													
		1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5
NUM	PT	50	24	24	17	10	5	7	7	8	11	14	10	15	15
	PP	36	18	16	12	10	4	8	8	10	14	20	10	19	19
HAD	PT	26	21	21	9	14	9								
	PP	32	28	26	10	14	8								
HSU	PT	14	10	16	9	4									
	PP	18	14	16	10	4									
VAD	PT	10	10	9	19	19	7	8	2	3	1				
	PP	14	12	12	22	20	6	10	2	4	2				
VSU	PT		9	8	15	22	10	13	3	3	1				
	PP		12	12	18	20	8	10	2	4	2				
EQN	PT		17	12	16	17	14	17	7	5	7	7	8	15	15
	PP		10	10	12	16	12	20	8	8	12	12	10	19	19
MEA	PT		9	10	8	8	11	7	7	5	5	5	5	5	5
	PP		6	8	6	6	6	8	8	8	8	8	6	6	6
HMU	PT				7	3	8	5	3	2					
	PP				10	6	14	10	6	8					
LAW	PT					3	5	5	3	3	3	1	1	8	8
	PP					4	6	6	4	6	6	2	2	10	10
VMU	PT						10	5	14	6	8	7	5	8	8
	PP						14	6	16	8	4	4	2	4	4
DIV	PT						15	22	34	48	33	40	13	14	14
	PP						18	10	16	16	6	8	2	3	3
FRA	PT						6	4	15	13	20	17	18	10	10
	PP						4	4	24	22	32	32	26	10	10
DEC	PT							7	5	4	11	7	36	10	10
	PP							8	6	6	14	10	38	10	10
NEG	PT											2	4	15	15
	PP											4	4	19	19

Note.--PT = proportion of time; PP = proportion of problems.

through a strand uses the pattern of correct and incorrect responses to insure a rate of movement that reflects performance. The movement scheme will be described in detail elsewhere (Suppes, Searle, & Lorton, in preparation). The overall determination is based on a model that assumes independence of responses for problems of a given equivalence class and is defined so that a student with average performance gains one year's GP in one school year of CAI time, which ranges from 6 to 10 minutes per school day. A detailed description of the strands curriculum is given by Suppes, Goldberg, Kanz, Searle, and Stauffer (1971).

DESIGN OF THE EXPERIMENT

Equipment

The central computer processor was the Institute's PDP-10 system located on the Stanford campus. On-line, real-time communication was maintained with the participating schools located in California, Florida, Oklahoma, Texas, and the District of Columbia by means of dedicated telephone lines.

The student terminals were KSR Model-33 teletypewriters. The teletypewriters communicate information to and from the central computer system at a rate of about 10 characters per second. All of the elementary mathematics exercises were typed at the terminal under computer control, and keyboard responses were given by the students. The details of exercise format and student responses are described in Suppes, Jerman, and Brian (1967) and Suppes and Morningstar (1972).

Students and Number of Sessions

The students participating in this experiment included the entire population receiving the elementary mathematics and language arts curriculum whose current, average GP on the mathematics strands curriculum was equal to or greater than 2.4 and equal to or less than 5.9, and who had taken at least 15 mathematics CAI sessions.

Five levels of intensity for number of CAI sessions were selected. Treatment groups 1, 2, 3, 4, and 5 were assigned 10, 30, 70, 100, and 130 sessions, respectively, and 77 students were randomly assigned to each group.

Session limits were imposed on a calendar basis so that students with low numbers of sessions received them distributed throughout the experimental period. A participating student had no control over the type of lesson, mathematics strands or language arts, he received. Whether he signed on for strands or language arts he was given a mathematics lesson if he was eligible for one. Otherwise, he received a language arts lesson.

Eligibility for a mathematics session was decided according to the following algorithm.

$$\text{If} \qquad NS_i < \left[D_i \left(\frac{TS_i}{TD_i} \right) + 2 \right] \text{ and } TS_i \leq TD_i \text{ ,}$$

then student i received a mathematics session; otherwise, he received a language arts session. In the algorithm,

NS_i = number of sessions taken during the experimental period by student i,

TS_i = total number of sessions student i was to receive during the experimental period,

TD_i = number of calendar days in the experimental period for student i,

D_i = number of calendar days elapsed for student i in the experimental period,

and the brackets denote the next greatest integer. For example, suppose a student was in a 70-sessions group ($TS_i = 70$) and the experimental period was 150 days ($TD_i = 150$). If on the 21st day he was in the experiment ($D_i = 21$) he had taken 11 sessions ($NS_i = 11$), then he would have received a mathematics session when he signed on, since

$$11 = NS_i < \left[D_i \left(\frac{TS_i}{TD_i}\right) + 2\right] = \left[21 \left(\frac{70}{150}\right) + 2\right] = 12 \ .$$

Some students signed on more than once a day in order to obtain the assigned number of sessions.

The actual number of mathematics sessions a student received was monitored daily. The assistance of teachers and proctors was sought to help students achieve the number of sessions they were assigned. Teachers were urged not to give compensatory off-line work to those students assigned to low numbers of on-line sessions, and, in general, not to alter the classroom work of any student because of his participation in the experiment.

THE MODIFIED STANFORD ACHIEVEMENT TEST

One aim of the experiment was to develop a reliable and valid test of achievement that could be administered 'on line', i.e., by computer, at student terminals. It was hoped that administering the test on line would standardize procedures for giving it, that the schools would be spared the difficulties of having yet another test to administer, and that the clerical task of getting test scores into the computer for data analysis would be minimized. The Arithmetic Computation Section of the Stanford Achievement Test (SAT) was used as a guide in developing a modified SAT (MSAT) that could be administered on line at student terminals.

Description of the SAT

The Arithmetic Computation sections of three SAT batteries, Primary II, Intermediate I, and Intermediate II, were used as models. For all three levels, forms W and X were examined.

Computation problems on the Primary II test are worked in the test booklet. All questions on the computation sections of the Intermediate level tests (I and II) are multiple choice, and responses are entered in the test booklet or on a standard answer sheet. The responses are letters between *a* and *j*. Each question has four or five possible responses; the letters *a* - *d* and *a* - *e* are used for one question, the letters *e* - *h* or *f* - *j* are used for the next question.

Several properties of the tests are shown in Table 3. The table contains, for each level, the number of problems, the time allowed for

Insert Table 3 about here

TABLE 3

Characteristics of Computation Section of the SAT

Test level	Number of problems	Time (min.)	Placement on form W			Placement on form X		
			.25	.50	.75	.25	.50	.75
Primary II	60	30	2.4	3.4	4.4	2.2	3.3	4.4
Intermediate I	39	35	3.5	4.6	5.8	3.5	4.7	6.0
Intermediate II	39	35	4.4	6.1	8.5	4.6	6.2	8.7

administration, and the normed GP corresponding to error rates of .75, .50, and .25. These last figures are shown separately for forms W and X.

Construction of the MSAT

Three sets of parallel items were constructed for each item in the computation section of the SAT. The method used to construct items depended on whether the type of SAT item being considered appeared in the CAI mathematics strands curriculum.

Each SAT problem of a type that occurs in the CAI curriculum was labeled with the equivalence class number into which it fitted. Both forms (W and X) were examined and labeled. If similar items for both SAT forms fell in the same equivalence class, three items from that equivalence class were used; if not, a new class definition was constructed that included the attributes of both classes, and three problems were written to fit this definition.

If a test problem was of a type that did not occur in the strands curriculum, a new class definition was written that described the salient features of the exemplars from both test forms. Then three items were written using the definition.

A careful analysis was made of the answer choices presented in Intermediate I and II. The distractors for each addition problem differ from the correct answer in one digit that occurs in any but the left-most position. The distractors for each subtraction problem differ from the correct answer in one digit that occurs in any but the left-most position, or the left-most digit is one less than that of the correct answer.

Two of the distractors for each multiplication problem in the SAT differ from each other in one digit and are both one digit too short. Other possible wrong answers differ from the correct answer in any but the left-most digit. In the case of a two-digit multiplier, one wrong answer for the MSAT was constructed by using only one of the digits of the multiplier to obtain a product. Distractors for decimal problems using dollar signs are the same as those for comparable integer problems. In other decimal problems in the SAT, the wrong answers differ from the correct answer in the placement of the decimal point; specifically, they are chosen from $(CA)(10^K)$, $K = \pm 1, \pm 2$, where CA is the correct answer.

Division problems in the SAT occur in three formats: $A/\overline{B}$, $B \div A$, and $1/A$ of B. Some choices have a remainder and others do not, regardless of the correct-answer remainder. Some remainders are larger than the divisor. The same method was used to construct MSAT answer choices for all three types of division problems. Distractors were constructed by changing one of the digits of the quotient, and where appropriate, choosing a random remainder.

One dilemma was encountered in constructing answers for division problems. Two problems on the X form of Intermediate I and one on the W form have two technically correct answers. The following item exemplifies these problems.

$4\overline{)208}$

A) 42
B) 50, REM 8
C) 52
D) 5, REM 8
E) NOT GIVEN

Although C is the 'correct' answer, an argument can be made that B is also correct. This answer choice was not used in constructing wrong answers for the MSAT. In both Intermediate I and Intermediate II there are four problems for which the correct answer is "not given."

Administration of the MSAT

The MSAT was administered by randomly selecting one of the three items written for each problem number. Thus, while approximately one-third of the items on one MSAT administration would be repeated on a second administration, it was unlikely that any two students sitting next to each other would receive the same test. A record was kept of the items that comprised each student's test.

The MSAT was administered as much like the SAT as possible. The major exception concerns the possibility of students taking the paper and pencil SAT to return to problems already completed or skipped over. Students may skip any problem on the MSAT simply by typing the return key on the teletype terminal. However, once a student answers or skips an MSAT problem there is no way for him to return to it.

Tests were administered at the terminal, preferably at one sitting. If a student did not complete the test in one sitting, he began where he left off at his next sign on. The maximum time allowed to complete the test was the time allotted for administration of the computation section of the SAT, plus one-half the time it takes to type the test at the terminal.

Response modes on the MSAT were as close as possible to those on the SAT. For multiple-choice items the student types a single letter. For constructed responses the student types the numerical answer as he would for an ordinary strands item with the strands error response suppressed. The student may alter his response using the rubout key. When he has completed his answer, he types the return key. If a student working a multiple-choice problem types a number instead of a letter, he is told to type a letter and given another chance. As soon as the return key is typed the program moves to the next problem. There is no time limit for individual problems. Many of the problems require pencil and paper computation.

The procedures that differed from the usual mode of responding at the teletype were explained to the student. A message at the beginning of each test mentions the following:

1. Use paper and pencil when necessary;
2. Answer multiple-choice questions with a letter (for Intermediate I and II);
3. Use the return key after a response (for Intermediate I and II);
4. Use the return key to skip a problem;
5. Use the rubout key to erase;
6. Be aware of the time limit for the entire test.

These six points and the printed instructions were explained and amplified by the CAI proctors before students took the MSAT. The instructions printed for Primary II and for the two intermediate forms of the MSAT are displayed in Figure 1. As in the SAT, students were given a

---

Insert Figure 1 about here

---

nongraded sample problem before beginning the MSAT. Students were not given any results at the end of a test session. A coded number containing the score was printed at the end of the session for use by teachers.

Choice of level of test. The level of the MSAT administered to a student is made with reference to his average GP in the mathematics strands curriculum. The choice algorithm is presented in Table 4.

Insert Table 4 about here

The MSAT was administered in January, 1972, at the beginning of the experimental period and again in May immediately after the experiment ended.

DATA ANALYSIS AND RESULTS

Analysis of Variance

As mentioned earlier, treatment groups 1, 2, 3, 4, and 5 were assigned 10, 30, 70, 100, and 130 sessions, respectively, and 77 students were randomly assigned to each of the five treatment groups. Complete data were obtained for 60 students in group 1, 62 students in group 2, 60 students in group 3, 60 students in group 4, and

Instructions for MSAT Primary II:

```
MODIFIED S. A. T. STRAND

THIS IS A SPECIAL TEST.  YOU HAVE 33 MINUTES TO WORK
ON IT.  YOU MAY USE PENCIL AND PAPER.

TYPE THE RETURN KEY TO GET THE NEXT PROBLEM.

HERE IS A SAMPLE PROBLEM.

4 + 9 = _13_
```

Instructions for MSAT Intermediate I and Intermediate II:

```
MODIFIED S. A. T. STRAND

THIS IS A SPECIAL TEST.  YOU HAVE 38 MINUTES TO WORK
ON IT.  USE PENCIL AND PAPER TO WORK THE PROBLEMS.

ALL QUESTIONS ARE MULTIPLE CHOICE.  TYPE A LETTER AND
THEN THE RETURN KEY.  USE THE RUBOUT KEY TO ERASE.  USE
THE RETURN KEY TO SKIP A PROBLEM.

HERE IS A SAMPLE PROBLEM.

4 + 9 =

A)  12
B)  5
C)  13
D)  49
E)  NOT GIVEN
_C
```

Fig. 1. Instructions printed for the Primary II and for the Intermediate forms of the MSAT.

TABLE 4

Algorithm for Choosing an MSAT Test Level

Test Level	Range of average GP on date of administration
Primary II	2.0 - 3.5
Intermediate I	3.6 - 4.8
Intermediate II	4.9 - 6.5

70 students in group 5. This information is summarized in Table 5, which also gives the means and standard deviations for number of

----
Insert Table 5 about here
----

sessions actually taken within each of the five treatment groups. The number of sessions taken fell well short of the number assigned in groups 3, 4, and 5, primarily because of difficulties in scheduling extra CAI sessions in the schools. However, the groups remained sufficiently distinct to warrant proceeding with analysis of variance.

Analyses of variance were performed taking MSAT scores and average GP of the mathematics strands as dependent measures. In order to make comparisons across all three MSAT battery scores, we used SAT scales to convert MSAT raw scores to GP scores. Analyses of variance were performed on pretreatment measures as well as post-treatment measures to check for any bias in the assignment of students to treatment groups.

The pretreatment analyses of variance for the strands average GP and MSAT scores are presented in Tables 6 and 7. Both F ratios are

----
Insert Tables 6 and 7 about here
----

small. In addition, the correlation between pretreatment, strands average GP and the number of sessions taken was .013, and the correlation between MSAT pretest GP and number of sessions taken was -.022. It is reasonable to conclude that random assignment of students to treatment groups was essentially achieved.

TABLE 5

Assigned and Obtained Number of Students and Sessions for

the Five Treatment Groups in the Experiment

Treatment group	Number of students assigned	Number of students obtained	Number of sessions assigned	Average number of sessions obtained	Standard deviation of sessions obtained
1	77	60	10	11.85	1.90
2	77	62	30	34.95	8.88
3	77	60	70	53.30	17.92
4	77	60	100	65.98	27.38
5	77	70	130	75.84	29.15

TABLE 6

Analysis of Variance for Pretreatment, Mathematics Strands

Average GP for the Five Treatment Groups

	Treatment group				
	1	2	3	4	5
Sample size	60	62	60	60	70
Mean	4.18	3.88	3.96	3.94	4.12
Standard deviation	.87	.87	.77	.84	.91

Analysis of variance

	Sum of squares	df	Mean square	F ratio
Between groups	4.1113	4	1.0278	1.408*
Within groups	224.1574	307	.7302	
Total	228.2687	311		

*Nonsignificant

TABLE 7

Analysis of Variance for Pretreatment, MSAT GP Scores

for the Five Treatment Groups

Treatment group

	1	2	3	4	5
Sample size	60	62	60	60	70
Mean	4.79	4.23	4.41	4.61	4.52
Standard deviation	1.50	1.31	1.21	1.72	1.64

Analysis of variance

	Sum of squares	df	Mean square	F ratio
Between groups	10.5828	4	2.6457	1.185*
Within groups	685.6444	307	2.2334	
Total	696.2272	311		

*Nonsignificant

The posttreatment analyses of variance for the strands average GP and MSAT scores are presented in Tables 8 and 9. The F ratio for the

---

Insert Tables 8 and 9 about here

---

strands GP scores is significant ($p < .01$, $df = 4/307$), and we note that the average GP improvement for the 10-sessions group 1 is only .15 compared with .96 for group 5. The F ratio for the MSAT scores is non-significant, but the average GP improvement for the 10-sessions group 1 is .42 compared with .76 for group 5.

Test of the Five Models

Parameters for the five models were generated twice, once using mathematics strands average GP as pretreatment and posttreatment achievement measures and once using MSAT GP scores. Models and parameters using strands average GP as the achievement measures are presented in Table 10.

Insert Table 10 about here

The linear model with interaction, Model II, accounts for more of the variance in the dependent variable (posttreatment average GP) than does any of the other models, but despite the inclusion of a term for the interaction of number of sessions with pretreatment GP, it represents only a slight improvement over Model I, the simple linear model. Assuming $N_i = 120$ or slightly less than one session per day for a school year and taking $a_2 = .0123$ from Model I, we can project $T_{i2} - T_{i1} = 1.48$. That is to say, if a student from this population takes about one strands session per day for an entire school year, we can expect his strands average

TABLE 8

Analysis of Variance for Posttreatment, Mathematics Strands

Average GP for the Five Treatment Groups

	Treatment group				
	1	2	3	4	5
Sample size	60	62	60	60	70
Mean	4.33	4.32	4.60	4.85	5.08
Standard deviation	.89	.84	.76	.94	.96

Analysis of variance

	Sum of squares	df	Mean square	F ratio
Between groups	28.3792	4	7.0948	9.088*
Within groups	239.6608	307	.7807	
Total	268.0400	311		

*Significant, $p < .01$.

TABLE 9

Analysis of Variance for Posttreatment, MSAT GP Scores for the Five Treatment Groups

	Treatment group				
	1	2	3	4	5
Sample size	60	62	60	60	70
Mean	5.21	4.82	4.89	5.29	5.28
Standard deviation	1.50	1.36	1.39	1.71	1.44

Analysis of variance

	Sum of squares	df	Mean square	F ratio
Between groups	12.9009	4	3.2252	1.464*
Within groups	676.4771	307	2.2035	
Total	689.3780	311		

*Nonsignificant

TABLE 10

Parameters Generated for the Five Models Using Mathematics Strands Average GP as Pretreatment and Posttreatment Measures

<u>Model I</u>: $E(T_{i2}) = a_0 + a_1 T_{i1} + a_2 N_i$.

Parameters: $a_0 = .305$, $a_1 = .930$, $a_2 = .012$.

Multiple correlation = .954.

<u>Model II</u>: $E(T_{i2}) = a_0 + a_1 T_{i1} + a_2 N_i + a_3 T_{i1} N_i$.

Parameters: $a_0 = .027$, $a_1 = .998$, $a_2 = .018$, $a_3 = -.001$.

Multiple correlation = .955.

<u>Model III</u>: $E(\ell n\ T_{i2}) = a_0 + a_1\ \ell n\ T_{i1} + a_2\ \ell n\ N_i$.

Parameters: $a_0 = .044$, $a_1 = .817$, $a_2 = .010$.

Multiple correlation = .941.

<u>Model IV</u>: $E(T_{i2}) = a_0 + a_1 T_{i1} + a_2\ \ell n\ N_i + a_3(\ell n\ N_i)^2 + a_4(\ell n\ N_i)^3$.

Parameters: $a_0 = .698$, $a_1 = .928$, $a_2 = .036$, $a_3 = -.126$.

Multiple correlation = .953.

<u>Model V</u>: $E(\ell n\ T_{i2}) = a_0 + a_1 N_i T_{i1}$.

Parameters: $a_0 = .363$, $a_1 = .0001$.

Multiple correlation = .632.

Note.--T_{i1} = pretreatment strands GP for student i,

T_{i2} = posttreatment strands GP for student i,

N_i = number of sessions taken by student i.

GP to increase by about a year and a half. Data presented later show that strands average GP underestimated both GP measured by paper and pencil administrations of the SAT and GP measured by the MSAT. This improvement of 1.48 can be compared with an expected GP increase over a school year of .3 to .4 in the SAT computation subtest for hearing-impaired students receiving ordinary instruction (Gentile & DiFrancesca, 1969).

Models and parameters using MSAT GP as pretreatment and posttreatment measures are presented in Table 11. The multiplicative model from econometrics

---

Insert Table 11 about here

---

that assumes weighted interaction of number of sessions with pretreatment GP, Model III, accounts for more of the variance in the posttreatment measure than does any of the other models, but, as with the strands average GP, it represents only a slight improvement over Model I, the simple linear model. Again, assuming $N_i = 120$ and taking $a_2 = .0084$ from Model I, we can project $T_{i2} - T_{i1} = 1.01$. That is to say, if a student from this population takes about one strands session per day for a school year of 120 net days, we can expect his MSAT GP to increase by about one year. Roughly, we can expect an increase of .1 in MSAT GP for every 12 sessions taken.

Overall, the models using MSAT GP as the achievement measure account for about 25 percent less variance in the posttreatment results than do those using strands GP as the achievement measure. In the case of Model I, this decrease might be due to less contribution from the pretreatment scores in accounting for posttreatment scores, less contribution from number of sessions, or both. A comparison of the a_1 parameter for

TABLE 11

Parameters Generated for the Five Models Using MSAT GP Scores as Pretreatment and Posttreatment Measures

---

<u>Model I</u>: $E(T_{i2}) = a_0 + a_1T_{i1} + a_2N_i$.

Parameters: $a_0 = 1.116$, $a_1 = .793$, $a_2 = .008$.

Multiple correlation = .811.

<u>Model II</u>: $E(T_{i2}) = a_0 + a_1T_{i1} + a_2N_i + a_3T_{i1}N_i$.

Parameters: $a_0 = .939$, $a_1 = .831$, $a_2 = .012$, $a_3 = -.001$.*

Multiple correlation = .812.

*The F ratio for deletion of a_3 is nonsignificant ($p < .005$).

<u>Model III</u>: $E(\ln T_{i2}) = a_0 + a_1 \ln T_{i1} + a_2 \ln N_i$.

Parameters: $a_0 = .101$, $a_1 = .711$, $a_2 = .006$.

Multiple correlation = .818.

<u>Model IV</u>: $E(T_{i2}) = a_0 + a_1T_{i1} + a_2 \ln N_i + a_3(\ln N_i)^2 + a_4(\ln N_i)^3$.

Parameters: $a_0 = 2.405$, $a_1 = .784$, $a_2 = .000$, $a_3 = -.330$, $a_4 = .069$.

Multiple correlation = .816.

<u>Model V</u>: $E(\ln T_{i2}) = a_0 + a_1N_iT_{i1}$.

Parameters: $a_0 = .370$, $a_1 = .0001$.

Multiple correlation = .477.

Note.--T_{i1} = pretreatment MSAT GP for student i,

T_{i2} = posttreatment MSAT GP for student i,

N_i = number of sessions taken by student i.

Model I, using strands GP ($a_1 = .930$), with Model I, using MSAT GP ($a_1 = .793$), and a comparison of the a_2 parameter for Model I, using strands GP ($a_2 = .0123$), with Model I, using MSAT GP ($a_2 = .0084$), imply that the pretreatment measure and the number of sessions both contribute less when Model I uses MSAT GP as the achievement measure than it does using strands average GP, a result that is not at all surprising.

Because of the interest in the contribution of number of sessions taken to the GP gain measured by the MSAT, we computed the 95 percent confidence level for coefficient a_2 for Model I (Table 11). The result is $a_2 = .0084 \pm .0032$. Consequently, the change in GP with 120 sessions should be between .624 and 1.392, admittedly a large interval. Assuming that the effect of 120 sessions is somewhat sublinear, we should conservatively expect an increase in GP of about .5 that would be due to the 120 strands CAI sessions. This conservatively estimated GP gain is still superior to the results for ordinary instruction of hearing-impaired students already cited. As is evident from Table 10, a larger gain in GP measured by the strands curriculum itself would be anticipated.

Comparison of SAT, MSAT, and Strands GP

In order to generalize the MSAT results, comparisons of the achievement measures used in this study with each other and with standardized tests are needed. Because neither the MSAT nor GP measured by the strands curriculum is a common measure, it was decided to estimate the concurrent validity of the MSAT GP and strands average GP by comparing them with each other and with paper and pencil administrations of the SAT.

Sixty students were drawn at random from among participating students in three of the residential schools. Selection of the students was stratified so that 4 were chosen for each of the 15 cells arising from the three MSAT forms (Primary II, Intermediate I, Intermediate II) and five treatment groups. Two of the 4 students were chosen at random and assigned to group I; the remaining 2 were assigned to group II. There were then 30 students (2 from each form-by-treatment cell times 15 cells) assigned to group I and 30 assigned to group II. Group I received paper and pencil administration of the SAT Arithmetic Computation Subtest (SAT-COMP), form W, before receiving the pretreatment MSAT, and group II received the SAT-COMP after receiving the pretreatment MSAT. The roles of groups I and II were reversed for the posttreatment measure. Group II received the SAT-COMP before the posttreatment MSAT, and group I received the SAT-COMP after the posttreatment MSAT. Pretreatment and posttreatment strands GP scores were also recorded.

Complete data were obtained for 44 of these students. The loss of 16 was solely due to such random factors as student illness, change of schools, and administrative errors. Means and standard deviations obtained by the 44 students for pretreatment and posttreatment SAT GP, MSAT GP, and strands GP are displayed in Table 12. It should be noted

---

Insert Table 12 about here

---

from Table 12 that the SAT consistently gave the highest estimate of GP for this group of students, the MSAT consistently gave the second highest GP estimate, and the strands GP consistently gave the lowest GP estimate.

TABLE 12

Means and Standard Deviations Obtained by 44 Randomly Chosen Students for Pretreatment and Posttreatment SAT GP, MSAT GP, and Mathematics Strands Average GP

	SAT GP pretreatment	SAT GP posttreatment	MSAT GP pretreatment	MSAT GP posttreatment	Strands GP pretreatment	Strands GP posttreatment
Mean	5.01	5.69	4.88	5.24	4.27	4.84
Standard deviation	1.13	1.40	1.53	1.46	.91	.96

Evidently, both the MSAT and the strands average GP measures underestimated GP measured by paper and pencil administration of the SAT.

A matrix of simple correlations for the GP scores obtained by the 44 students on pretreatment and posttreatment SAT, MSAT, and mathematics strands is given in Table 13. These correlations are fairly large, but

---

Insert Table 13 about here

---

they are not sufficiently large to identify SAT GP, MSAT GP, and strands GP as parallel measures. According to Lord and Novick (1968), two distinct measurements X_{gs} and X_{hs} are parallel if for every subject s in the population, $\pi_{gs} = \pi_{hs}$ and $\sigma(E_{gs}) = \sigma(E_{hs})$, where π indicates measurements with the same true scores but possibly different error variances. More intuitively, two measurements are parallel if their expectations are equivalent and their observed score variances are equal. This is not true of the three GP measures.

Difference Scores

Table 14 displays a matrix of simple correlation coefficients obtained from number of sessions, pretreatment and posttreatment MSAT GP, and

---

Insert Table 14 about here

---

pretreatment and posttreatment strands GP for the 312 students who participated in the experiment. Table 14 also includes correlation coefficients for the difference scores, MSAT Δ GP (posttreatment minus pretreatment MSAT GP) and strands Δ GP (posttreatment minus pretreatment strands average GP).

TABLE 13

Matrix of Simple Correlation Coefficients for GP Scores Obtained by 44 Randomly Chosen Students on Pretreatment and Posttreatment SAT, MSAT, and Mathematics Strands Observations

	SAT pre-treatment	SAT post-treatment	MSAT pre-treatment	MSAT post-treatment	Strands pre-treatment	Strands post-treatment
SAT pre-treatment	1.000	.761	.787	.689	.685	.733
SAT post-treatment		1.000	.773	.827	.758	.794
MSAT pre-treatment			1.000	.833	.800	.796
MSAT post-treatment				1.000	.764	.807
Strands pre-treatment					1.000	.860
Strands post-treatment						1.000

TABLE 14

Matrix of Simple Correlation Coefficients for Number of Sessions, Pretreatment MSAT GP, Posttreatment MSAT GP, MSAT △ GP, Pretreatment Strands GP, Posttreatment Strands GP, and Strands △ GP for the 312 Students Participating in the Experiment

	Sessions	MSAT pre-treatment	MSAT post-treatment	MSAT △ GP	Strands pre-treatment	Strands post-treatment	Strands △ GP
Sessions	1.000	-.022	.155	.273	.013	.416	.796
MSAT pre-treatment		1.000	.793	--	.748	.711	.041
MSAT post-treatment			1.000	--	.689	.762	.247
MSAT △ GP				1.000	-.097	.072	.320
Strands pre-treatment					1.000	.863	--
Strands post-treatment						1.000	--
Strands △ GP							1.000

Note.--The difference score, MSAT △ GP, is posttreatment MSAT GP minus pretreatment MSAT GP. The difference score, Strands △ GP, is posttreatment strands GP minus pretreatment strands GP.

The data reported in Tables 13 and 14 allow comparisons of the 44-student sample with the full sample of 312. The correlation of the strands pretreatment GP and the strands posttreatment GP is .860 in the 44-student sample compared with .863 in the 312-student sample. The largest difference in correlation was obtained for the pretreatment MSAT GP with the posttreatment strands GP--the 44-student sample correlation being .711 compared with .796 for the 312-student sample. As might be expected, the correlations for the 312-student sample are in every case higher than their counterparts in the 44-student sample. Given these small differences in correlation coefficients, it seems reasonable to conclude that, with respect to the measures taken, the 44-student sample is representative of the full 312-student population.

The anticipated increase in correlation occurs when difference scores replace posttreatment scores as the dependent measures in regressions using number of sessions as the independent variables. The correlation for sessions and MSAT posttreatment GP is only .155 compared with a correlation of .273 for sessions and MSAT Δ GP. More striking, but also expected, is the correlation for sessions and strands posttreatment GP of .416 compared with .796 for sessions and strands Δ GP.

Means and standard deviations for GP change (posttreatment GP minus pretreatment GP) obtained by the five treatment groups are displayed in Table 15. With the exception of average GP change for group 3, a fairly

---

Insert Table 15 about here

---

steady increase in GP change with increasing number of sessions is evident in the data. It should be noted that the GP change for group 5 whose

TABLE 15

Sample Size, Means, and Standard Deviations of GP Change (Posttreatment GP Minus Pretreatment GP) Measured by Strands Average GP and MSAT for the Five Treatment Groups

		Treatment group				
		1	2	3	4	5
Sample size		60	62	60	60	70
Strands Δ GP	Mean	.15	.45	.64	.91	.96
	Standard deviation	.01	.03	.03	.05	.05
MSAT Δ GP	Mean	.42	.58	.48	.68	.76
	Standard deviation	.10	.09	.08	.10	.11

members averaged 75.84 sessions is .96 for strands average GP and .76 for MSAT GP. Both of these measures underestimated GP measured by paper and pencil administration of the SAT to the 44-student sample. Using the smaller measure, the group 5 students achieved an increase in mathematics computation GP of .76 during the experimental period of approximately five months. This improvement is about double the GP gain indicated by Gentile and DiFrancesca (1969) for hearing-impaired students after a full school year of traditional classroom instruction.

SUMMARY AND CONCLUSIONS

From the analyses of data given above, we conclude that the mathematics strands CAI curriculum can lead to substantial increases in mathematics computation GP when used by hearing-impaired students. The increases are sufficient to bring the students to GP gains expected of normal-hearing students. Moreover, these gains can be achieved by students working intensely for only a few minutes a day in a supplementary drill-and-practice program. The actual time spent at a computer terminal by each student ranged from 6 to 10 minutes for each session.

In addition, a simple linear model of student achievement gives a good account of the posttreatment distribution of GP measured either by the MSAT or by the strands GP. The investigation of other models, including models with interaction terms, did not lead to any substantial improvement in accounting for posttreatment GP variance. The results of the analysis, including the application of the linear model, indicate that greater numbers of CAI sessions are beneficial for all students, across all levels of pretreatment achievement. Further investigation using demographic and other psychological variables would be desirable in providing a more detailed analysis of the extent to which posttreatment distribution of GP is primarily affected only by pretreatment distribution of GP and amount of time at computer terminals.

The inequality-averting properties of the mathematics strands curriculum used as compensatory education for disadvantaged hearing students has already been noted by Suppes and Morningstar (1970b) and by Jamison, Fletcher, Suppes, and Atkinson (1973). These properties seem to obtain to about the same degree for the population of hearing-impaired students used in this study.

REFERENCES

Clark, C. H., Beatty, L. S., Payne, J. N., & Spooner, G. A. Elementary mathematics, grades 1-6. New York: Harcourt, Brace & World, 1966.

Cronbach, L. J., & Furby, L. How we measure "change"--or should we? Psychological Bulletin, 1970, 74, 68-80.

Eicholz, R. E., & O'Daffer, P. G. Elementary school mathematics, grades 1-6. Palo Alto, Calif.: Addison-Wesley, 1968.

Gentile, A., & DiFrancesca, S. Academic achievement test performance of hearing impaired students. Series D, Number 1. Washington, D. C.: Office of Demographic Studies, Gallaudet College, 1969.

Jamison, D., Fletcher, J. D., Suppes, P., & Atkinson, R. C. Cost and performance of computer-assisted instruction for education of disadvantaged children. In J. Fromkin and R. Radner (Eds.), Education as an industry. New York: National Bureau of Economic Research, Columbia University Press, 1973 (in press).

Lord, F. M. Elementary models for measuring change. In C. W. Harris (Ed.), Problems in measuring change. Madison, Wisc.: University of Wisconsin Press, 1963.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Suppes, P. Sets and numbers, grades 1-6. New York: Singer, 1966.

Suppes, P., Goldberg, A., Kanz, G., Searle, B., & Stauffer, C. Teacher's handbook for CAI courses. Technical Report No. 178. Stanford: Institute for Mathematical Studies in the Social Sciences, Stanford University, 1971.

Suppes, P., Jerman, M., & Brian, D. _Computer-assisted instruction: The 1965-66 Stanford arithmetic program_. New York: Academic Press, 1968.

Suppes, P., & Morningstar, M. Computer-assisted instruction. _Science_, 1969, _166_, 343-350.

Suppes, P., & Morningstar, M. Four programs in computer-assisted instruction. In W. H. Holtzman (Ed.), _Computer-assisted instruction, testing, and guidance_. New York: Harper & Row, 1970. (a)

Suppes, P., & Morningstar, M. Technological innovations: Computer-assisted instruction and compensatory education. In F. Korten, S. Cook, and J. Lacey (Eds.), _Psychology and the problems of society_. Washington, D. C.: American Psychological Association, 1970. (b)

Suppes, P., & Morningstar, M. _Computer-assisted instruction at Stanford, 1966-68: Data, models, and evaluation of the arithmetic programs_. New York: Academic Press, 1972.

Suppes, P., Searle, B., & Lorton, P., Jr. _The strands arithmetic CAI program_. Stanford, Calif.: Institute for Mathematical Studies in the Social Sciences, 1973, in press.