AN EVALUATION OF THE STANFORD CAI PROGRAM IN INITIAL READING (GRADES K THROUGH 3)

by

J. D. Fletcher and R. C. Atkinson

TECHNICAL REPORT NO. 168

March 12, 1971

PSYCHOLOGY SERIES

Support for this research was sponsored by the National Science Foundation, Grant No. NSF-GJ-443X.

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

AN EVALUATION OF THE STANFORD CAI PROGRAM IN INITIAL READING (GRADES K THROUGH 3)

J. D. Fletcher and R. C. Atkinson

Stanford University

Stanford, California 94305

Computer-Assisted Instruction (CAI) in initial reading, grades K through 3, has been under development at Stanford University for the last seven years. Initially, the aim of this effort was to implement a complete CAI reading curriculum which would depend only minimally on ordinary classroom activity. These early efforts were successful (Atkinson, 1968), but it became clear that the cost of such a program was prohibitive. Economically and pedagogically, some aspects of initial reading seemed better left to the classroom teacher. The aim of the reading project during the last three years has been to design and implement a low cost CAI curriculum that would act as a supplement to normal classroom instruction. A student terminal in the current program consists only of a "Model 33" teletypewriter equipped with audio. There is no graphic or photographic capability at the student terminal, and the character set of the teletypewriter includes only upper case letters. On the other hand, audio for the Stanford CAI system is quite flexible in that it is digitized and stored on magnetic disks. This system provides for rapid (30 milliseconds) random access to any one of 6,000 recorded sounds. The Stanford CAI system is more fully documented in Atkinson, Fletcher, Chetin, and Stauffer (1971).

Learning to read may be divided into two basic tasks variously referred to as decoding and communication. For present purposes, decoding may be defined

as the rapid, if not automatic, association of phonemes or phoneme groups with their respective graphic representations; communication may be denoted as reading for meaning, aesthetic enjoyment, emphasis, and the like. Of the two, decoding skills are more easily defined and, consequently, are more amenable to CAI than are communication skills. The major emphasis of the Stanford CAI program is on decoding, with the view that other aspects of reading instruction can be left to the classroom teacher. Because of this emphasis, the CAI program can supplement classroom instruction using any basic vocabulary or textbook series.

Decoding is, however, not the sole concern of the program. Instruction is divided into seven content areas or strands. Strand 0, the readiness strand, provides practice with the manual skills required for interaction with the CAI program and instruction on a series of fairly standard "reading readiness" tasks. Strand I, the letter strand, provides practice in copying, recognition, and recall of the letters of the alphabet. The initial pass through the alphabet presents letters singly and in maximally contrasting groups, for example (RTO); later passes through the alphabet present letters in minimally contrasting groups, for example (MNW). Strand II, the word strand, provides for the development of a sight word vocabulary. Seven K through 3 reading vocabulary lists were analyzed in developing this strand. Of the words used in Strands II through IV, those that do not include regular grapheme-phoneme correspondences are presented only in this strand. Strand III, the spelling strand, provides for recognition and recall of orthographically regular monosyllabic words arranged in groups which emphasize a single spelling pattern, for example (ran, fan, man) or (fat, fan, fad). Strand IV, the phonics strand, provides for direct practice in copying

and recognition of the spelling patterns themselves as well as the "construction" of monosyllabic words from given consonant clusters and spelling patterns. Strand V, the comprehension categories strand, attempts to provide practice with the meaning of words by emphasizing the semantic categories of words. Exercises in this strand ask the student to select the word of those displayed that is an animal or that is a color, etc. Strand VI, the comprehension sentence strand, provides practice in reading sentences by requiring the student to select a word to fill an empty "slot" in the sentence. On any given day, a student's lesson may involve exercises drawn from one to five different strands. A more complete description of the program as well as the rationale underlying it is presented in Atkinson, Fletcher, Chetin, and Stauffer (1971).

The CAI program is highly individualized so that each student is exposed to a sequence of instructional materials that maximizes his progress. The basic approach has been to develop mathematical models for the acquisition of various decoding skills and to use these models to specify optimal procedures. These procedures require a rapidly accessible sufficient history of response information to be maintained for each student. As the student progresses through the curriculum, his history is continually updated and interrogated in order to specify the curriculum items to be presented next. A discussion of optimization procedures developed for the CAI reading program can be found in Atkinson and Paulson (1971).

The first tryout of the program occurred in the 1968-69 school year with students in the grades K, 1, 2 and 3. As expected, many problems of curriculum design and system operation were identified and had to be corrected during this period. By the summer of 1969, however, the system and curriculum had stabilized

to a satisfactory level of operation, and an evaluation of the program was undertaken during the 1969-70 school year. The purpose of this report is to briefly summarize some of the more important findings obtained from the evaluation.

Method

The problems of evaluating a new curriculum are many, and it is difficult if not impossible to deal with all of them. The design adopted for this evaluation has its faults, but within the economic and administrative constraints of the situation it appeared to be the most reasonable choice. A matched-pairs design was used in which compensation for differences between experimental and control groups is achieved by matching on the basis of pretest scores.

Although over a hundred students were run for varying periods of time on the CAI reading curriculum, the evaluation was limited to a group of 50 matched pairs. Prior to receiving any exposure to CAI, 25 pairs of first-grade boys and 25 pairs of first-grade girls were matched on the basis of the Metropolitan Readiness Test (MET). The MET was administered in early November to groups of 10 or less pupils by trained test personnel. The Numbers and Draw-A-Man subtests of the MET were not administered. Matching was achieved so that the MET scores for a matched pair of students were no more than two points apart. Moreover, in matching the students an effort was made to insure that both members of a pair had classroom teachers of roughly equivalent ability. The mean MET score for the boys participating in the evaluation was 56.6 and the mean for the girls was 55.1.

The experimental member of each matched pair of students received eight to ten minutes of CAI instruction per school day roughly from the first week in

January until the second week in June. The control member of each pair received no CAI instruction. Except for the eight to ten minute CAI period there is no reason to believe that the activities during the school day were any different for the experimental and control subjects.

Three post-tests were administered to all subjects in late May and early June, 1970. Four subtests of the Stanford Achievement Test (SAT), Primary I, Form X, were used. These subtests were: word reading (S/WR), paragraph meaning (S/PM), vocabulary (S/VOC), and word study (S/WS). Second, the California Cooperative Primary Reading Test (COOP), Form 12A (grade 1, spring) was administered. Only the total raw scores were used from this test. Both the SAT and the COOP were administered to classroom groups by teachers under the supervision of district testing personnel. Finally, a test (DF) developed at Stanford and tailored to the goals of the CAI reading curriculum was administered individually to all subjects. The DF items fell into eight groups yielding the following eight subtests: upper case letters (D/LU), lower case letters (D/LL), upper case words (D/WU), lower case words (D/WL), spelling patterns (D/SP), monosyllabic words comprising these spelling patterns (D/SW), and nonsense monosyllables comprising these spelling patterns (D/SN). The words for the D/WU and D/WL subtests were chosen at random from first-grade vocabulary lists. The spelling patterns for the D/SP, D/SW, and D/SN subtests had all been taught in the CAI curriculum, but none of the words or nonsense syllables in the D/SW and D/SN subtests had been taught. In administering the test, an item printed in primary type on a 3 x 5 index card is shown to a subject who then has 10 seconds to read the item aloud.

The CAI subjects were expected to score higher on all subtests of DF than were the non-CAI subjects, since each subtest represented a specific goal of the curriculum. However, because the Model 33 teletypewriters did not provide for lower case printing, the CAI subjects' gains were expected to be greater for the two upper case subtests (D/LU and D/WU) than for the related lower case subtests (D/LL and D/WL). The three spelling pattern subtests were all presented in upper case, so the question of upper versus lower case did not arise for these presentations.

Some predictions can also be made for the SAT subtests. The greatest difference between CAI and non-CAI subjects was predicted for the S/WS since this subtest deals directly with sounds and spelling patterns (Kelley, Madden, Gardner, & Rudman, 1964). Very little difference between CAI and non-CAI subjects was expected for the S/PM subtests. Teletypewriter presentation is not suited to large amounts of text output in an instructional situation, and there were no paragraph exercises in the CAI curriculum. Some differences, but not of significant magnitude, were predicted for the S/WR and S/VOC subtests.

Earlier results had shown that boys in CAI initial reading do about as well as girls despite the almost universally expected superior performance of girls in conventional initial reading (Atkinson, 1968). For this reason male and female matched pairs were kept separate to see if this result would hold for the current evaluation.

Results and discussion

During the course of the school year, an equal number of pairs was lost from the female and male groups; complete data were obtained for 22 pairs of boys and 22 pairs of girls.

Means, standard deviations, and t-values for differences in SAT, COOP, and DF total scores are presented in Table 1 for the matched pairs of boys and

---
Insert Table 1 about here
---

the matched pairs of girls. In this table, standard deviations are displayed in parentheses, and t-values in brackets. The t-values calculated are for non-independent samples, and those that are significant ($p < .01$, one-tailed) are starred.

The results of these analyses are heartening. Of the six post-test comparisons, only one (COOP for matched pairs of girls) failed to indicate a significant difference in favor of the CAI reading subjects. These differences are also important from the standpoint of improvement in estimated grade placement. Table 2 displays the mean grade placement of the four groups on the SAT

Insert Table 2 about here
---

and COOP. The differences between CAI and non-CAI groups in estimated grade placement range from .4 to .7 school years.

Means, standard deviations, and t-values for the differences on the four SAT subtests are presented in Table 3 for male and female matched pairs. As

---
Insert Table 3 about here
---

in Table 1, standard deviations are displayed in parentheses, and t-values in brackets; t-values that are significant ($p < .01$, one-tailed) are starred.

Table 1. Means, standard deviations (in parentheses), and t-values (in brackets) for the Stanford Achievement Test (SAT), the California Cooperative Primary Test (COOP), and the CAI Reading Project Test (DF)

		SAT		COOP		DF	
	CAI	109.7		33.2		64.9	
		(24.1)		(8.6)		(7.0)	
Boys			[t=3.60*]		[t=4.70*]		[t=7.01*]
	non-CAI	90.2		23.4		53.0	
		(19.5)		(8.9)		(10.4)	
	CAI	115.7		33.7		64.1	
		(26.2)		(10.4)		(7.6)	
Girls			[t=2.55*]		[t=1.65]		[t=3.10*]
	non-CAI	96.5		28.9		56.6	
		(30.5)		(10.8)		(13.5)	

*$p < .01$, df = 21

Table 2. Average grade placement on the Stanford Achievement Test (SAT) and the California Cooperative Primary Test (COOP)

		SAT	COOP
Boys	CAI	2.2	2.5
	non-CAI	1.8	1.8
Girls	CAI	2.4	2.6
	non-CAI	2.0	2.2

Table 3. Means, standard deviations (in parentheses), and t-values (in brackets) for the word reading (S/WR), paragraph meaning (S/PM), vocabulary (S/VOC), and word study (S/WS) subtests of the Stanford Achievement Test

		S/WR	S/PM	S/VOC	S/WS
Boys	CAI	25.8	21.8	21.8	40.3
		(5.6)	(8.2)	(5.8)	(8.4)
		[t=4.42*]	[t=3.33*]	[t=-.25]	[t=3.33*]
	non-CAI	18.9	15.1	22.1	34.0
		(5.4)	(6.7)	(6.4)	(5.7)
Girls	CAI	27.3	24.2	21.3	42.9
		(6.2)	(8.6)	(6.3)	(8.3)
		[t=3.03*]	[t=2.62*]	[t=.64]	[t=2.19]
	non-CAI	21.4	17.6	20.3	37.3
		(7.2)	(11.0)	(5.7)	(10.7)

*$p < .01$, df = 21

These SAT subtests reveal some interesting results. The S/WS differences are fairly large as expected, but, in the case of the boys, they fall slightly short of statistical significance ($p < .01$). Of the four SAT subtests, the S/WS was expected to reflect most clearly the goals of the CAI curriculum; yet greater differences between CAI and non-CAI groups were obtained for both the S/WR and S/PM subtests. Also notable is the lack of any real differences for the S/VOC. One explanation for this result is that the vocabulary subtest measures a pupil's vocabulary independent of his reading skill (Kelley, et al., 1964); since the CAI reading curriculum is primarily concerned with reading skill and only incidentally with vocabulary growth, there may have been no reason to expect a discernable effect of the CAI curriculum on the S/VOC. Most notable, however, are the S/PM results. In both the male and female groups, the CAI students performed significantly better on paragraph items than did the non-CAI students, despite the absence of paragraph items in the CAI program and the relative dearth of sentence items. These results for phonics-oriented programs are not unprecedented as Chall's (1967, pp. 106-107) survey shows. Nonetheless, for a program with so little emphasis on connected discourse, they are surprising.

Means, standard deviations, and t-values for five of the seven DF subtests are presented in Table 4 for male and female matched pairs. As in Tables 1 and 3,

---

Insert Table 4 about here

---

standard deviations are in parentheses, and t-values in brackets; significant t-values ($p < .01$, one-tailed) are starred. All subjects obtained perfect or nearly perfect scores on the DF letter subtests (D/LL and D/LU); the male and

Table 4. Means, standard deviations (in parentheses), and t-values (in brackets) for the upper case words (D/WU), lower case words (D/WL), spelling patterns (D/SP), spelling pattern words (D/SW), and spelling pattern nonsense syllables (D/SN) subtests of the CAI Reading Project Test

		D/WU	D/WL	D/SP	D/SW	D/SN
Boys	CAI	10.4	9.9	7.0	7.0	6.6
		(2.1)	(1.6)	(1.5)	(1.4)	(1.6)
		[t=5.20*]	[t=4.72*]	[t=3.62*]	[t=4.91*]	[t=5.80*]
	non-CAI	7.4	7.3	5.5	5.1	4.4
		(2.7)	(2.6)	(2.1)	(1.9)	(2.0)
Girls	CAI	10.5	10.2	7.4	6.3	6.5
		(1.7)	(1.9)	(1.0)	(1.8)	(2.0)
		[t=3.37*]	[t=2.63*]	[t=3.35*]	[t=1.44]	[t=2.56*]
	non-CAI	8.3	8.7	5.5	5.5	4.8
		(3.2)	(2.8)	(2.7)	(2.9)	(2.8)

*$p < .01$, df = 21

female CAI groups obtained slightly higher scores on the subtests than did their non-CAI counterparts, but the performance was so high and the variability so low that comparisons are not justified. These data suggest that a lesser emphasis on letter teaching in the CAI program is in order, and that the time devoted to letter teaching should be shifted to other strands of the curriculum.

The DF word subtests presented in Table 4 show the expected superior performance of the CAI groups on the upper case presentations (D/WU). However, the differences favoring the CAI groups for lower case presentations are also statistically significant though of a lesser magnitude than those for upper case presentations. Evidently, the lack of lower case letters in the CAI program is a handicap of only minor importance.

The DF spelling pattern subtests reflect goals which are at the heart of the CAI curriculum. The greatest effect of the curriculum was expected to be on the ability of students to recognize spelling patterns (D/SP), to pronounce orthographically regular but unfamiliar words comprising the spelling patterns (D/SW), and to pronounce orthographically regular nonsense syllables comprising the spelling patterns (D/SN). Five of the six obtained differences on these spelling pattern subtests are statistically significant.

It was expected that some measure of performance on the system would correlate fairly highly with the SAT, COOP, and DF total scores. These correlations are presented in Table 5. The measure of performance on the CAI curriculum used

---

Insert Table 5 about here

---

here is the total number of curriculum units brought to criterion by a student divided by the total amount of time accumulated on the system. It will be noted

Table 5. Correlations of the number of CAI reading items completed per unit time with the Stanford Achievement Test (SAT), the California Cooperative Primary Test (COOP), and the CAI Reading Project Test (DF)

	SAT	COOP	DF
Boys	.74	.68	.48
Girls	.84	.77	.49

are so substantial suggests that the CAI treatment administered over several years could well have dramatic results. Although we have no systematic data on students who have been on the program for several years, the size of the effects observed in this short-term study are in accord with our impressions of the improvements achieved by students with more extended experience.

Other analyses of these data have been run but will not be reported here. Our main purpose in this paper is to briefly report a few of the more important results from the evaluation, but not to offer any firm conclusions or interpretations. We recognize that some readers will not be happy with our multiple use of paired t-tests. They are presented not as definitive measures of statistical significance but rather as rough indexes of the influence of CAI on the various dependent measures. Also, we expect that there will be readers interested in analyzing aspects of these data not reported here. With this in mind, we have recorded our data in a format that can be readily sent to those who request it.

References

Atkinson, R. C. Computerized instruction and the learning process. American Psychologist, 1968, 23, 225-239.

Atkinson, R. C., Fletcher, J. D., Chetin, H. C., and Stauffer, C. M. Instruction in initial reading under computer control: The Stanford Project. Educational Technology Research Reports Series, Englewood Cliffs, N. J.: Educational Technology Publications, 1971.

Atkinson, R. C., and Paulson, J. A. An approach to the psychology of instruction. Psychological Bulletin, 1971 (in press).

Chall, J. Learning to read: The great debate. New York: McGraw-Hill, 1967.

Kelley, T. L., Madden, R., Gardner, E. G., and Rudman, H. C. Stanford achievement test: Directions for administering. New York: Harcourt Brace Jovanovich, 1964.