Volume Title: Education as an Industry

Volume Author/Editor: Joseph T. Froomkin, Dean T. Jamison and Roy Radner, eds.

Volume Publisher: NBER

Volume ISBN: 0-88410-476-1

Volume URL: http://www.nber.org/books/jami76-1

Publication Date: 1976

Chapter Title: Cost and Performance of Computer-Assisted Instruction for Education of Disadvantaged Children

Chapter Author: Dean T. Jamison, J. Dexter Fletcher, Patrick Suppes

Chapter URL: http://www.nber.org/chapters/c4492

Chapter pages in book: (p. 199 - 248)

PART TWO Compensatory Education

5

DEAN T. JAMISON
The World Bank

J. DEXTER FLETCHER
Navy Personnel Research and Development Center

PATRICK SUPPES
Stanford University

RICHARD C. ATKINSON
National Science Foundation

Cost and Performance of Computer-Assisted Instruction for Education of Disadvantaged Children

I. INTRODUCTION

This paper discusses the potential role of computer-assisted instruction (CAI) in providing compensatory education for disadvantaged children. All CAI involves, to one extent or another, the interaction of students

NOTE: This research was supported in part by National Science Foundation Grant NSF-GJ-443X3 to the Institute for Mathematical Studies in the Social Sciences, Stanford University. A. Kelley, S. Michelson, and D. Wiley provided helpful comments on an earlier draft of this paper.

with computers. Curriculum material is stored by a computer which is provided with decision procedures for presenting the material to individual students. Typically students work at terminals, usually teletypewriters, which are located at school sites and are connected by telephone lines to a central computer. Using time-sharing techniques, a single computer may simultaneously serve more than 1,000 students at diverse and remote locations. Advances in time-sharing techniques, coupled with reductions in hardware costs and increasing availability of tested curriculum material, are beginning to make CAI economically attractive as a source of compensatory education. Pedagogically, the value of CAI is established by its capacity for immediate evaluation of student responses and detailed individualization of treatment based on accurate and rapid retrieval of performance histories.

A number of institutions in the United States have computer-assisted programs under way in varying scales of complexity. Zinn (1970) and Lekan (1971) have provided overviews of these efforts. Stanford University's Institute for Mathematical Studies in the Social Sciences (IMSSS) has been engaged in such development efforts for over ten years and now operates one of the largest CAI centers in the country. This paper discusses the Institute's efforts to use CAI to provide compensatory education for disadvantaged students. Before turning to these efforts, however, it is worthwhile to place our work in the context of the large national effort in compensatory education that has been financed, primarily, by Title I of the Elementary and Secondary Education Act of 1965.

For a number of years the federal government has spent about one billion dollars annually to provide compensatory education for disadvantaged children in the United States. Unfortunately, much of the available evidence suggests that these federally funded Title I programs have met with little success. During the period 1966–68 Piccariello (1969) conducted a large-scale evaluation of Title I reading programs and, in more than two instances out of three, found no significant achievement differences between children in control groups and children in one of the Title I programs. Further, only slightly more than half of the significant differences obtained were in a positive direction. In his widely discussed paper on IQ and scholastic achievement, Jensen (1969) surveyed a large number of studies indicating a general failure of compensatory education.

Rather than studying the typical compensatory education program, Kiesling (1971) undertook a study of those compensatory education programs that had been most successful in the State of California. Kiesling concluded that there were a number of common elements in all these successful programs, and that one could learn from their success and replicate them. Thus, while compensatory education may have been, on the average, unsuccessful in the past, Kiesling concluded that there is no reason to repeat these failures. Success could be achieved by tailoring future compensatory programs around those that have previously proven themselves. Kiesling presented a number of paradigmatic compensatory programs for both arithmetic and reading and estimated their annual cost per student to be on the order of $200 to $300 per year in addition to the normal school allotment for that student.

A different interpretation from Kiesling's of the failure of compensatory education is that what goes on in schools has little effect on the achievement of students. This view received considerable support from Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld, and York (1966), and is consistent with the views of Jensen (1969). Coleman et al. concluded that factors within the schools seem to affect achievement much less than do factors outside the schools; these somewhat disheartening conclusions have been the source of vigorous debate since their initial publication. A number of recent view interpreting the data of the Coleman Report may be found in Mood (1970). The general drift of the papers in this book is that schooling is rather more important than one would conclude from the initial Coleman Report; nevertheless, there is an increasing consensus, since publication of the report, that input factors in the schooling process seem to have a good deal less effect on the outputs than had been thought previously (see Jamison, Suppes, and Wells, 1973).

Our own work, however, has led to more hopeful conclusions concerning the potential capability of the schools to affect scholastic performance. We have found strong and consistent achievement gains by disadvantaged students when they are given CAI over a reasonable fraction of a school year. Thus we are more inclined to accept Kiesling's general conclusions that compensatory education can work than the less hopeful interpretations of the Coleman Report. As Bowles and Levin (1968) pointed out: "The findings of the Report are particularly inappropriate for assessing the likely effects of radical changes in the level and compositions of resources devoted to schooling, because the range of variation in most school inputs in this sample is much more limited than the range of policy measures currently under discussion." Our evaluations of CAI provide detailed information about the output effects of a much broader variety of school inputs than the Coleman Report was able to consider.

This paper reports on the performance of three CAI programs that have performed well with underachieving children. Section II of the paper describes those programs—one in elementary arithmetic, one in initial reading, and one in computer programming for high school students. Section III reports on an evaluation of the performance of these programs. We considered two aspects of performance: achievement gain and the degree to which the program enabled disadvantaged students to close the gap between themselves and more advanced students. In order to examine this latter distributional effect, we relied in part on Gini coefficients derived from Lorenz curve representations of achievement data. We also examined the results in the light of several alternative mathematical formulations of "inequality-aversion." Section IV of the paper provides a discussion of costs. In particular, we examined the problem of making computer-assisted instruction available in rural as well as urban areas and attempted a realistic assessment of those costs. Our cost projections were for systems having about 1,000 student terminals; this number of terminals would allow 20,000 to 30,000 students to use the system per day. We computed not only dollar costs but also opportunity costs for using CAI in order to estimate the increase in student-to-teacher ratios that would be required if CAI were introduced under the constraint that expenditures per student should remain constant.

II. DESCRIPTION OF THREE PROGRAMS

A. Arithmetic

Development of elementary-school mathematics (grades 1 through 6) CAI was begun by the Institute in 1965. The intent of the program is to provide practice in arithmetic skills, especially computation, as an essential supplement to regular classroom instruction. Concepts presented by the CAI program are assumed to have been previously introduced to the students by their classroom teacher.

In the version of the mathematics curriculum discussed in this report, curriculum material for each of the six elementary-school grades was arranged sequentially in 20–27 concept blocks that corresponded in order and content to the mathematical concepts presented in several textbook series surveyed during the development of the curriculum. Each concept block consisted of a pretest, five drills divided into five levels of difficulty, and a posttest. The pre- and posttests were comprised of equal numbers of items drawn from each of the five difficulty levels in the drills. Each block contained approximately seven days of activity, one day each for the pre- and posttests and five days for the five drills. As part of each day's drill, a student also received review items drawn from previously completed concept blocks. Review material comprised about a third of a day's drill.

The level of difficulty for the first drill within a block was determined by a student's pretest performance for the block. The level of difficulty for each successive day's drill was determined by the student's performance during the preceding day. If a student's performance on a drill was 80 per cent or more correct, his next drill was one level of difficulty higher; if his performance on a drill was 60 per cent or less correct, his next drill was one level of difficulty lower.

The drill content, then, was the same for all students in a class, with only the difficulty levels varying from student to student. The content of the review material, however, was uniquely determined for each student on the basis of his total past-performance history. His response history was scanned to determine the previously completed concept block for which his posttest score was lowest, and review exercises were drawn from this block. Material from the review block was included in the first four drills for the current block, and a posttest for the review block was given during the fifth drill. The score on this review posttest replaced the previous posttest score for the review block and determined subsequent review material for the student.

Student terminals for the arithmetic curriculum were Model-33 teletypewriters without the random audio capability required for the reading program. These teletypewriters were located at school sites and were connected by telephone lines to the Institute's central computer facility at Stanford University. Students completed a concept block about every one and one-half weeks. This version of the arithmetic curriculum is described extensively in a number of publications including Suppes and Morningstar (1969) and Suppes, Jerman, and Brian (1968).

A more highly individualized mathematics strand program in arithmetic has been developed over the past several years and has replaced the program just described. Performance data in this paper are for the earlier program; a description of the more recent program was given by Suppes and Morningstar (1970).

B. Reading

CAI in initial reading (grades K–3) has been under development by IMSSS since 1965. The original intent of the reading program was to implement a complete CAI curriculum using cathode-ray tubes (CRT), light pen and typewriter input, slides, and random access audio. These efforts, described in Atkinson (1968), were successful but prohibitively expensive. Economically and pedagogically, some aspects of initial reading seemed better left to the classroom teacher. Subsequent efforts of the reading project were directed toward the development of a CAI reading curriculum that would supplement, but not replace, classroom reading instruction.

The current reading curriculum requires only the least expensive of teletypewriters and some form of randomly accessible audio. No graphic or photographic capabilities are needed and only uppercase letters are used. Despite these limitations, an early evaluation of the curriculum indicated that it is of significant value (Fletcher and Atkinson, 1972).

The version of the reading curriculum used in this report, more fully described by Atkinson and Fletcher (1972), emphasized phonics instruction. There were two primary reasons for this emphasis. First, it enabled the curriculum to be based on a relatively well-defined aspect of reading theory, making it more amenable to computer presentation. Second, the phonics emphasis on the regular grapheme-phoneme correspondences (or spelling patterns) which occur across all English orthography insured that the program appropriately supplemented classroom instruction using any initial reading vocabulary.

Instruction was divided into seven content areas or strands: O—machine readiness; I—letter identification; II—sight-word vocabulary; III—spelling patterns; IV—phonics; V—comprehension categories; and VI—comprehension sentences.

The term strand in the reading program defined a basic component skill of initial reading. Students in the reading program moved through each strand in a roughly linear fashion. Branching or progress within strands was criterion dependent; a student proceeded to a new exercise within a strand only after he had attained some (individually specifiable) performance criterion in his current exercise. Branching between the strands was time dependent; a student moved from one strand to take up where he left off in another after a certain (again, individually specifiable) amount of time, regardless of what criterion levels he had reached in the strands. Within each strand there were two to three progressively more difficult exercises that were designed to bring students to fairly high levels of performance. This criterion procedure was explained in more detail by Atkinson and Fletcher (1972).

Entry into each strand was dependent upon a student's performance in earlier strands. For example, the letter-identification strand started with a subset of letters used in the earliest sight words. When a student in the letter-identification strand exhibited mastery over the set of letters used in the first words of the sight-word strand, he entered that strand. Initial entry into both the phonics and spelling pattern strands was controlled by the student's placement in the sight-word strand. Once he entered a strand, however, his advancement within it was independent of his progress in other strands. On any given day, a student's lesson might draw exercises from one to five different strands.

Most students spent two minutes in each strand and the length of their daily sessions was ten minutes. A student could be stopped at any point in an exercise, either by the maximum-time rule for the strand or by the session time limit; however, sufficient information was saved in his history record to assure continuation from precisely the same point in the exercise when he next encountered that strand.

C. Computer Programming

Development of computer-assisted instruction in computer programming was begun by the Institute in 1968 and was initially made available to students at an inner-city high school in February, 1969. Requisite knowledge of computer languages and systems varies greatly among applications, and for this reason, general concepts of computer operations rather than knowledge of the specific languages or systems used were emphasized in the curriculum. To achieve this generality, the curriculum ranged from problems in assembly-language coding to symbol manipulation and test-processing. The three major components of the curriculum were SIMPER (Simple Instruction Machine for the Purpose of Educational Research), SLOGO (Stanford LOGO), and BASIC. Associated with each component were interpreters, utility routines, and curriculum material.

Basically, computers understand only binary numbers. These numbers may be either data or executable instructions. A fundamental form of programming is to write code as a series of mnemonics, which bear a one-to-one relationship to the binary number-instructions executable by a computer; this type of coding is called assembly-language programming. The instructions of higher order languages, such as BASIC and SLOGO, do not bear a one-to-one relationship to the instructions executed by a computer and, therefore, obscure the fundamental operations performed during program execution. The intent of SIMPER was to make available to students using teletypewriters a small computer that could be programmed in a simple assembly language. The SIMPER computer was, of course, mythical, since giving beginning students such sensitive access to an actual time-sharing computer would be both prohibitively expensive and potentially disastrous.

As simulated, SIMPER was a two-register, fixed-point, single-address machine with a variable size memory. There were sixteen operations in its instruction set. To program SIMPER, a student typed the pseudo operation "LOC" to tell SIMPER where in its memory to begin program execution, and then entered the assembly-language code that comprised his solution to an assigned problem. During execution of the student's program, SIMPER typed the effect of each instruction on its memory and registers. In this way, students received special insight into how each instruction operated and how a series of computer instructions is converted into meaningful work.

SLOGO, the Institute's implementation of LOGO, was the second major component of the curriculum. LOGO is a symbol manipulation and string-processing language developed by a major computer utilities company expressly for teaching the principles of computer programming. It is suitable for manipulating data in the form of character strings, as well as for performing arithmetic functions, and its most powerful feature is its capacity for recursive functions. It was thought that the computer applications most characteristic of the employment available to these students would be the inventory control problems that arise in filing and stock-room management, and it was these problems that were stressed in the SLOGO component of the curriculum. Students were taught not only the SLOGO languages, but the data structures needed for applications such as tree searches and string editing.

SIMPER and SLOGO were more fully documented by Lorton and Slimick (1969). They were written for the Institute's PDP-10 computer and made available to students in the spring and fall of 1969. Mixed with the usual, well-documented enthusiasm of all students for CAI was some disappointment among the computer programming students that they were not learning a computer language generally found in industry. For this reason, the ubiquitous BASIC programming language was prepared for the Institute's PDP-10 computer and made available to the students in the spring of 1970.

The BASIC course, as the SIMPER and SLOGO courses before it, was designed to permit maximum student control. Most of this control concerned the use of such optional material as detailed review, overview lessons, and self-tests. Students were aware that they would be graded only on homework and tests, and it was emphasized that their course grades would not include wrong answers made in the BASIC teaching program.

The course consisted of 50 lessons, each comprising 20 to 100 problems and each requiring one to two hours to complete. The lessons were organized into blocks of five. Each lesson was followed by a review printout and each block of five lessons was followed by a self-test and overview lesson. Students received review printouts, self-tests and overview lessons at their option. Each block was terminated by a short graded test that was evaluated partly by computer and partly by the supervising teacher.

Students were given as much time as needed to answer each problem. Since the curriculum emphasized tutorial instruction rather than drill material, students could spend several minutes thinking or calculating before entering a response; hence, there was no time limit. Because the subject matter of the course was a formal language which was necessarily unambiguous to a computer, extensive analysis of students' responses was possible and highly individualized remediation could be provided for wrong, partially wrong, or inefficient solutions to assigned problems. Significantly, individual errors and misconceptions could be corrected by additional instruction and explanation without incorporating unnecessary exposition in the mainstream of the lesson.

III. PERFORMANCE

We conceive compensatory education to have two broad purposes with respect to student achievement. The first is, of course, to increase the student's achievement level over what it would have been without compensatory education. We discuss achievement gains in III.A. The second purpose of compensatory education is to decrease the spread among students or to make the distribution of educational output more nearly equitable. The notion of equality in education has received considerable attention in recent years, and we made no attempt to review that literature here; Coleman (1968) provided a useful overview of some of the issues. Michelson (1970) discussed inequality in real inputs in producing achievement, and in a later paper (Michelson, 1971) discussed inequality in financial inputs. Our treatment differs in focusing on output inequality and, methodologically, in utilizing tools recently developed by economists for analyzing distribution of income. Section III.B discusses our results in this area.

A. Achievement Gain

Gains in Arithmetic

During the 1967–68 school year, approximately 1,000 students in California, 1,100 students in Kentucky, and 600 students in Mississippi participated in the arithmetic drill-and-practice program. Sufficient data were collected to permit CAI and non-CAI group comparisons for both the California and Mississippi students. The California students were drawn from upper-middle-class schools in suburban areas quite uncharacteristic of those for which compensatory education is usually intended. The Mississippi students, on the other hand, were drawn from an economically and culturally deprived rural area and provided an excellent example of the value of CAI as compensatory education.

The Mississippi students (grades 2 through 6) were given appropriate forms of the Stanford Achievement Test (SAT) in October, 1967. The SAT was administered to the Mississippi first-grade students in February, 1968. All the Mississippi students (grades 1 through 6) were posttested with the SAT in May, 1968. Twelve different schools were used; eight of these included both CAI and non-CAI students, three included only CAI students, and one included only non-CAI students. Within the CAI group, 1 to 10 classes were tested at each grade level, and within the non-CAI group, 2 to 6 classes were tested at each grade level. Achievement gains over the school year were measured by the differences between pre- and posttest grade placements, estimated by the SAT computation subscale. Average pretest and posttest grade placements, calculated differences of these averages, t values for these differences, and degrees of freedom for each grade's CAI and non-CAI students are presented in Table III-1. Significant t values ($p < .01$) are starred. The performance of the CAI students improved significantly more over the school year than that of the non-CAI students in all but one of the six grades. The largest differences between CAI and non-CAI students occurred in grade 1 where, in only three months, the average increase in grade placement for CAI students was 1.14, compared with .26 for the non-CAI students.

On other subscales of the SAT, the performance of CAI students, measured by improvement in grade placement, was significantly better than that of the non-CAI students on the SAT concepts subscale for grade 3 [$t(76) = 3.01$, $p < .01$], and for grade 6 [$t(433) = 3.74$, $p < .01$], and on the SAT application subscale for grade 6 ($t(433) = 4.09$, $p < .01$). In grade 4, the non-CAI students improved more than the experimental group on the concepts subscale [$t(131) = 2.25$, $p < .05$].

Appropriate forms of the SAT were administered to all the California students (grades 1 through 6) in October, 1967, and again in May, 1968. Seven different schools were used. Two of the schools included both CAI and non-CAI students, two included only CAI students, and three included only non-CAI students. Within the CAI group, 5 to 9 classes were tested at each grade level, and within the non-CAI group, 6 to 14 classes were tested at each grade level. Average pretest and posttest grade placements on the SAT computation subscale, calculated differences of these averages, t values for these differences, and degrees of freedom for each grade's CAI and non-CAI students are presented in Table III-2. As in Table III-1, significant t values ($p < .01$) are starred. The performance of the CAI students improved significantly more over the school year than that of the non-CAI students in grades 2, 3, and 5.

TABLE III-1 Average Grade-Placement Scores on the Stanford Achievement Test: Mississippi 1967–68[a]

	Pretest		Posttest		Posttest–Pretest			
Grade	Experimental	Control	Experimental	Control	Experimental	Control	t	Degrees of Freedom
1	1.41 (52)*	1.19 (63)	2.55	1.45	1.13	0.26	9.63**	113
2	1.99 (25)	1.96 (54)	3.37	2.80	1.38	0.84	4.85**	77
3	2.82 (22)	2.76 (56)	4.85	4.04	2.03	1.26	4.87**	76
4	2.34 (56)	2.45 (77)	3.36	3.14	1.02	0.69	2.28	131
5	3.09 (83)	3.71 (134)	4.46	4.60	1.37	0.89	3.65**	215
6	4.82 (275)	4.35 (160)	6.54	5.49	1.72	1.13	4.89**	433

* Values in parentheses are numbers of students.

** $p < .01$.

[a] The assumptions underlying this test of significance are, first, that the two distributions compared are distributed normally and, second, that their variances are equal. Robustness of the t test is discussed by Boneau (1960) and Elashoff (1968) among others.

TABLE III-2 Average Grade-Placement Scores on the Stanford Achievement Test: California 1967–68[a]

Grade	Pretest		Posttest		Posttest–Pretest		t	Degrees of Freedom
	Experimental	Control	Experimental	Control	Experimental	Control		
1	1.39 (58)*	1.31 (259)	2.62	2.51	1.23	1.21	0.20	315
2	2.06 (65)	2.16 (238)	3.20	2.89	1.14	0.73	4.96**	301
3	3.00 (136)	2.85 (210)	4.60	3.86	1.60	1.02	6.70**	344
4	3.40 (103)	3.49 (185)	4.87	5.00	1.46	1.51	−0.41	286
5	4.98 (149)	4.44 (90)	6.41	5.31	1.43	0.88	4.06**	237
6	5.42 (154)	5.70 (247)	7.43	7.59	2.01	1.90	0.84	399

* Values in parentheses are numbers of students.

** $p < .01$.

[a] The assumptions underlying this test of significance are, first, that the two distributions compared are distributed normally and, second, that their variances are equal Robustness of the t test is discussed by Boneau (1960) and Elashoff (1968) among others.

On other subscales of the SAT, the CAI students improved significantly more over the school year than did the non-CAI students on the concepts subscale for grade 3 [$t(344) = 4.13$, $p < .01$] and on the application subscale for grade 6 [$t(399) = 2.14$, $p < .05$].

A comparison of the California students with the Mississippi students suggests at least two observations worth noting. First, when significant effects were examined for all six grades, the CAI program was more effective for the Mississippi students than for the California students. Second, changes in performance level for the CAI groups were quite similar in both states, but the non-CAI group changes were very small in Mississippi relative to the non-CAI group changes in California. These observations suggest that CAI may be more effective when students perform well below grade level and are in need of compensatory education, as in the rural Mississippi schools, than when the students receive an adequate education, as in the suburban California schools.

These data do not fully reflect the breadth of educational experience permitted by CAI. Some of the Mississippi students took the Institute's beginning course in mathematical logic and algebra, which had been prepared for bright fourth- to eighth-grade students whose teachers were not prepared to teach this advanced material. At the end of the 1967–68 school year, two Mississippi Negro boys placed at the top of the first-year mathematical logic students, almost all of whom came from upper-middle-class suburban schools.

Gains in Reading

The reading data used in this report were also discussed by Fletcher and Atkinson (1972). In November of the school year, 25 pairs of first-grade boys and 25 pairs of first-grade girls were matched on the basis of the Metropolitan Readiness Test (MET). Matching was achieved so that the MET scores for a matched pair of subjects were no more than two points apart. Moreover, an effort was made to insure that both members of a matched pair had classroom teachers of roughly equivalent ability.

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.

Four schools within the same school district were used. Two schools provided the CAI students and two different schools provided the non-CAI subjects. The schools were in an economically depressed area eligible for federal compensatory education funds.

Three posttests were administered to all subjects in late May and early June. Four subtests of the Stanford Achievement Test (SAT), Primary I, Form X, were used: 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. Third, a test (DF) developed at Stanford and tailored to the goals of the CAI reading curriculum was administered individually to all subjects.

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 and t values for differences in SAT, COOP, and DF total scores are presented in Table III-3. In this table t values are displayed in brackets. The t values calculated are for nonindependent samples, and those that are significant ($p < .01$, one-tailed) are starred.

The results of these analyses were encouraging. All three indicated a significant difference in favor of the CAI reading subjects. These differences were also important from the standpoint of improvement in estimated grade placement. Table III-4 displays the mean grade placement of the two groups on the SAT and COOP.

Means and t values for the differences on the four SAT subtests are presented in Table III-5. As in Table III-3, t values are displayed in brackets; t values that are significant ($p < .01$, one-tailed) are starred.

These SAT subtests revealed some interesting results. 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 was the lack of any real differences for the S/VOC. One explanation for this result is that the vocabulary subtest measures a pupil's

TABLE III-3 Means and t Values[a] for the Stanford Achievement Test (SAT), the California Cooperative Primary Test (COOP), and the CAI Reading Project Test (DF)[b]

	SAT	COOP	Degrees of Freedom
CAI	112.7	33.4	64.5
	[4.22*]	[4.04*]	[6.46*]
Non-CAI	93.3	26.2	54.8

* $p < .01$, df = 43.

[a] t values are shown in brackets.

[b] The assumptions underlying this test of significance are, first, that the two distributions compared are distributed normally and, second, that their variances are equal. Robustness of the t test is discussed by Boneau (1960) and Elashoff (1968) among others.

TABLE III-4 Average Grade Placement on the Stanford Achievement Test (SAT) and the California Cooperative Primary Test (COOP)

	SAT	COOP
CAI	2.3	2.6
Non-CAI	1.9	2.1

vocabulary independent of his reading skill (Kelley et al., 1964); since the CAI reading curriculum was primarily concerned with reading skill and only incidentally with vocabulary growth, there may have been no reason to expect a discernible effect of the CAI curriculum on the S/VOC. Most notable, however, were the S/PM results. 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 were surprising.

The effect of CAI on the progress of boys compared with the progress of girls is interesting. The Atkinson (1968) finding that boys benefit more from CAI instruction than do girls was corroborated by these data. On the SAT the relative improvement for boys exposed to CAI versus those not exposed to CAI was 22 per cent; the corresponding figure for girls was 20 per cent. On the COOP the percentage improvement due to CAI was 42 for boys and 17 for girls. Finally, on the DF the improvement was 32 per cent for boys and 13 per cent for girls. Overall, these data suggested that both boys and girls benefit from CAI instruction in

TABLE III-5 Means and t Values[a] for the Word Reading (S/WR), Paragraph Meaning (S/PM), Vocabulary (S/VOC), and Word Study (S/WS) Subtests of the Stanford Achievement Test[b]

	S/WR	S/PM	S/VOC	S/WS
CAI	26.5	23.0	21.6	41.6
	[5.18*]	[4.17*]	[.35]	[3.78*]
Non-CAI	20.1	16.3	21.2	35.7

*$p < .01$, df = 43.
[a] t values are shown in brackets.
[b] The assumptions underlying this test of significance are, first, that the two distributions compared are distributed normally and, second, that their variances are equal. Robustness of the t test is discussed by Boneau (1960) and Elashoff (1968) among others.

reading, but that CAI is relatively more effective for boys. Explanations of this difference were discussed by Atkinson (1968).

Achievement Gains in the Computer Programming Course

Eight weeks prior to the end of the 1969–70 school year, students who received CAI instruction in BASIC were given the SAT's mathematical computation and application sections. A control group of students from the same school was given the same test. At semester's end, the test was repeated and the following additional data were gathered: (i) verbal achievement scores from the ninth-grade-level test of the Equality of Educational Opportunity Survey, and (ii) responses to the socioeconomic status questionnaire of the EEO survey.

Sufficient pre- and posttest scores were obtained for 39 CAI students and 19 non-CAI students. Average pre- and posttest scores for the SAT computation and application subscales, average gains, and t values for differences in the average gains achieved by CAI and non-CAI students are presented in Table III-6.

The SAT tests were used here in the absence of a standardized achievement test in computer programming; gains in arithmetic achievement were, then, only a proxy for gains in the skills taught in the course. Presumably, students gained in arithmetic skill because they spent more than the usual time working on quantitative problems.

There was also a good deal of textual output at the teletype that the students needed to read and comprehend, and it was the unanimous impression of the teachers who worked with the students that they were better able to read as a result. However, scores on verbal achievement tests administered at the end of the school year showed virtually no differences between the CAI and control groups in reading ability.

In order to identify some of the sources of achievement gain, we ran a

TABLE III-6 Arithmetic Achievement for Computer Programming Course[a]

	CAI			Control				Degrees of
	Pre	Post	Gain	Pre	Post	Gain	t	Freedom
SAT computation	7.97	9.11	1.14	7.97	8.41	.44	1.68	55
SAT application	7.74	8.61	.86	8.33	8.38	.05	1.73	55

[a] The assumptions underlying this test of significance are, first, that the two distributions compared are distributed normally and, second, that their variances are equal. Robustness of the t test is discussed by Boneau (1960) and Elashoff (1968) among others.

stepwise linear regression of gain scores (posttest minus pretest) against pretest scores, verbal scores, and various items from the socioeconomic status (SES) questionnaire. The dependent variable was the sum of the gain scores on the computation and applications sections of the test. Table III-7 below lists the independent variables and the coefficients estimated for them.

The results in the table are self-explanatory, but we make two comments in conclusion. First, even though the regression coefficient on CAI was not significant at standard levels, in magnitude it was substantial; failure to have had CAI during this eight-week interval would remove about .5 years (one-half of .99) of arithmetic achievement. Naturally, it would be desirable to replace the 0–1 CAI variable with actual amount of time on the system; the regression coefficient would then have a good deal more practical value. Second, the mathematics pretest had a negative coefficient; when CAI and control regressions

TABLE III-7 Determinants of Achievement Gain[a,b]

Independent Variable	Mean	Standard Deviation	Regression Coefficient	Standard Error
Constant term			4.40	
CAI {0 CAI group; 1 control group}	.35	.48	−.99	.96
Sum of pretest scores on computation and application	15.3	4.22	−.26	.14
Raw score on verbal test	27.6	9.9	.17	.06
Age in years	15.9	2.5	−.23	.20
Race {0 Caucasian; 1 Other}	.23	.42	−1.44	1.18
Number of people living in child's home	5.63	1.86	.13	.29
Total years of schooling of both parents	15.5	10.52	−.02	.05
Educational aspiration of student, in years of schooling	15.4	4.45	.07	.11
Previous math grade placement achievement of student	2.40	1.30	−.11	.39

[a] Dependent variable is the sum of students' gain scores on arithmetic and computation sections of SAT.
[b] $R^2 = .26$.

were run separately, this coefficient was negative for CAI and positive for control. This result implies that CAI in sufficient quantity would have an equalizing effect, a point to be further discussed in the next subsection. We plan, in a later paper, to analyze in more detail the interaction of CAI and student background characteristics as determinants of scholastic achievement (Wells, Whelchel, and Jamison, 1973).

B. Reduction in Inequality

Our second criterion of performance concerned the extent to which CAI is inequality reducing. Clearly, any compensatory program that has positive achievement gain, if applied only to those sectors of the population who perform least well, will have a tendency to reduce inequality. Often, however, entire schools receive the compensatory education and it is less likely that the program is inequality reducing. Our purpose in this subsection was to use techniques developed for analyzing inequality in the distribution of income to provide concrete measures of the extent to which CAI is inequality reducing. These measures are as applicable when an entire student population receives the compensatory treatment as when only some subset of the population does.

We first used a traditional measure of inequality—the Gini coefficient based on the Lorenz curve—to examine before and after inequality in CAI and control groups and to examine inequality in achievement gains. Use of the Gini coefficient as a measure of inequality has, however, a number of shortcomings that were reviewed by A. Atkinson (1970). Prominent among these is that it is not purely an empirical measure but contains an underlying value judgment concerning what constitutes more inequality. Newbery (1970) showed that it is impossible to make this value judgment explicit by means of any additive utility function. Therefore, we also used the inequality measure proposed by A. Atkinson that makes explicit the underlying value judgments.

Use of either the Atkinson measure or Gini coefficients implies that achievement test scores should be measured on a ratio scale (i.e. the achievement measure must be unique up to multiplication by a positive constant). If, for example, achievement measures were only unique up to a positive linear transformation, the Gini coefficient could be made arbitrarily small by adding an arbitrarily large amount to each individual's achievement test score. The reader is cautioned that our assumption that achievement is measured on a ratio scale is quite strong; on the other hand, a ratio scale is implicit in the assumption that one test score is better than another if—and only if—the number of problems correct on the one test is greater than the number correct on the other.

Inequality Measured by the Gini Coefficient

Consider a group of students who have taken an achievement test; each student will have achieved some score on the test, and there will be a total score obtained by summing all the individual scores. We may ask, for example, what fraction of the total score was obtained by the 10 per cent of students doing most poorly on the test, what fraction was obtained by the 20 per cent of students doing most poorly, and so on. The Lorenz curve plots fraction of total score earned by the bottom x per cent of students as a function of x.

These concepts may be expressed more formally in the notation of Levine and Singer (1970) as follows. Let $N(u)$ be the achievement-score density function. The $N(u)du$ represents the number of individuals scoring between u and $u + du$. The total number of students, N, and their average score, A, are given by

$$N = \int_0^\infty N(u)du$$

and

$$A = \frac{1}{N} \int_0^\infty uN(u)du$$

The fraction of students scoring a or less is given by

$$f(a) = \frac{1}{N} \int_0^a N(u)du$$

and the fraction of the total score obtained by students scoring a or less is

$$g(a) = \frac{\int_0^a uN(u)du}{NA}.$$

The Lorenz curve plots $g(a)$ as a function of $f(a)$, and a typical Lorenz curve for our results is shown in Figure III-1 below. The $f(a)$, $g(a)$ pairs are obtained by computing these functions for all values of a. If there were a perfectly equitable distribution of achievement (everyone having identical achievement) the Lorenz curve would be the 45° line depicted in Figure III-1. The more $g(a)$ differs from the 45° line the more inequitable is the distribution of achievement. The Gini coefficient is an aggregate measure of inequality that is defined as the ratio of the area between $g(a)$ and the 45° line to the area between the 45° line and the abscissa. If the Gini coefficient is zero, the distribution of achievement is completely uniform; the larger the Gini coefficient, the more unequal the distribution.

In order to examine the extent to which the different CAI programs described in Section II of this paper were in fact inequality reducing, we

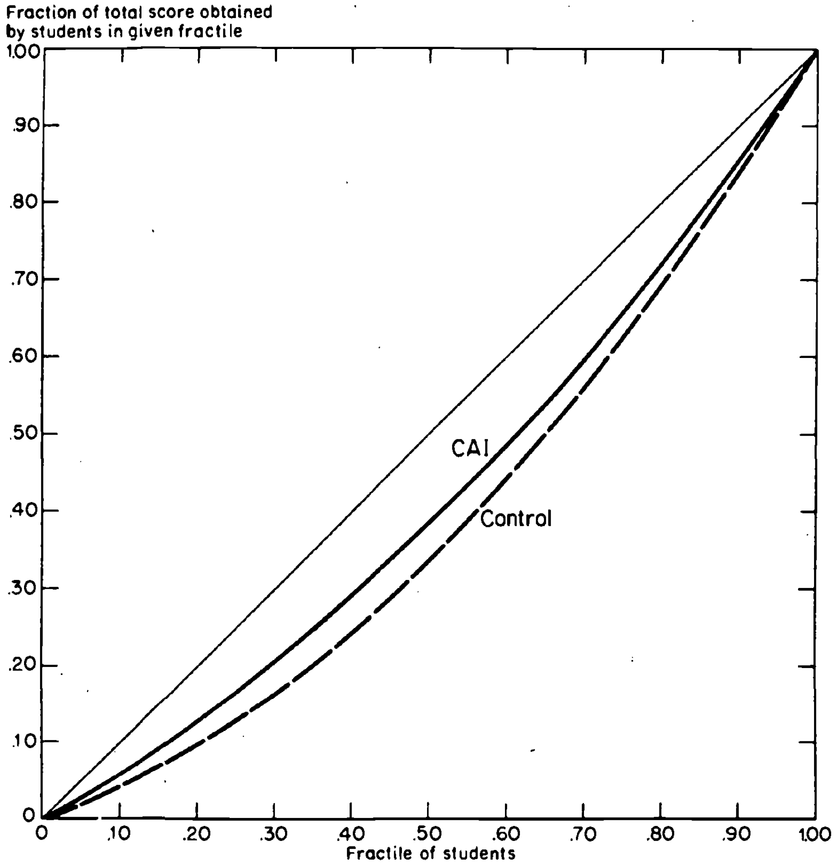

FIGURE III-1 Typical Lorenz Curves: Distribution of Reading Posttest Achievement (COOP)

computed Gini coefficients for the distribution of achievement before and after the CAI was made available for both the CAI and the control groups. In Table III-8 these Gini coefficients are presented for both the high school level computer programming course and the elementary arithmetic course in Mississippi and California grades 1 through 6. For each group at each grade level, we give the Gini coefficients for the pretest for the group as a whole, the Gini coefficients for the posttest for the group as a whole, and the difference between those two Gini coefficients. Similar information is given for the control group. In the final column of the table the difference between columns 3 and 6 of the table is shown; if this difference is positive, it indicates that there is more of a reduction in inequality in the CAI group than in the control

TABLE III-8 Gini Coefficients for CAI and Control Groups

Group	CAI Pre (1)	CAI Post (2)	CAI Pre-Post (3)	Control Pre (4)	Control Post (5)	Control Pre-Post (6)	Col. (3) Minus Col. (6)
Computer Programming							
SAT COMP R.S.[a]	.113	.087	.026	.108	.096	.012	.014
SAT APPL R.S.[b]	.119	.111	.008	.084	.097	−.013	.021
SAT COMP G.P.[c]	.079	.066	.013	.075	.070	.005	.008
SAT APPL G.P.[d]	.080	.079	.001	.059	.069	−.010	.011
Math Drill[e] and Practice							
Mississippi 1967–68							
Grade 1	.057	.067	−.010	.037	.062	−.025	.015
Grade 2	.064	.039	.025	.055	.050	.005	.020
Grade 3	.016	.032	−.016	.035	.038	−.003	−.013
Grade 4	.080	.053	.027	.084	.065	.019	.008
Grade 5	.095	.070	.025	.078	.079	−.001	.026
Grade 6	.068	.077	−.009	.078	.084	−.006	−.003
California 1967–68							
Grade 1	.058	.077	−.019	.054	.075	−.021	.002
Grade 2	.075	.056	.019	.073	.062	.011	.008
Grade 3	.042	.063	−.021	.050	.060	−.010	−.011
Grade 4	.067	.053	.014	.065	.058	.007	.007
Grade 5	.056	.048	.008	.055	.068	−.013	.021
Grade 6	.077	.073	.004	.065	.070	−.005	.009

[a] Gini coefficients from Stanford Achievement Test, computations subscale, raw scores.
[b] Gini coefficients from Stanford Achievement Test, application subscale, raw scores.
[c] Gini coefficients from Stanford Achievement Test, computation subscale, grade placements.
[d] Gini coefficients from Stanford Achievement Test, application subscale, grade placements.
[e] Gini coefficients for all math drill and practice from Stanford Achievement Test, computation subscale, grade placements.

group. For the high school CAI group we computed the Gini coefficients for both raw scores and grade placement scores and the differences between those two computations can be seen in the table. We applied a sign test to the 12 arithmetic cases and the two computer programming cases that used grade placement scores to test the significance of the hypothesis that inequality was reduced more in the CAI groups than in the control groups. From column 7 of Table III-8 it can be seen that in only 3 of the 14 cases was the CAI less inequality-reducing than no CAI. The sign test then implied an acceptance of the hypothesis that CAI is inequality reducing at the .05 level.

In Table III-9 we show the Gini coefficients for CAI and control

TABLE III-9 Gini Coefficients for Reading Achievement Posttests[a]

	CAI	Control	Control − CAI
SAT	.134	.174	.040
COOP	.183	.266	.083
DF	.068	.152	.084
S/WR[(1)]	.140	.209	.069
S/PM[(2)]	.226	.396	.170
S/WS[(3)]	.119	.149	.030
S/VOC[(4)]	.170	.183	.013

[a] Due to careful matching of CAI and control groups by pretest achievement (on the Metropolitan Readiness Test—see Section III.A), pretest Gini coefficients are not shown.

groups for the various sections of the reading achievement posttests. We did not include the pretest scores since different tests were used and the results were not directly comparable. In all 7 cases in Table III-9 the Gini coefficient was less for the CAI group than for the control group; the hypothesis that CAI is inequality reducing was substantiated in this case at the .01 level. The widely held subjective impression that no students in the reading CAI groups are "lost" seems, then, to be strongly supported by these data. It is reasonable to expect that the effect of CAI on posttests would correlate positively with the Gini coefficient differences obtained from the CAI and non-CAI subjects. The difference in Gini coefficients should be greatest where the CAI treatment is greatest and this seems to be the case. The effect of CAI was statistically significant on the S/WR, S/PM and S/WS, and for these subtests the Gini coefficient differences were fairly large. There was only a slight positive effect of CAI in the S/VOC, and the Gini coefficient differences for this subtest were correspondingly small.

Value Explicit Measures of Inequality

In this section we consider a measure of inequality proposed by A. Atkinson (1970) that makes explicit the value judgment entering into the comparison of the inequality of two distributions. Atkinson drew, in his discussion of greater and lesser inequality, on a close parallel between the concept of greater risk (or greater spread) in a probability distribution and the concept of greater inequality in a distribution of income. He was thus able to apply certain results concerning the ordering by riskiness of probability distributions to ordering by degree of inequality of income distributions. He showed that a variety of conventional measures of inequality—including variance, coefficient of variation, relative mean deviation, Gini coefficient, and standard deviation of logarithms—are necessarily consistent with the ordering implied by concave utility functions. That is, one can, in general, find a concave utility function that is consistent with the ordering induced by any of the above measures.

Atkinson then proposed that the overall utility, W, of a distribution of achievement scores, $N(u)$, be represented by the following formula

$$W = \int_0^{\bar{u}} U(u)\, N(u)\, du$$

when $\bar{u}$ is the maximum score achieved on the test. It is assumed in the above that $U(u)$ is increasing and concave, i.e., that $U'(u)$ is greater than 0 and that $U''(u)$ is less than 0.[1] The concavity implies, for that particular population, that there is an aversion to inequality. Given this aversion to inequality there exists a level of achievement, u_e, that is *lower* than the average level of achievement in the population under consideration such that if everyone in the population had exactly a u_e level of achievement, the overall level of social welfare would remain constant at W. Following Atkinson, we call u_e the "equally distributed equivalent" level of achievement. Clearly, u_e, in general, depends on the form of U; however, by direct analogy with the theory of choice under uncertainty, u_e is invariant with respect to positive linear transformations of U.

If μ is the average level of achievement in the society, then a reasonable measure of inequality, I, is given by the following formula

$$I = 1 - \frac{u_e}{\mu}$$

The lower I is, the more equal is the distribution of achievement; to put this another way, as u_e gets closer to μ, the cost of having inequality gets lower. The I ranges between 0 for complete equality and 1 for complete inequality and tells us, in effect, by what percentage total achievement could be reduced to obtain the same level of W if the achievement level were equally distributed.

In order to apply the measure I we need an explicit formulation of U. In this paper we consider two classes of functions U. The first of these was suggested by Atkinson and has the property of "constant relative inequality aversion." Constant relative inequality aversion means that multiplying all achievement levels in the distributions by a positive constant does not alter the measure I of inequality. If there is constant relative inequality aversion, it is known from the theory of risk aversion that $U(u)$ must have the following form

$$U(u) = a + B\,\frac{u^{1-\epsilon}}{1-\epsilon} \text{ if } \epsilon \neq 1$$

and

$$U(u) = ln(u) \text{ if } \epsilon = 1$$

Another possibility that Atkinson considered was that of constant *absolute* inequality aversion, which means that adding a constant to each achievement level in the distribution does not affect the measure of inequality. A theorem of Pfanzagl (1959) shows that if there is constant absolute inequality aversion, then $U(u)$ must have one of the following two forms

$$U(u) = au + b$$

or

$$U(u) = a\,\lambda^{u} + b$$

Strict concavity implies the latter of these two and that $0 < \lambda < 1$.

We thus have two families of utility functions, one indexed by ϵ and the other by λ, which include a large number of qualitatively important alternatives for U. In Figure III-2, $U(u)$ is shown for several values of ϵ and in Figure III-3 $U(u)$ is shown for several values of λ. Since transforming the functions depicted in Figures III-2 and III-3 by a positive linear transformation does not affect the measure I, the height and location of the functions in those two figures is arbitrary.

It is clear from the preceding, that the measure I of inequality for any fixed distribution of achievement varies with ϵ or λ. In Figure III-3 we have constrained $U(u)$ to pass through 0 and 1 for all values of λ, implying that $U(u) = (1 - \lambda^{u})/(1 - \lambda)$. For λ very close to 1, inequality is close to 0; as λ gets smaller, inequality gets larger for any fixed distribution. The way in which I varies with ϵ is just the opposite; low values of ϵ give a low measure of inequality, whereas large values of ϵ give large values for I.

Figures III-4 and III-5 illustrate I plotted as a function of ϵ and as a function of λ for one particular CAI group and its control. The distributions $N(u)$ are of posttest scores and they are for a case where there was little difference in inequality on the pretest, as measured by the Gini coefficients of the CAI and control groups.

One benefit in having a measure of inequality indexed by some parameter describing degree of inequality aversion (such as λ or ϵ) is that the control group may be judged to be more equal for some values of λ and ϵ but less equal for others. Table III-10 shows such reversals as a function of ϵ under the assumption of constant relative inequality aversion. Table III-11 shows the same information as a function of λ.

We have attempted to provide explicit measures of the extent to which the three CAI programs we reviewed are inequality-reducing. We used recent work on measurement of inequality that has appeared in the economics literature to show that, ultimately, measurement of inequality rests on either an implicit or explicit value judgment. We showed measures of inequality for CAI and control groups for several explicit

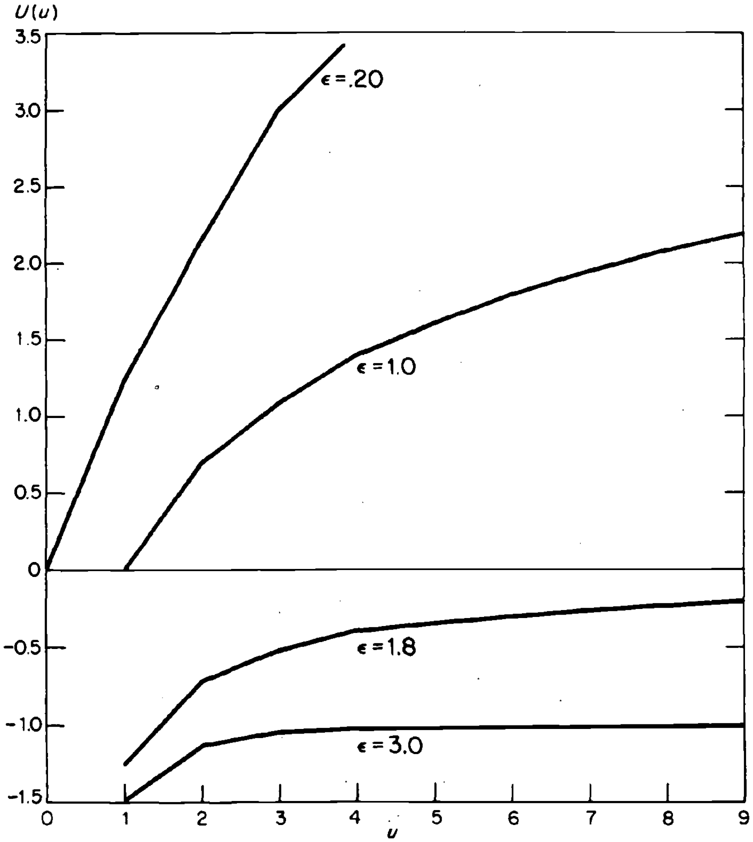

FIGURE III-2 $U(u)$ **for Several Values of** ϵ

classes of value judgments concerning distribution of achievement. It is perhaps worth stressing that as we were actually designing and implementing our CAI programs, we did not have inequality-reduction in mind as an explicit goal; our results, literally, just turned out this way.

The next step to take at this point, we feel, is to design patterns of CAI presentation that are optimal by some utility function U maximized to a variety of constraints. One constraint could be the distribution of prior achievement in the population to which we are providing CAI; another constraint could be the total number of terminal hours per month available to that population; still another constraint could be possible impositions from the school administration that no students get less than a certain amount of CAI or more than a certain amount of CAI per day on an average; and a final fundamental constraint could be the production function that relates time on the system and other factors to gains in

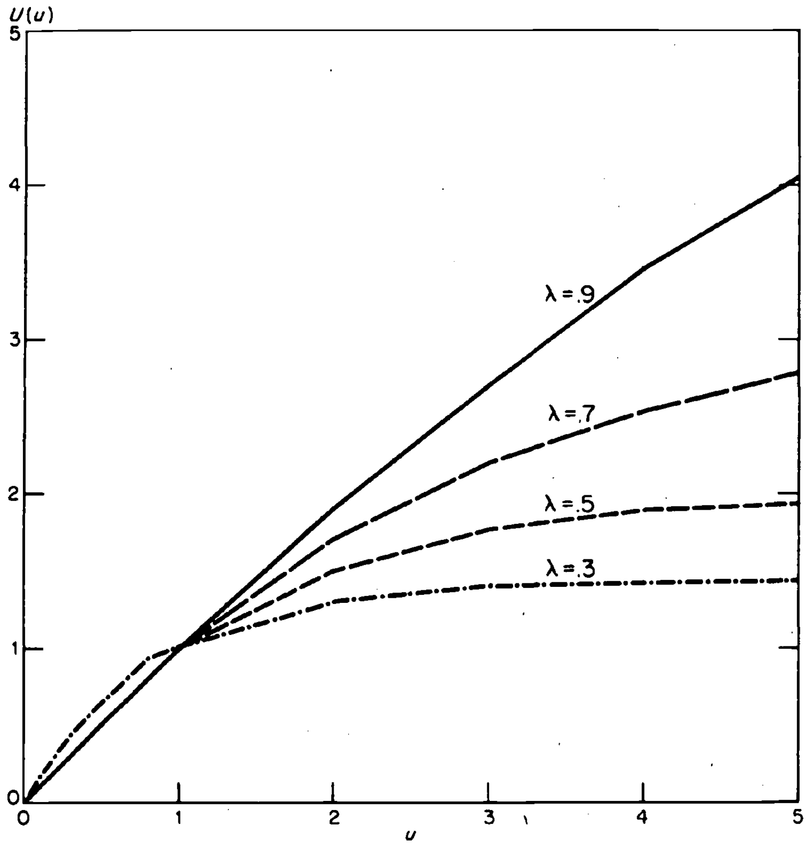

FIGURE III-3 ***U(u)* for Several Values of λ**

student achievement. What we plan to examine in the future is how the solution to this optimization problem varies with U when the constraints vary. We can then design patterns of instruction that are explicitly tailored to several separate U's and empirically examine the extent to which we obtain the stated objectives. We hope that in this fashion any tradeoffs that might exist between total achievement gain and inequality-reduction can be made explicit both in terms of the underlying technology and the underlying value structure.

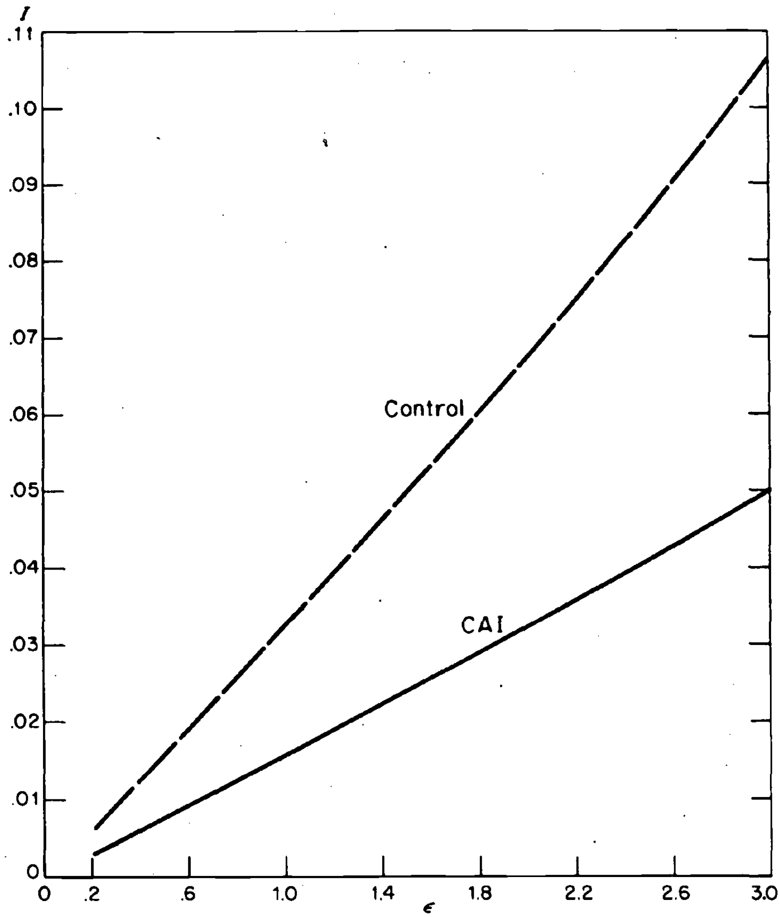

FIGURE III-4 *I* **as a Function of** ε **for Fifth-Grade Arithmetic, California, 1967–68**

IV. COST OF COMPUTER-ASSISTED INSTRUCTION

A. General Considerations

It is useful to place CAI costs into three broad categories. The first category comprises the terminal equipment used by the students. Terminals vary in complexity from a simple touch-tone pad used for telephones to a CRT with keyboard, light pen, audio, and random-access slide screen, and costs vary accordingly. The second cost category com-

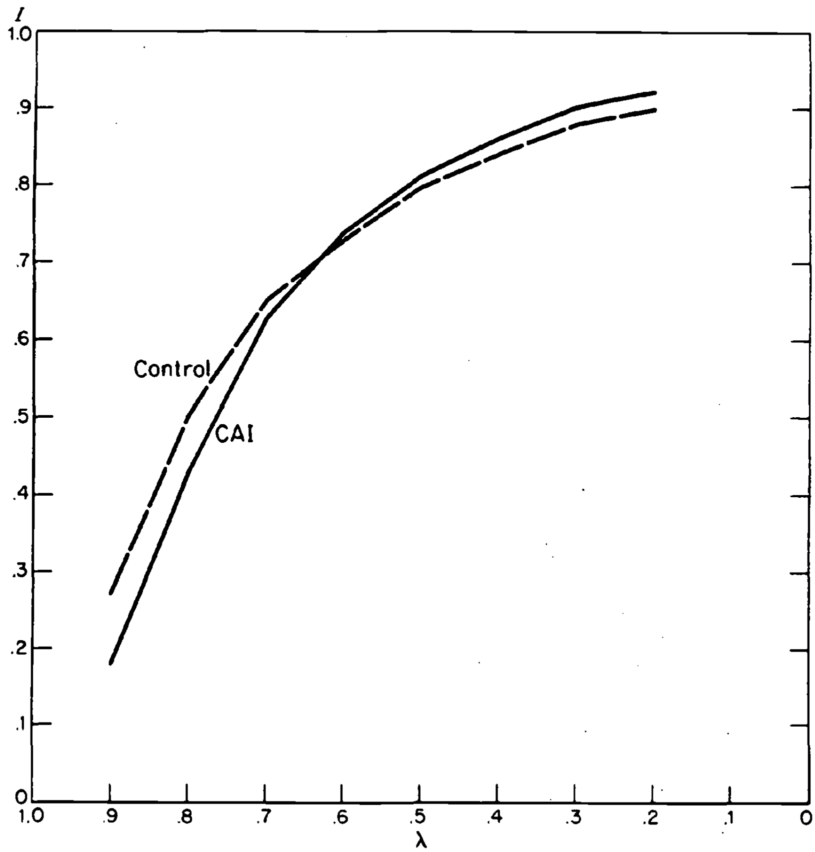

FIGURE III-5 ***I* as a Function of λ for Fifth-Grade Arithmetic, California, 1967–68**

prises the computer system that decides on and stores instructional presentations and evaluates student responses. This category includes the central processing unit, disc and core storage, high-speed line units, and peripheral equipment. The final cost component is the multiplexing and communication system that links the student terminals to the main computer system. This communication system can be reasonably simple when the terminals are located within a few hundred feet of the computer. If the terminals are dispersed, the communication system may include a communication satellite as well as one or more small computers that assemble and disassemble signals.

Up to this point we have mentioned only the cost components necessary to *provide* CAI and have assumed that the curriculum to be used has already been programmed. Only the cost of provision is considered

TABLE III-10 CAI Inequality Reduction: Constant Relative Inequality Aversion[a]

Student Group (Math Drill and Practice)	ϵ .20	.60	1.0	1.4	1.8	2.2	2.6	3.0
Mississippi 1967–68								
Grade 1	.001	.002	.004	.005	.006	.007	.007	.007
Grade 2	.004	.012	.020	.030	.041	.054	.068	.084
Grade 3	−.002	−.005	−.008	−.012	−.015	−.019	−.024	−.029
Grade 4	.002	.005	.009	.014	.020	.028	.038	.050
Grade 5	.005	.012	.019	.023	.026	.027	.025	.022
Grade 6	.000	−.002	−.003	−.004	−.006	−.007	−.009	−.010
California 1967–68								
Grade 1	.000	.000	.000	.000	.000	.001	.002	.002
Grade 2	.002	.004	.007	.009	.011	.014	.016	.019
Grade 3	−.002	−.006	−.010	−.015	−.021	−.027	−.035	−.045
Grade 4	.004	.001	.001	.000	−.003	−.007	−.013	−.022
Grade 5	.003	.010	.017	.025	.034	.042	.052	.062
Grade 6	.002	.006	.010	.015	.022	.030	.039	.051

[a] The numbers shown in the table are $I_A - I_B$ as a function of ϵ. I_A is the difference in inequality between CAI and control after treatment (i.e., on the posttest) and I_B is the difference before treatment. If the difference is greater after treatment than before, CAI is inequality-reducing.

TABLE III-11 CAI Inequality Reduction: Constant Absolute Inequality Aversion[a]

Student Group (Math Drill and Practice)	λ .90	.80	.70	.60	.50	.40	.30	.20
Mississippi 1967–68								
Grade 1	−.001	−.005	−.009	−.011	−.013	−.005	.011	.030
Grade 2	.010	.041	.090	.127	.146	.148	.139	.120
Grade 3	−.131	−.180	−.237	−.297	−.331	−.331	−.300	−.246
Grade 4	−.013	.016	.050	.054	.044	.033	.024	.017
Grade 5	.048	.006	−.010	−.007	.000	.004	.009	.016
Grade 6	−.083	−.108	−.098	−.078	−.060	−.046	−.037	−.030
California 1967–68								
Grade 1	.032	.069	.086	.086	.081	.078	.077	.076
Grade 2	−.018	−.038	−.041	−.031	−.020	−.012	−.006	.001
Grade 3	−.078	−.116	−.158	−.173	−.160	−.246	−.118	−.096
Grade 4	.050	.044	.012	−.010	−.024	−.031	−.033	−.036
Grade 5	.092	.071	.021	.002	−.004	−.004	−.005	−.006
Grade 6	−.020	.045	.045	.038	.034	.031	.029	.027

[a] The numbers shown in the table are $I_A - I_B$ as a function of λ. I_A is the difference in inequality between CAI and control after treatment (i.e., on the posttest) and I_B is the difference before treatment. If the difference is greater after treatment than before, CAI is inequality-reducing.

here. Of course, unless ways are found to share a single curriculum among many users, the per student cost of curriculum preparation can be prohibitively high. Levien (1972) discussed how to provide incentives and how to recoup costs for CAI curriculum preparation. Since a reasonably large body of tested curriculums already exists, we consider those costs sunk and do not include them here.

There appear to be two trends in design philosophy for the computer component of a CAI system. One trend is toward large, highly flexible systems capable of simultaneously providing curriculums in many subjects to a large number of simultaneous users. The other trend is toward small, special-purpose computer systems capable of providing only one or two curriculums to a few students. A large, general-purpose computer system might have 1,000 or more student terminals simultaneously in use. The proposed PLATO IV system of the University of Illinois is now aiming for about 500–3,000 fewer than originally planned (Stifle, 1972); the small special-purpose system is likely to have 8 to 32 terminals. Naturally, the number of terminals per computer has important implications for the communication system. In order to make a large system worthwhile, an extensive communication system is almost inevitable. On the other hand, even a moderate-sized elementary school could use a 16-terminal system, and only simple communications would be required. The potential scale economies of a large computer system, its broader range of offerings, and its easy updating must be balanced, then, against the lower communication costs of special-purpose systems.

Jamison, Suppes, and Butler (1970) examined the cost of providing CAI in urban areas by way of a small special-purpose computer system, the first of which is now in operation. Rather than review those costs here, we refer the reader to that paper. Costs per student per year are approximately $50 above the normal cost of educating the child, assuming that the school system in no way attempts to reduce other costs (for example, by increasing the student-teacher ratio) as a result of introducing CAI.

B. Cost of Providing CAI in Rural Areas

The most distinctive aspect of providing CAI in rural areas is that the students to be reached are highly dispersed and thus tend to be reasonably distant from a central computer. One could use small computers for rural areas at costs somewhat higher than Jamison et al. estimated for urban areas. To obtain the advantages of a large central system, however, the communication system must be sophisticated. In this section, we examine the cost of providing large-scale CAI in rural areas. To obtain per student annual-cost figures we examine each of the three cost areas mentioned above and then combine them to give the final figures. Our costs were based on the CAI system at IMSSS, using the curriculum already available; other systems could have different costs.

Terminal Costs

The cost of a Model-33 teletypewriter, including modifications, is about $850. To provide the teletypewriter terminal with a computer-controlled audio cassette would increase the cost about $150, but since this is not operational now, the additional $150 was not included in our estimates. An alternative would be to lease the teletypewriters. That cost is about $37 per teletypewriter per month and includes maintenance.

Computer Facility Costs

Cost estimates are provided for a system capable of running about 1,000 students at a time. The system would be run at "4/5 diversity," i.e., 1,250 terminals would be attached to the system under the assumption that no more than 4/5 of the 1,250 would run at any one time. The assumption of 4/5 diversity is conservative, given our past experience.

The system would comprise two PDP-10 computers, each with a 300 million 8-bit byte disc, 512K 36-bit words of core memory, a swapping drum, and appropriate I/O and interfacing devices. The system would essentially be a doubled 500-terminal system; if, however, appreciably more terminals were desired, other designs would be appropriate.

Table IV-1 shows the initial costs of the system and Table IV-2 shows annual costs. Overhead is *not* included.

In order to express all costs as annual costs, we multiplied the $3,260,000 by .15, assuming about a ten-year equipment lifetime and 10 per cent social discount rate. Thus, the annual cost of the initial equip-

TABLE IV-1 Initial Costs, Computer Components of CAI System[a]

Component	Cost
Computer system	$2,560
Spare parts and test equipment	200
Planning and installation	350
Building	150
Total	$3,260

[a] Costs in thousands of dollars.

TABLE IV-2 Annual Costs, Computer Components of CAI System[a]

Component	Annual Cost
System operation	$150
System maintenance	175
Building maintenance	20
Supplies	35
Total	$380

[a] Costs in thousands of dollars.

ment purchase is $490,000. When added to the direct annual costs, the total is $870,000 per year. With 1,250 terminals, the central facility cost is $690 per terminal per year.

Communication Costs

In an unpublished paper, Jamison, Ball, and Potter (1971) examined in some detail the cost of communication between a central computer facility and rural terminals.[2] They considered two types of systems—one using commercial phone services and one using a single transponder on a communication satellite. Costs of communicating by satellite are independent of distance whereas phone costs are distance-dependent. Thus, for longer distances, satellites are increasingly attractive. Figures IV-1 and IV-2, taken from Jamison, Ball, and Potter, show the annual cost of communication and multiplexing for satellite and terrestrial systems. Both assume that the terminals are clustered in groups of eight. The graphs assume "best estimate" satellite and phone-service costs in 1975 and eight-year equipment lifetime with 10 per cent cost of capital. They also include maintenance and system installation but do not include overhead.

The present engineering cost estimates for G, the satellite ground-station cost, is $10,000, but this is the estimate for a feasible, not optimal system—we expect much engineering improvement. Thus, Figure IV-1 shows that the annual communication cost for a satellite distribution system would be about $800,000. From Figure IV-2 we see that if D, the average distance between the central computer facility and the terminals, exceeds about 550 miles, then communication via satellite is cheaper than via telephone.[3] Since the average distance to the terminals is quite likely to exceed 550 miles, $800,000 is our estimate of communication and multiplexing cost. This comes to $640 per terminal per year.

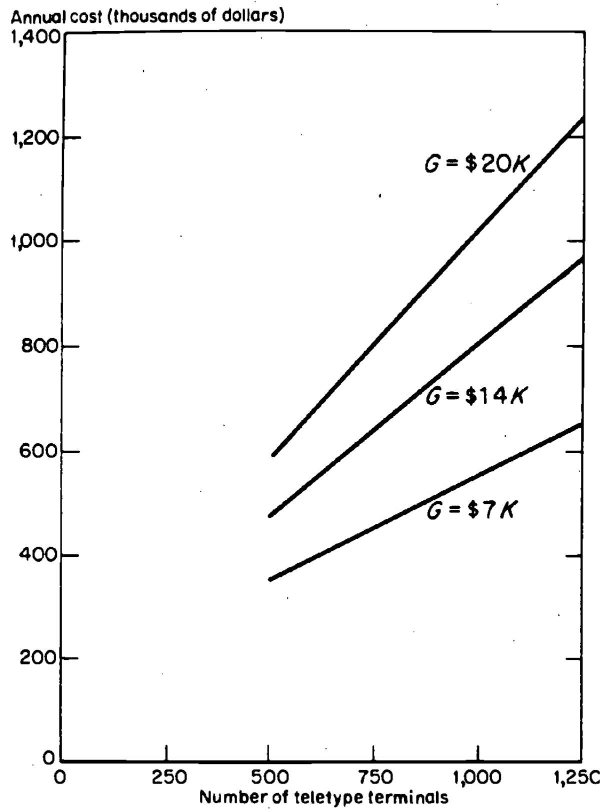

FIGURE IV-1 Annual Communication and Multiplexing Cost, Satellite System

Per Student Costs

To obtain the annual cost of the terminal we multiplied its purchase price ($850) by .15 to obtain $130 and added 10 per cent of its purchase price to cover maintenance. The total is $215 per year. Teacher training must also be included and is typically a one-week course given at the school at a cost of about $500, plus transportation, per person. Continuing our assumption of eight terminals per school, and assuming that the course will be repeated for at least four years and that transportation costs average $300 per session, the per terminal annual charge of teacher training is $25. A final cost to be considered is that of the terminal room proctor. Much of this cost can be covered by volunteers and inexpensive help and would cost not more than $2,000 per school year or $250 per

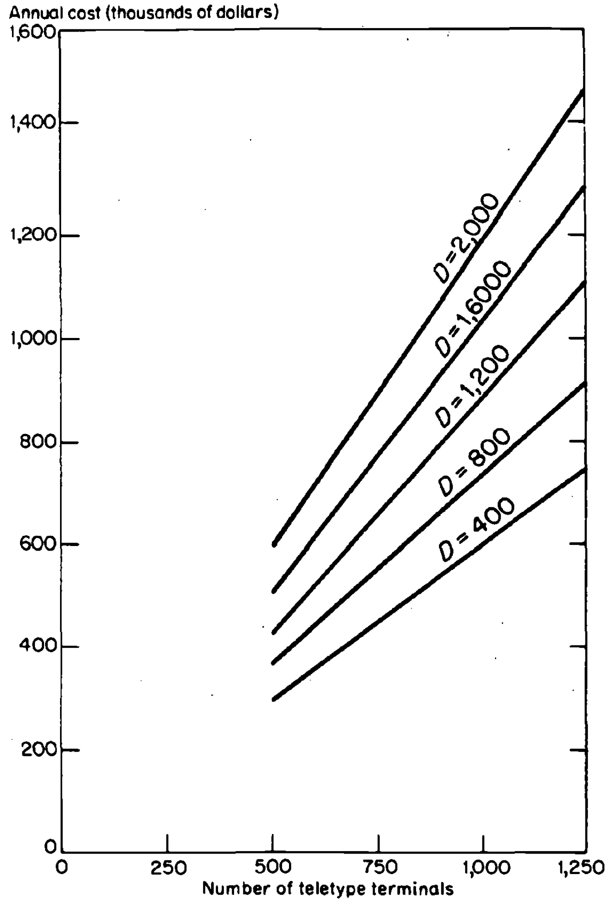

FIGURE IV-2 Annual Communication and Multiplexing Cost, Commercial Telephone System

terminal per year. We assume space available in the schools due to a declining rural population.

Table IV-3 shows the annual costs per terminal. A utilization rate of twenty-five students per terminal per day is typical with this sort of system, so that the cost per student per year would be on the order of $75. Overhead costs might increase this to as much as $125. If the number of terminals per school were increased from eight to ten, there would be no increase in communication and multiplexing, teacher train-

TABLE IV-3 Annual Cost in 1975 of Rural CAI per Terminal

Item	Cost
Teletype terminal	$ 215
Computer facility cost	690
Communication and multiplexing	640
Teacher training	25
Proctoring	250
Supplies and miscellaneous	25
Total	$1,845

ing, or proctoring costs, so our estimates are conservative in that respect.

Kiesling's (1971) estimates for conventional compensatory education at about the quality provided by CAI were $200 to $300 per student per year in urban and suburban areas. It would presumably be more expensive to provide this quality of compensatory education to rural areas, and salary inflation would also increase his estimates. We thus feel that CAI is a low-cost alternative for providing compensatory education to rural areas.

A possible pattern of development for rural compensatory education is to begin with satellite or long-line communications to a large central system, and then, after a cadre of experienced personnel has been trained, to convert to less expensive special-purpose systems located in the area.

C. Opportunity Cost of CAI

In the preceding discussion of cost we estimated ceteris paribus costs of adding CAI to the school curriculum. We indicated that the add-on costs of CAI are sufficiently less than those of alternative compensatory education programs so that, if additional funds were available for compensatory education, CAI appears an attractive alternative. If add-on funds are unavailable—and this is common in the present financial environment—then CAI can be introduced only at the cost of providing less of some other school resource to the students. The amount of the other resources foregone represents, then, the opportunity cost of providing CAI to the school. As teacher costs constitute by far the largest component—on the order of 70 per cent—of school costs, our purpose in this section is to examine what must be given up in terms of teacher resources in order to provide CAI for students.

The amount of teacher time required per child per year depends on average class size, average number of days per school year, and average number of class hours per school day. We assumed that length of school day and length of school year are less flexible than average class size, and examined only the effect on class size of introducing CAI. The other two variables could, however, be introduced into the analysis in a straightforward way.

Let the "instructional" cost per year for a class be the cost of its teacher's salary plus the cost of whatever CAI the class receives. Let S be the class size before CAI is introduced, T be the teacher's annual salary, and C be the cost per student per year of CAI, including all costs previously indicated in Table IV-3. We wish to compute A, the number of additional students in the class that are required to finance the CAI. With no CAI, the annual instructional cost for the class is T; with CAI, the cost is $T + C(S + A)$. We require that the per student cost with CAI be no greater than the cost without it, that is

$$\frac{T}{S} = \frac{T + C(S + A)}{S + A}$$

Solving this equation for A we obtain

$$A = CS^2/T - CS)$$

The partial derivatives of A with respect to T, C, and S are also of interest, and those are given below

$$\frac{\partial A}{\partial C} = TS^2/(T - CS)^2$$

$$\frac{\partial A}{\partial S} = CS(2T - CS)/(T - CS)^2$$

and

$$\frac{\partial A}{\partial T} = -CS^2/(T - CS)^2$$

Table IV-4 below shows A, $\partial A/\partial S$, $\partial A/\partial C$, and $\partial A/\partial T$ for $C = \$50$ (urban) and \$75 (rural) under the assumptions that $T = \$11{,}000$ and $S = 26$.

A number of interesting points emerge from the table. First, even if $C = \$75$, the student-to-teacher ratio only goes from 26 to 31.6 in order to provide CAI. If the Coleman Report is correct in concluding that student performance is insensitive to student-to-teacher ratio, this would seem to be an attractive reallocation to the extent that it can be made politically feasible.[4] Second, from the values for $\partial A/\partial C$ we see that a \$10 increase in C would require about a .8 increase in A if C is \$75. Third, from the value of $\partial A/\partial S$ we see that an increase of 1 in S causes

TABLE IV-4 Increment in Class Size Required to Finance CAI

Variable	Expression[a]	Cost of CAI per Student per Year $50	$75
A	$CS^2/(T - CS)$	3.5	5.6
$\partial A/\partial C$	$TS^2/(T - CS)^2$	.079	.091
$\partial A/\partial S$	$CS(2T - CS)/(T - CS)^2$	.286	.477
$\partial A/\partial T$	$-CS^2/(T - CS)^2$	−.00036	−.00062

[a] S is initial class size and it is assumed to be 26; T is annual teacher salary and it is assumed to be $11,000; C is cost per student per year of CAI, and A is the increment in class size required to finance CAI if there are to be no increases in per student annual costs.

an increase of .286 in A if C = $50 but an increase of .477 if C = $75. Finally, the last row in the table shows that a $1,000 annual increase in teacher salary would decrease A by about .36 if C is $50; it decreases A by almost twice that amount if C is $75. In general, the partial derivatives in the table seem quite sensitive to C.

We conclude this section by observing that the cost of CAI seems to have decreased sufficiently to make CAI quite attractive compared to alternative compensatory techniques with roughly similar performance. This holds whether one considers CAI as an add-on cost or as a substitute for teacher time.

NOTES

1. Sen (1972) criticized the restrictiveness of the additive functional form that Atkinson assumed for determining W. Sen provided a definition of inequality similar to Atkinson's based on a more general functional form. However, the practical usability of the additive form remains a strong argument in its favor.
2. Ball and Jamison (1973) presented updated and more detailed cost estimates for all aspects of a CAI system designed to serve rural populations; their cost estimates differ only a little from the more preliminary ones used here.
3. A further, and very important, advantage of using satellites is that it eliminates the necessity of working with poorly equipped rural telephone services. IMSSS has experienced many delays and unexpected costs as a result of working with such services in Kentucky and elsewhere.
4. Jamison, Suppes, and Wells (1974) reviewed additional evidence that indicates student performance to be insensitive to the student-to-teacher ratio; their review summarized the literature on the effectiveness of various educational technologies as well as various forms of conventional instruction.

REFERENCES

Atkinson, A. B. "On the Measurement of Inequality." *Journal of Economic Theory* 2 (1970): 221–224.

Atkinson, R. C. "Computerized Instruction and the Learning Process." *American Psychologist* 23 (1968): 225–239.

Atkinson, R. C., and Fletcher, J. D. "Teaching Children to Read with a Computer." *Reading Teacher* 25 (1972): 319–327.

Ball, J., and Jamison, D. "Computer-Assisted Instruction for Dispersed Populations: System Cost Models." *Instructional Science* 1 (1973): 469–501.

Boneau, C. A. "The Effects of Violations of Assumptions Underlying the *t*-Test." *Psychological Bulletin* 57 (1960): 49–64.

Bowles, S., and Levin, H. "The Determinants of Scholastic Achievement: An Appraisal of Some Recent Evidence." *Journal of Human Resources* 3 (1968): 3–24.

Chall, J. *Learning to Read: The Great Debate*. New York: McGraw-Hill, 1967.

Coleman, J. *The Evaluation of Equality of Educational Opportunity*. P-3911. Santa Monica, Calif.: Rand Corporation, 1968.

Coleman, J. S.; Campbell, E. Q.; Hobson, C. J.; McPartland, J.; Mood, A. M.; Weinfeld, F. D.; and York, R. L. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Office of Education, 1966.

Elashoff, R. M. "Effects of Errors in Statistical Assumptions." *International Encyclopedia of the Social Sciences*. New York: Macmillan, 1968, Vol. 5, pp. 132–142.

Fletcher, J. D., and Atkinson, R. C. "An Evaluation of the Stanford CAI Program in Initial Reading (Grades K through 3)." *Journal of Educational Psychology* 63 (1972): 597–602.

Jamison, D.; Ball, J.; and Potter, J. "Communication Economics of Computer-Assisted Instruction for Rural Areas: Satellite vs. Commercial Phone Systems." Presented at the 22nd International Astronautical Congress of the International Astronautical Federation, Brussels, September, 1971.

Jamison, D.; Suppes, P.; and Butler, C. "Estimated Costs of Computer-Assisted Instruction for Compensatory Education in Urban Areas." *Educational Technology* 10 (1970): 49–57.

Jamison, D.; Suppes, P.; and Wells, S. "The Effectiveness of Alternative Instructional Media: A Survey." *Review of Educational Research* 44 (1974): 1–67.

Jensen, A. "How Much Can We Boost I.Q. and Scholastic Achievement?" *Harvard Educational Review* 39 (1969):1–123.

Kelley, T. L.; Madden, R.; Gardner, E. G.; and Rudman, H. C. *Stanford Achievement Test: Directions for Administering*. New York: Harcourt Brace Jovanovich, 1964.

Kiesling, H. *Input and Output in California Compensatory Education Projects*. R-781-CC/RL. Santa Monica, Calif.: Rand Corporation, 1971.

Lekan, H. A., ed. *Index to Computer-Assisted Instruction*. 3rd ed. New York: Harcourt Brace Jovanovich, 1971.

Levien, R. *The Emerging Technology: Instructional Uses of the Computer in Higher Education*. New York: McGraw-Hill, 1972.

Levine, D. B., and Singer, N. M. "The Mathematical Relation between the Income Density Function and the Measurement of Income Inequality." *Econometrica* 38 (1970): 324–330.

Lorton, P., and Slimick, J. "Computer-Based Instruction in Computer Programming—a Symbol Manipulation–List Processing Approach." *Proceedings of the American Federation of Information Processing Societies 1969 Fall Joint Computer Conference* 33 (1969): 535–543.

Michelson, S. "The Data: Equal Protection and School Resources." *Inequality in Education* 2 (1970): 4–16.

———. "The Political Economy of Public School Inequalities." In M. Carnoy, ed., *The Political Economy of Education: New Approaches to Old Problems*, forthcoming.

Mood, A., ed. *Do teachers make a difference?* Washington, D.C.: GPO, 1970.

Newbery, D. "A Theorem on the Measurement of Inequality." *Journal of Economic Theory* 2 (1970): 264–266.

Pfanzagl, J. "A General Theory of Measurement—Applications to Utility." *Naval Research Logistics Quarterly* 6 (1959): 283–294. Reprinted in R. Luce, R. Bush, and E. Galanter, eds., *Readings in Mathematical Psychology*, Vol. II. New York: John Wiley, 1965.

Piccariello, H. "Evaluation of Title I." Unpublished paper, 1969.

Sen, A. "On Economic Inequality." The Radcliffe Lectures at Warwick University, draft, 1972.

Stifle, J. *The PLATO IV Architecture*. Technical Report No. X-20, May, 1972. University of Illinois-Urbana, Computer-Based Education Research Laboratory.

Suppes, P., and Morningstar, M. "Computer-Assisted Instruction." *Science* 166 (1969): 343–350.

———. "Technological Innovations: Computer-Assisted Instruction and Compensatory Education." In F. Korten, S. Cook, and J. Lacey, eds., *Psychology and the Problem of Society*. Washington, D.C.: American Psychological Association, 1970.

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

Wells, S.; Whelchel, B.; and Jamison, D. "The Impact of Varying Levels of Computer-Assisted Instruction on the Academic Performance of Disadvantaged Students." Technical Report, Stanford University, Institute for Mathematical Studies in the Social Sciences, 1973.

Zinn, K. L. *An Evaluative Review of Uses of Computers in Instruction*. Final Report, U.S.O.E. Contract No.: OEC-5-9-32-509-0032, December, 1970, University of Michigan, Center for Research on Learning and Teaching.

5 COMMENTS

Allen C. Kelley
Duke University

For more than a decade Patrick Suppes and his associates at Stanford have been pioneering the application of computer technology to elementary and high school education. Their imaginative and ambitious projects have included basic research in learning theory, the formulation and evaluation of programmed-instruction techniques, and the development of computer and communications systems for disseminating programmed materials. Thousands of students across the United States have been "plugged into" the Stanford CAI facility; moreover, recent advances in satellite communications technology have greatly expanded the geographic dispersion of their instructional programs. Based on the extensive data resulting from these applications, it has been demonstrated that CAI and programmed instruction are effective, under selected circumstances, in increasing student achievement; furthermore, CAI is technologically feasible in a wide range of educational settings. However, resources are scarce. Technological feasibility is only one of the conditions necessary for the utilization of CAI; this technique must also be shown to be economically efficient. The present paper by Jamison, Fletcher, Suppes, and Atkinson represents one of several recent studies which examines the economic viability of CAI.

The paper before us presents a benefit-cost analysis of CAI in three subject-matter areas—arithmetic, reading, and computer programming—with particular attention to the role of CAI in compensatory education. Several notable conclusions are reached:

1. The CAI programs have both a statistically significant and quantitatively large impact on increases in student achievement.
2. The program's benefits appear greatest at relatively low achievement levels.
3. CAI tends to narrow the inequality in the distribution of educational outputs.
4. CAI is economically efficient in selected settings, e.g., compensatory education in rural schools.

Given the high failure rate of the compensatory-education programs in the United States, the CAI system reported in the present paper might be considered as a striking success story. Not only are extremely impressive gains in student achievement shown to result from the CAI format—about a 50 per cent increase in output, but these impressive benefits are obtained at only 30 per cent of the cost of even the most successful of the currently operational programs. It sounds too good to be true; unfortunately, it probably is.

For several reasons, we must be somewhat cautious in accepting without qualification the conclusions of this study. Consider first the results showing a positive impact of CAI on student achievement. The critical question in interpreting this result relates to identifying what the experiment measures; that is, what the control group represents. In the arithmetic and reading programs, students in the control and experimental groups obtained basic instruction in concepts, application, and drill from their regular classroom teachers. In addition, those in the experimental group participated in a CAI curriculum, which, in arithmetic, emphasized drill; and which, in reading, focused on phonics. The CAI curricula were a supplement to the regular classroom arithmetic and reading programs. The impact of the CAI programs, then, must be considered *relative* to the activities of the control group during the period when the experimental students were engaged in CAI. The net value of the output of CAI is thus the difference between the *value* of the

CAI output and the output obtained by students in the control group when the experiment was taking place. To the extent that the value of the control group's time was positive, then the statistical results are an overstatement of the impact of CAI on the economic value of the student's total instruction. Indeed, it is possible that if the quantity or price of the forgone output was high enough, the statistically significant positive CAI impact could be translated into a decrease in the economic value of educational output. I realize that designing and implementing an experiment which takes into account the opportunity cost of the experimental group's time is difficult. On the other hand, the benefits of this modification in experimental design could be great. For example, if an experiment were implemented where matched pairs of students were assigned to similar activities at similar times—one with CAI mathematics drill, another with nondirected mathematics drill from, say, a programmed text—we would then be in a position to interpret the economic, as distinct from only the statistical, significance of the results.

Another qualification to the results relates to the assignment of control and experimental groups. The authors report neither the assignment procedures nor do they report a test of the homogeneity of student attributes. While a random selection of control and experimental groups is not necessary, knowledge of the assignment procedures is required to interpret the results. Similarly, since their evaluation procedures do not generally take into account the impact of CAI for various types of students, it would be appropriate to examine in detail the composition of the classes involved.

From an economics point of view, the most interesting aspect of this paper relates to the distribution of the benefits of CAI. Clearly, an assessment of the economic efficiency of any public good must examine not only the output level but also its distribution. This is because the value of output will typically vary according to who receives the output. Thus, in obtaining the total value of output deriving from an educational technique, it is necessary to examine its distribution and assign or identify values for that output for alternative students. Both of these issues, seldom confronted in the education literature, are recognized by the authors. They are to be commended for stressing its importance. However, their analytical and statistical treatment of the distribution issues requires some qualification.

The main statistical model employed in this study to assess the aggregate impact of CAI is a test on the differences in the mean performance of the "typical" student in the experimental and control groups. This model, unfortunately, conceals any of the distributional effects of CAI. For example, while a positive net increase resulting from CAI was revealed, this may have been the result of positive increases for some students and negative results for others. Paradoxically, the benefits of CAI are on a *priori* grounds attributed to individualization of instruction, yet the statistical models employed for analyzing its impact deal with "average" or "typical" students and abstract from any differences in individual behavior. Regression or other statistical techniques where distributional effects are explicitly considered would appear to be more appropriate to the basic research question under consideration.

Unfortunately, in the single instance where a regression was run—a stepwise regression for the computing course, the theoretical model underlying the statistical exercise was misspecified. The regression assumes that the distribution of educational achievement is invariant to the use of CAI—that is, all interaction terms between CAI and student attributes are omitted. This model is not only inconsistent with learning theory and a substantial empirical literature in education, but also with the results of the authors in previous studies. While I suspect that the authors do not believe that CAI is an equally beneficial educational technique for all students, the hypotheses they test and the statistical models they employ abstract from the key distributional questions of who gets how much and why.

The primary method employed by the authors to examine the impact of CAI on output distribution is familiar to economists: Lorenz curves and Gini coefficients. However, the results are somewhat difficult to interpret. First, without a specific norm, one cannot determine what constitutes an important change in the Gini coefficients. Second, while the authors conclude that CAI is generally egalitarian in output distribution, I am uneasy about the finding that in the Mississippi arithmetic experiment, CAI *increased* the inequality of output distribution in three out of six cases. Moreover, CAI's distributional impact was negative in the very instance where it had its most dramatic positive effect on aggregate student achievement—the Mississippi first graders. Third, even though a comparison of the CAI and control groups shows that the direction of change toward egalitarianism was statistically significant, it should be underscored that this result is specific to the output measures of the study—arithmetic drill skills and phonics. But if we assume that the opportunity cost to the students engaged in CAI was nonzero, then clearly a complete assessment of the distributional impact of CAI would appropriately employ more comprehensive output measures, including those specific study areas relevant to the control group during the period of CAI instruction. Finally, and most important, Gini coefficients and the other measures of output distribution employed in this study are not particularly interesting from either an educational or a policy point of view. They obscure the nature of the relevant redistribution taking place. For example, a zero Gini coefficient difference, given the results presented in this paper, is fully consistent with a simultaneous reallocation of educational benefits from girls to boys and from high achievers to low achievers.

While I have qualms regarding the statistical and theoretical analysis underlying the authors' examination of the distributional impact of CAI, I do support their concern with this issue. Their interest in output distribution, as well as that of government officials and the general public, is based primarily on normative grounds. As an economist interested in issues of economic efficiency, however, I am uneasy about promoting CAI or any educational technique as a *means* for redistributing output. Put differently, CAI may or may not represent an efficient technique for redistributing output, even given preferences for such a redistribution. For example, to the extent that different students respond differently to alternative production techniques—a widespread empirical finding in the education literature—then the most efficient redistribution procedure may be one which uses several different production techniques and allocates resources among them so as to maximize output subject to a distributional goal. CAI used mainly for compensatory education, given the evidence provided, may in some settings be relatively efficient both in terms of total output and its distribution. On the other hand, we must stress again that distributional goals are also multidimensional. We may wish to redistribute over some student attributes but not others. Thus, the use of CAI for all students requiring compensatory education may be inefficient, even if CAI produces greater output and allocates this toward bright youngsters who have been low achievers, if at the same time CAI reallocates output from girls to boys (probably an unpopular outcome). Again, at the empirical level, this argues for a research design and statistical models which highlight differences in individual learning styles.

Finally, a temporal dimension might also be employed in assessing the efficiency of output redistribution. While an educational technique may distribute more output to low-achieving students, it is also relevant to identify whether this output so redistributed is retained. If, for example, retention rates are lower for low-achieving students, especially in areas such as recall and recognition where the CAI drill programs appear most effective, and if *retained* value-added is a primary objective in the education industry, then we must include in our distribution analysis not only normative judgments but also long-run technological possibilities.

The shortest section of this study, and of my comments, applies to the cost analysis of CAI. The authors have examined the cost per student of *providing* the CAI curriculum to the classroom, i.e., computer and communications installation, maintenance, and operation costs. Even though they focus on the rural school setting, I would suggest that their assumptions must be considered as generally providing a lower bound on CAI costs. While I shall not examine these costs in detail, I would urge the authors to examine and present a sensitivity analysis of the cost figures to alternative key assumptions. They assume, for example, that the terminals and central processing units are highly utilized, that floor space and rooms in schools devoted to CAI have no opportunity cost, that CAI curriculum-development costs should be excluded, that costs for CAI proctoring are low or zero through the use of volunteers, and that the opportunity cost of the student's time engaged in CAI is low or zero. With these assumptions they arrive at a per student cost of $75 which compares favorably with Kiesling's study showing per student costs of comparable compensatory-education programs ranging between $200 and $300. For at least two reasons, this comparison must be qualified. First, Kiesling's cost figures are more comprehensive than those applied to CAI. Second, Kiesling's cost function varies according to the amount of output produced. If, for the several reasons discussed above, we conclude that the CAI output estimates represent an upper bound on likely long-run production possibilities of this technique, then the appropriate Kiesling cost figures must be adjusted downward.

The paper concludes with an interesting discussion of the financial feasibility of CAI in compensatory education. Recognizing that schools are un-likely to increase significantly their budgets for compensatory education, the authors compute the increase in the student-teacher ratio required to implement a CAI program without any increase in the total per pupil cost of education. Using their cost figure of $75, they find that class sizes averaging 26 increase "only" 20 per cent. If overhead is included in their CAI calculations, the class-size increase would be 40 per cent. They conclude that an increase in class size would seem to be a "quite attractive reallocation" under the assumption that student performance is insensitive to the student-to-teacher ratio. The latter assumption is critical. On the one hand, the authors attribute to individualized instruction the impressive increase in arithmetic and reading skills provided by a short ten-to-twenty minute daily contact with CAI. On the other hand, they are willing to assume, on the basis of little evidence, that a reduction in individualized instruction through a 20 per cent to 40 per cent increase in class size for the remaining five and one-half hours of instruction per day will have a negligible impact on the size and distribution of educational output. If they are right that individualized instruction provided by small class sizes in the compensatory-education setting is unimportant—and this is a testable hypothesis—then we have a long way to go to explain why the brief exposure to the computer programs is able to bestow such significant achievements through individualized instruction. Short of such an explanation, and given the nagging possibility of Hawthorne effects, we must defer our conclusion that CAI is presently, or will be in the future, an economically viable production technique.

David E. Wiley

University of Chicago

The paper by Jamison, Fletcher, Suppes, and Atkinson has three major parts: one devoted to a description of the technology, the second to an evaluation of the performance of this technology, and the last, to an assessment of its costs of application. The second section is divided into two parts; one concerning differences in level of performance for computer-assisted and conventional instruction; and the other, differences in the distribution of performances.

SOME COMMENTS ON THE ASSESSMENT OF DIFFERENTIAL PERFORMANCE

1. Mathematics

The experiment was performed in two states (California and Mississippi) in all six elementary grades. In Mississippi, twelve schools were used (8, 3, 1);

while in California, 7 were used (2, 2, 3). The gain in grade equivalents for each treatment was compared for all six grades in both California and Mississippi. Because the assignment of treatments to schools was not random, and because the process of student selection for participation in CAI in a school is not clear, some care must be taken in the interpretation of the results. The assignment and selection process has seemingly produced some differences in the previous performances of the CAI and conventional groups. The use of gain scores (in grade equivalent units) was an attempt to eliminate some of these initial differences and to increase the sensitivity of the hypothesis test. The appropriate question relates to differences between the conditional distributions of final performance for various levels of initial performance. The gain score analysis assumes that the mean of the conditional distribution is a linear function of initial status and that the slope of the regression line is one. Both assumptions seem unreasonable for these data. The authors note that the gains in Mississippi were greater than those in California. They suggest that gains may be greater for CAI when pupils are below grade level. This would imply a nonlinear relation. Also, the way in which the CAI treatments allocate resources to pupils (more resources are allocated to slow learners in an attempt to achieve more equal outcomes) would presumably make the relation between initial and final performance smaller than for conventional instruction. The metric problem is also rather bothersome. It is not clear that equal amounts of instructional effort will produce equal amounts of gain in grade-equivalent units. This problem is distinct from the one mentioned above, since the treatments are not scaled in resource or effort units. Also the grade equivalents are based on typical performances of individuals in various grades and do not represent equal amounts of learning. Finally, the utility of various final performances would presumably vary, depending on the objectives of instruction. The problem of measurement error is also crucial for an appropriate treatment of this problem.

2. Reading

The reading study used matched pairs of individuals, one of whom had received CAI; and the other, conventional instruction. Since the CAI pupils came from two schools and the conventional pupils came from two different schools, if the school averages are different, the measurement errors in the initial performance levels have different biases for each member of the pair. Since the measurement errors are large for these achievement tests, large differences in school means could produce very different results when the errors in these variables are taken into account.

3. Computer Programming

The interesting thing about the results for computer programming is the discrepancy between the results based on the gain scores and those based on the regression analysis. The regression coefficient for the sum of the pretest scores in the gain analysis is −.26 with a standard error of .14, and not zero, as is assumed in the original analysis. Also the verbal score has a significant coefficient indicating that the groups differed in verbal ability initially. None of the other coefficients is significantly different from zero. This indicates that when initial differences are taken into account there is no detectable effect of CAI (coefficient is −.99 with a standard error of .96).

COMMENTS ON THE ANALYSIS OF DIFFERENCES IN THE DISTRIBUTION OF ACHIEVEMENT

There are conceptual problems in using test scores as quantities. The total number of items on a test is fixed in any application and the score is probably better viewed as equivalent to the proportion of items correct rather than a count variable. This is especially true in using standardized achievement tests, because the items are selected so that about half the respondents will select the correct alternative for each item. These characteristics make the relations among test scores nonlinear. Recent work in test theory has produced rather strong models for test scores, which indicate that transformations of raw scores are necessary to produce variables that have appropriate ratio scale properties. It would be useful to explore the invariance of the Gini coefficients and the Atkinson models with respect to these transformations. The raw test-score metric has some characteristics which might cause some problems with these procedures. If one of the groups is relatively close in average value to the ceiling of the test, the dispersion of that group must be less than that of the other group. This linkage between mean level and dispersion may often be removed by an appropriate transformation. These problems are related to, but are conceptually distinct from, problems of invariance with respect to classes of utility functions.

Some problems exist with respect to the sign tests performed in comparing the Gini coefficients for CAI and conventional instruction. In Table III-8 the two grade-placement differences for the computer programming experiment are correlated. This violates the assumptions of the sign test. Note that all differences are correlated in Table III-9.