DOCUMENT RESUME

ED 055 764 24 RE 003 912

AUTHOR Strang, Harold R.
TITLE An Automated Audio-Visual Approach to Remediate Reading Problems. Final Report.
INSTITUTION Virginia Univ., Charlottesville.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
BUREAU NO BR-0-C-020
PUB DATE 15 Jul 71
GRANT OEG-3-70-0007(010)
NOTE 36p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Autoinstructional Programs; Behavioral Objectives; *Disadvantaged Youth; *Grade 6; *Programed Tutoring; Reading Comprehension; Reading Difficulty; Reading Improvement; Reading Programs; *Reading Research; Rewards

ABSTRACT

Three experiments conducted over a 2-year period with disadvantaged sixth graders are reported. In Experiment 1, the effectiveness of exposure to graded reading materials and audiovisual tutoring was tested in an automated reading program with specific behavioral objectives and a self-contained reward system. Twenty-one students formed three groups of equal reading proficiency. The group receiving specific audiovisual tutoring showed substantial gains in reading accuracy over the groups receiving trial-and-error training in reading and in mathematics. Tutored students' gains were highest on successive comprehension and untimed standardized tests. During the second year, in Experiment 2, the audiovisual tutoring was administered to 19 students, and gains were compared to those students not receiving any machine instruction. Improvement in tutored students paralleled that of Experiment 1. In Experiment 3, students participating in Experiment 2 intermittently received automated instruction on several everyday life reading skills. The tutored students' improvements across the year were significantly greater than those of students receiving no tutoring. Tutored students also showed some generalization to skills not included in the automated tutoring. included. (Author/AW)

O-C-020

OE

PA 24

JUL 2 1971

Final Report

Grant No. OEG-3-70-0007 (010)

Harold R. Strang
Department of Foundations of Education
Peabody Hall
University of Virginia
Charlottesville, Virginia 22903

AN AUTOMATED AUDIO-VISUAL APPROACH TO REMEDIATE READING PROBLEMS

July 15, 1971

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

Region III

1/2

Abstract

The reported research comprises three experiments conducted over a two-year period with sixth grade disadvantaged children. In Experiment I the effectiveness of exposure to graded reading materials and audio-visual tutoring was tested in an automated reading program possessing specific behavioral objectives and a self-contained reward system. Twenty-one students were assigned so as to form three groups equated on reading proficiency. The group that received specific audio-visual tutoring on graded reproductive, interpretive, and successive comprehension readings showed substantial gains in reading accuracy over the groups receiving trial-and-error training in reading and in math. Tutored students' gains were highest on successive comprehension and untimed standardized test measures. During the second year Experiment 2 consisted of administering the audio-visual tutoring to a larger number of students (N = 19) and contrasting gains recorded by these students to those of students not receiving any machine instruction. Improvement in tutored students paralleled that of Experiment 1. Experiment 3 also was conducted during the second year. Those students participating in Experiment 2 intermittently received automated instruction on a number of reading skills used in everyday life (e.g., using telephone directories, using mail order catalogs, and reading newspapers and airline schedules). The tutored students' improvements across the year were significantly greater than those recorded by students receiving no tutoring. Also, tutored students showed some generalization to skills not included in the automated tutoring.

3

Final Report

Grant No. OEG-3-70-0007 (010)

AN AUTOMATED AUDIO-VISUAL APPROACH TO REMEDIATE READING PROBLEMS

Harold R. Strang
Department of Foundations of Education
Peabody Hall
University of Virginia
Charlottesville, Virginia 2__03

July 15, 1971

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education

Region III

Preface

The research herein reported was made possible through the efforts of the following people: William C Owen, Pamela Johnson, Jean Piquet, Georgeanna Bond, Terry Connolly, and Patricia S. Tabb. I would like to thank these people, along with Dr. William C. Berry, Assistant Superintendent for Curriculum and Instruction for the Charlottesville, Virginia, public schools, and Mrs. Pauline Garrett, Principal of the Jefferson Elementary School in Charlottesville, for their assistance.

Finally, I would like to thank my wife, Alice, for the many hours she contributed both in the operation of the project and in the preparation of this report.

Harold R. Strang

Table of Contents

List of Tables

List of Figures

Background

Running counter to the more global approaches to reading research has been an attempt by a number of investigators to delineate and test specific variables relating to reading success. As Katz and Deutsch (1967) have stressed, "Insufficient attention has been devoted to two issues of major theoretical importance: (a) the relationships between audition and vision in various skills, and (b) the developmental aspects of such modality functioning."

Several programs have attempted to automate the administration of audio and visual teaching aids. The Electronic Program Laboratory in Millis, Massachusetts (1966), planned a program based on the auditory mode. In the Tampa, Florida, public schools, Malpass (1967) developed a visually oriented teaching procedure whereby training in word recognition and paragraph reading were implemented through the use of rear-screen film strip projections. A remedial program developed by Gibson and Richards (1965) included both auditory and visual training components. A mobile laboratory visited the participating public schools and offered under-achieving youngsters a variety of audio-visual instructional aids including film strips, sound films, and Language Master cards.

Turning to a more sophisticated audio-visual approach, Atkinson (1968) developed a computer-assisted program wherein students received both tutoring as well as drill-and-practice in math and reading.

The tutor incorporated into the current studies also included automated audio-visual aids. Its approach, however, was somewhat different from the main impetus of former approaches, in that it attempted not only to present the audio-visual help but to present it in a way that would approximate what a live tutor would do. The stimulus for this direction was furnished by the impressive findings of Staats et al. (1967) who pursued the notion of establishing the one-to-one student-teacher model in training reading skills in children showing gross reading deficiencies.

In addition to concern over establishing an individualized tutorial function, an effort was made to develop the tutor's programs around specific behavioral objectives such as Popham (1970) suggested.

Finally, the work of Phillips (1968) and Wolf, Giles, and Hall (1968) suggests that extrinsic reward systems are highly applicable in motivating poverty children. Incorporated within the audio-visual tutor was a token reward system linked to extrinsic back-up rewards.

Strang and Wolf (1971), using a tutor similar in function to the one cited in the current studies, effected sizeable improvement in a group of sixth grade ghetto children. The current studies attempt to better assess this automated approach by such measures as including a larger number of pertinent improvement measures, a tighter control group, a larger spectrum of readings, and a more efficient administration of the audio-visual help. Also, for the first time the tutor was tested in a public school.

This report represents three experiments conducted at an all sixth grade school located in Charlottesville, Virginia. The first experiment was conducted during the 1969-1970 school year. Experiments 2 and 3 were conducted during the 1970-1971 school year.

Experiment 1

Objectives

The specific research objectives of the first experiment were to test the effectiveness of an automated instructional procedure in increasing students' accuracy in answering:

(1) constructed reading comprehension questions relating to recognizing the sequence of events in, reproducing facts from, and making interpretations from graded reading passages;

(2) standardized reading test items.

This automated procedure included exposure to graded reading materials, audio-visual applications of tutoring whenever necessary, and extrinsic rewards for accuracy. The group receiving these training procedures will hereafter be referred to as Group A.

To test whether students would benefit without the aid of the audio-visual tutoring, a control group was constructed (Group B). Students in this trial-and-error group received the automated exposure to the graded reading materials and extrinsic rewards for accuracy but did not receive the audio-visual tutoring.

To test whether students would show improvement on reading measures merely by being actively included in the study but not receiving exposure to any of the graded reading material, a second control group was established (Group C). Students in this group received automated exposure to math instead of reading materials, received no audio-visual tutoring, but earned extrinsic rewards for performance.

To test whether students would show improvement without the aid of extrinsic rewards, a control group receiving no extrinsic rewards would have had to have been constructed. Since the research was conducted at one location, it was deemed impractical to construct such a control group.

Equipment

The objectives were realized through the use of audio-visual tutors completely automating both the testing and the remediation phases of the program.

Three tutoring consoles were located on a table 120 inches long and were separated one from the other by wood partitions. During testing and training, a student was seated in front of a console so that he faced a 10" x 13" plexiglass screen. A rear-mounted projector presented visual materials while a tape recorder presented the auditory elements. These functions were governed by the operation of a relay rack located at the rear of each console.

Each console was also equipped with an electrical answering panel so that students could register one of four possible answers into the system.

The awarding of reward points for the student's performance was also automated. A digital counter located at the right of the view-screen gave the student a constant appraisal of the number of points he had earned during that session.

The apparatus automatically recorded both student accuracy and time.

Subjects

The automated reading program was conducted in a classroom of the Jefferson Elementary School, Charlottesville, Virginia. Throughout the school year students came daily at an assigned time from their regular classrooms to the reading program. Students came in groups of three to one of the seven 50 minute sessions that ran each day. Every attempt was made to draw students during their language arts periods.

At the onset of the program, a group of sixth grade students were selected on the bases of poverty backgrounds, I.Q. scores, and teacher recommendations. Specifically, the type of student sought was the underachiever who was reading far below grade level and yet showed academic potential as judged by I.Q. scores and teacher reports. Final selection was based on the results of a battery of reading tests administered to these students. The tests included a 100-question fourth grade level constructed reading test, Part G of Form W of the elementary level California Achievement Reading Test, and Form A of the reading portion of the Sequential Test of Educational Progress. Also, Form Am of the Intermediate Reading Test portion of the Metropolitan Achievement Test was administered. During testing, rewards were used to maximize performance.

Three groups of seven students each were then constructed. To establish pre-instruction equality, students were assigned so that the resulting groups possessed similar mean scores on the 100-question constructed fourth grade reading tests and on the standardized reading tests. The groups were then randomly assigned to the A, B, or C condition (see Table 1).

Neither sex nor race were criteria for original student selection or for grouping. The boy-girl ratio in Groups A, B, and C was 5 : 1, 4 : 3, and 3 : 3 respectively. The Negro-white ratio in Groups A, B, and C was 2 : 4, 5 : 2, and 3 : 4 respectively.

Procedures

All students proceeded through five sequential programs, each of which consisted of a pretest, a period of instruction, and a posttest. While all groups received identical testing, instruction differed for each of the three groups.

During the reading pretest, all students were presented with 32 projected frames, each consisting of several paragraphs of reading followed by a multiple choice question pertaining to that passage. The questions were constructed so that there were four alternative answers, and the student, by differentially touching his answer stylus to one of the four answer locations on the console, selected the one he deemed to be correct. On all pretests as well as posttests, the programming equipment set a minimum of 30 seconds' exposure to each frame before an answer could be made.

While all questions related to the student's comprehension of the reading passages, three specific comprehension skills were tested. One type required the student to reproduce facts from the passage (reproductive comprehension). A second tested the student's ability to discern the sequence of events in the reading passage (successive comprehension). The third required the student to extract the main idea or to make interpretations regarding the text (interpretive comprehension).

The student progressed through the testing phase at his own pace. Immediately following each answer, the student was exposed to the next reading passage for 30 seconds before having the opportunity to answer the accompanying question. During testing, the student was not informed after each answer as to his accuracy. In attempting to maintain attention to the task throughout this phase, however, following every fourth answer, the student received points for his correct answers to the four preceding questions. The student completed a test in two daily sessions.

After testing, the students proceeded to the instruction phase. Here, for the next 13 consecutive 50 minute sessions, students in the A and B reading groups were exposed daily to 12 reading passages with accompanying questions. The 32 pretest items were included in the 156 frame program. All textual materials were taken from current graded reading workbooks; all questions were constructed.

As in the testing conditions, each text-question frame had a set minimum time delay before an answer could be registered. Unlike the conditions in the test phase, after a student had answered a question, he knew immediately whether or not he had been correct.

If a student in Instruction Group A (the audio-visual tutoring group) answered a question correctly, he was awarded a number of points and then went on to the next text-question frame. If, however, he made an incorrect choice, a light momentarily flashed "wrong," and the automatic programming equipment immediately administered individual audio-visual tutoring.

This tutoring varied for comprehension types. If an interpretive error were made, those parts of the passage relating to making a correct response were underlined in red, and a tape track administered parallel verbal help. If a reproductive error were made, tutoring consisted of the presentation of a second slide of the same reading passage with specific areas of the text relating to the question underlined. And if a successive error were made, a tape track gave specific help relating to that question.

After tutoring, if the student made a correct response, he received a lesser number of points than he would have received had he not needed tutoring. Then he proceeded to the next text-question frame. In this manner, the students in Group A proceeded through 12 text-question frames during each of the 13 days of instruction.

During instruction, the students in instruction Group B (the trial-and-error reading group) received the same 156 text-question frames as those students in instruction Group A. The reward system was also identical. The after-error condition, however, differed. As in the A condition, if a student committed an error, he immediately recognized this, since he received no points, a "wrong" light panel flashed, and he did not advance to his next text-question frame. Other than this, no help was given. During the time in which A students would be receiving audio-visual tutoring, B students received a comparable imposed exposure to the missed text-question frame. If subsequent errors were made to the same question, the above procedure was repeated. As in the A condition, a correct answer advanced the student to the next text-question frame.

During the instruction phase, the students in Group C (the math control group) were not given any programmed materials relating to reading. Instead, they received machine training on a series of mathematics problems. The mode of presentation and the reward system paralleled that of the reading instruction B condition.

After the 13 day program phase, all students received the 32-frame reading posttest. The text portions of the frames were identical to those of the pretest. The question portions, however, while requiring the same information for correct answering, were rephrased, and the alternative answers were often reworded and sequenced differently from those of the pretest.

This program emphasized accuracy rather than speed in reading, in that students were rewarded in tests as well as during the program for accuracy and not for rate of correct answering. A complete record, however, including a daily account of the time that each student spent in all aspects of the program, was kept.

Rewards

Based on the findings of Wolf, Giles, and Hall (1967), the study employed a system of both short-term and long-term extrinsic rewards. The points displayed on the digital counter were awarded for correct responding, the greatest "pay-off" being given for lengthy chains of correct answers. At the end of the day, if a student had accumulated enough points, he could redeem them for short-term rewards such as potato chips, candy, or peanuts. Or, if he wished, he could save his points until he had accumulated enough to buy small, inexpensive items of his own choosing at the local stores.

Results

Gross measures of materials covered and amounts of time spent. The A and B students proceeded through five reading programs in which 780 reading passages totaling approximately 179,244 words were presented.

During this time, C students received 780 problems consisting of 56,160 addition or subtraction computations.

In Group A an average of 42.1 hours per student were spent in reading training. Assuming 1,000 minutes equal one month of program time, the Group A students averaged 2.53 months of training.

Group B spent an average of 44.3 hours (or 2.7 months) in reading training.

Control math students spent an average 35.05 hours in both the testing and training phases of the program.

Immediate training effects of the A and B conditions. One important measure across the A and B conditions was the immediate effect of the tutoring in aiding the student to correct missed questions. A ratio was derived by comparing the number of errors committed after an initial error (subsequent errors) to the number of initial errors. Figure 1 depicts these ratios across the five programs.

For all programs, the mean subsequent error - first error ratio was substantially lower for the A students, indicating that the tutoring given to the A group was consistently more effective in eliciting a subsequent correct answer than was the simple trial-and-error training given to the B group.

In viewing subsequent error - first error ratios across individual students, of the six A students, five had lower ratios than any of the B students.

The effect of tutoring upon answering new questions. A broader intra-program measure of difference in the A and B students' reading improvement was found by continually comparing the frequencies across programs of correct answering to questions upon their initial presentation. For the A students the mean total number of these initial correct answers increased steadily across the five 156 question programs, even though the grade levels increased from fourth to fifth and the average number of words per passage increased from program blocks I - II to IV - V by 7.4%. The 91.8 average of Programs I - II rose significantly to 113.9 in Programs IV - V ($T = 0$, $N = 6$, $p < .025$).* Initial mean gains of the B students tapered off as materials became more difficult. On Programs I - II B students averaged 87; on Programs IV - V, 92.6 (see Figure 2).

In viewing improvement in individual students on the correct answering measure, all six A students showed increases in accuracy across program blocks I - II to IV - V.

The B students showed a more diffuse pattern. Four B students showed an increase from Programs I - II to IV - V, while three showed a decrease.

An analysis of the students' improvement in the correct answering measure across types of comprehension included in the training revealed that A students made significant mean gains from Programs I - II to IV - V in all three comprehension categories.

---

*Due to the inability to meet assumptions relating to the use of parametric statistics, the Wilcoxon Rank Sum Test was used for all cross conditions, whereas the Wilcoxon Signed Rank was used for intra-condition changes, unless otherwise cited. (F. Wilcoxon and R. Wilcox, Some Rapid Approximate Statistical Procedures, 1964)

Almost all A students reflected gains across the three comprehension types. While B students, too, showed mean gains across the three comprehension categories, their statistically insignificant gains were universally lower, and in each category fewer members of the group showed improvement (see Table 2).

Time spent in completing the programs. A comparison of the time spent in various parts of the training phase of the program revealed large A - B differences.

Directly comparing time measures between the two groups across the programs, for program block I - II both A and B students averaged 2.29 minutes before making answers to new questions. In the IV - V block, however, the A students averaged .42 minutes more before making these initial answers than did the B students (T_u = 57, M = 6, N = 7, p $\angle$.05).

In comparing the average time spent before answering questions after errors had been committed, A students, in program block I - II, averaged .16 minutes more than did B students (T_u = 58, M = 6, N = 7, p $\angle$.05). Viewing program block IV - V, the A students not only spent a significant 1.02 minutes more than the B students before answering (T_u = 58, M = 6, N = 7, p .02), but their increase from blocks I - II to IV - V was in itself significant over that of the B students (T_u = 56, M = 6, N = 7, p $\angle$.05).

Pre-program test results. Analysis of the tests preceding each program revealed differential improvement across the three conditions (see Figure 3).

The A's upward movement in correct responding over the tests was significant over the B - C movement (T_u = 87, M = 6, N = 7, p $\angle$.05).

In addition to these results of average test improvement, analyses of individual students' performance yielded group differences. In Group A five of the six students showed progress from the pre-reading test to Pretest V. Of the seven B students, however, only one showed improvement. Furthermore, all seven C students showed a decrease in test performance on Pretest V.

Changes in the time spent per question across the tests also differed for the three groups. Group A spent 1.93 minutes per question on both tests. Group B dropped from 2.02 minutes per question on the pre-reading test to 1.40 minutes on Pretest V. Group C dropped from 2.13 minutes to 1.35.

Pretest - posttest changes across the conditions. A measure of average intra-program improvement was also derived from test results. Collectively looking at the sums of pretest - posttest improvements across the five programs, A students averaged 16% improvement on posttests; B students averaged 9%, and C students averaged 1%. Not only did the A group achieve the highest pretest - posttest gain in percentage of correct answering, but this improvement was significantly higher than the combined B - C groups' percentage (T_u = 93, M = 6, N = 8, p $\angle$.01).

Standardized test results. The standardized tests used to establish pre-instruction equality among students were also administered at the conclusion of the school year to determine changes in standardized reading test performance. A "carnival wheel" contingency, as described by Strang and Wolf (1971), was used to maximize performance during the end-of-program standardized testing. Here children had the opportunity to work for extra points, snacks at a local drive-in, or pennies and nickels.

The standardized test reading measure was defined as the average of the three standardized reading tests: Part G of Form W of the elementary level California Achievement Reading Test, Form A of the reading portion of the Sequential Test of Educational Progress, and Form Am of the Intermediate Reading Test portion of the Metropolitan Achievement Test (see Table 3).

A second measure utilizing standardized test performance was derived. Throughout the program and unit-by-unit testing, Group A students universally achieved higher accuracy than did B or C students when unlimited amounts of time were allowed for answering. It was therefore decided to measure accuracy across the groups on an untimed standardized test. Since both timed pre-training and post-training scores of the G section of Form W of the California Achievement test had been obtained, an untimed parallel form (Z) of the G section of the CAT was given. During the administration of this Z form, however, no time limit for student completion was set.

Table 4 depicts the three groups' performance on the timed Form W pretest, on the timed Form W posttest, and on the untimed Form Z posttest.

The A and B students showed significantly greater improvement across the timed tests than did the C students ($T_L = 37.5$, $M = 6$, $N = 13$, $p < .05$). The B students' improvement slightly exceeded that of the A students.

When untimed posttest performance was compared with the timed pretest performance, again the A and B students collectively showed significantly higher gains than did the control students ($T_L = 29$, $M = 6$, $N = 13$, $p < .01$). On this measure, however, the A students demonstrated superiority over the B students. A gross perspective as to the grade level changes represented by the raw scores can be realized by noting that each raw score point equals 2 to 3 months.

The average time spent by the A students in completing the untimed test was 40.40 minutes. The B group averaged 28.92 minutes; the C group, 27.02 minutes.

That these group differences in test times were merely reflecting a time variable uncontrolled in original grouping seems unlikely, since, during the constructed 100-question reading test administered at the beginning of the program, B students averaged 6% more time in answering questions than A students, and C students averaged 2% more time in answering than did A students.

Cost of the incentive system. The monetary value of rewards expended in the program was $132.74. The average daily value of rewards earned equaled $.07 per student, ranging from $.03 to $.13 for individual pupils. The specific token redemption patterns of the students are cited by Strang (1971).

Discussion

This project demonstrated that previously non-productive students could be maintained in academic tasks for an entire school year with the aid of extrinsic rewards averaging only $.07 daily per student.

Not only were the students attentive to task, but they also showed real reading improvements, especially if they received audio-visual tutoring. All program-derived measures indicated that by the end of the five graded programs, the tutored students showed improvements in reading accuracy far above the other students. With regard to individual progress, virtually all students receiving the audio-visual tutoring recorded substantial gains in accuracy.

Coupled with this increase in accuracy, the tutored students spent ever-increasing amounts of time both in answering initial program questions and in using the audio-visual tutoring aids. These findings strongly support the conclusion that the tutored students were, indeed, developing more reflective learning tempos characterized by the careful reading of new materials and by a strong receptiveness to remedial aids when administered.

Although not showing the accuracy gains of those students receiving the audio-visual tutoring, the students in the trial-and-error reading group apparently did benefit from their exposure to the programmed materials. Their across-program reading speed increased with no corresponding decrease in accuracy.

Both the tutored and the trial-and-error students, while showing unequal quantitative improvement, did show similar patterns of differential gain in successive, reproductive, and interpretive comprehension. As might be surmised, both groups showed greatest improvement in sequencing facts in readings and least improvement in making interpretations from readings.

Further support for the validity of the intra-program findings was furnished through an analysis of pretesting patterns across the two reading groups. While these results did reflect the same differential patterns of tutored versus trial-and-error improvement, it is interesting to note that a nearly uniform difference in the magnitude of improvement in the two groups unequivocally favored program over test measures. This finding is interpreted as relating to two factors: (1) the students could earn slightly more reward points for program performance; (2) there were differences in the immediacy of reward feedback. During testing, reward feedback was administered after every fourth answer, whereas during the program, reward feedback was administered after every answer. While the literature documents that differential reward and immediacy of feedback variables affect rate changes of simple operants such as bar pressing or marble dropping, it is revealing to find that these variables may also strongly influence the emission of more complex intellectually oriented behaviors such as reproducing facts from a reading or even making interpretations from readings. This issue becomes even more compelling when it is realized that this study recorded consistent 15 to 30 percent decrements in performance on grade equivalent materials when rewards were decreased by only a small fraction and knowledge of rewards earned was delayed for only several minutes.

The control students' progressive decline in accuracy on all program-derived measures clearly depicts a condition in which students, without the benefit of instructional help, show a progressive deterioration in performance as a function of increasingly difficult lessons.

That the differential reading gains of students receiving reading training (Groups A and B) versus math control students resulted from a Hawthorne effect seems unlikely since all students attended the same number of sessions during the year. Control students also received the same rewards for performance. Their program sessions differed only in that while the reading students were exposed to reading lessons, the control students were exposed to math lessons.

Although the results derived from tests and intra-program measures clearly demonstrate the effects of tutoring, these effects relate only to three comprehension behaviors. The question arises as to the audio-visual tutoring versus trial-and-error effects on improvement on more generalized measures of reading, such as standardized reading tests. While both the audio-visual tuto[illegible]ed and the trial-and-error reading groups averaged fewer than 60 class sessions [illegible] the actual program (that part where feedback and/or tutoring was administered), their improvement over the year more than doubled the improvement made by the control students. Unlike the improveme t patterns shown by the constructed program and test measures, however, the trial-and-error students actually achieved an average gain of two months in excess of that made by the tutored students. Inspection of individual tests revealed that the tutored students' slightly inferior test scores related not to the number of errors committed but to the tutored students' lack of completing the test in the stipulated time.

16

When a standardized testing measure was obtained under untimed conditions such as those found during all of the training, the tutored students again displayed superior improvement. Although they did, indeed, take more time to answer questions, this amount of time averaged only 41 seconds per test question over the average length of time a student could spend on each question if he were to finish the test in the allotted 20 minutes.

In viewing the progress made by the tutored and control students in the program, the possible differential influence of regular classroom language arts instruction must be considered. Due to the necessity for establishing student participation times early in the year prior to pretesting and subsequent group matchings, it was decided to schedule all students during their language arts periods whenever possible. Of course, this precluded the possibility of guaranteeing that the math controls could be drawn from classes other than language arts. Thus, the question arises as to whether the lack of gain in control students merely reflects lack of any reading instruction during the school year. Two findings reflect upon this. First, those students in the math group who did miss language arts classes missed only an average of 7 percent of the yearly 500 hours allotted by the school to language arts. Second, a comparison of those control students drawn from language arts to those drawn from classes other than language arts revealed virtually no difference in reading decrements across the constructed pre-reading - 5th level reading tests. All control students showed a loss. In addition, the four students drawn from language arts classes averaged an 18.8 percent loss; the three students drawn from classes other than language arts averaged a 17 percent loss.

It would appear, then, that the brief exclusions from language arts class had little influence upon the performance of the control students on the program-administered reading tests.

Another aspect of classroom time missed in language arts is that those students receiving the automated reading instruction (Groups A and B) accumulated an average loss of only 13 percent. These data, coupled with those relating to the control group, clearly establish a perspective for viewing the extent of the time loss dimension. It is concluded that the tutored students' gain over the control students' was achieved with very little interference to the students' regular classroom instruction in language arts.

The improvements recorded in this study must be attributed to the automated tutors' interaction with the students, since during the periods of tutoring no human intervention was administered. That the impressive gains made by students were not mere statistics is indicated by extremely favorable teacher reports telling of several notable improvements in classroom performance.

Experiment 2

Both Experiments 2 and 3 were carried out simultaneously during the second year. Participating students spent two-thirds of their time in Experiment 2 and one-third of their time in Experiment 3.

Basically, Experiment 2 was a replication of the first experiment, in that its major purpose was to assess the effects of automated audio-visual tutoring upon increasing accuracy on reproductive, interpretive, and successive comprehension questions as well as on standardized reading tests. Several important changes in methodology, however, were initiated in Experiment 2:

(1) As a result of the powerful effect of the audio-visual tutoring on the six students in Experiment 1, it was decided to test this condition's effectiveness on a larger number of students (N = 19). In order to schedule this number of students each day, the trial-and-error group of Experiment 1 was eliminated.

(2) Since control students' virtual lack of improvement during Experiment 1 offered little support for a Hawthorne effect, the five control students in Experiment 2 were exposed only to the testing phases of Experiment 2.

(3) In order that the tutored students could receive more instruction during the year, all tests, except those administered at the beginning and at the end of the school year, were eliminated from Experiment 2. One test which was administered to all students at the beginning and again at the end of the school year was a 45-item constructed test including fourth, fifth, and sixth level interpretive, successive, and reproductive comprehension questions.

(4) Fewer standardized reading tests were administered. Two tests were given both at the beginning and again at the end of Experiment 2. One, the X form of Section G of the California Achievement Reading Test, was untimed so that the students were not restricted by a set time to complete the test. The second test, the Y form of Section G of the California Achievement Reading Test, was timed.

(5) The programs of the previous year also were revised so that students progressed through five 156-lesson programs sequenced in the following manner: Programs I and II consisted of all fourth grade materials; Program III, of a combination of fourth and fifth grade materials; Program IV, of all fifth grade materials; and Program V, of a combination of fifth and sixth grade materials. Regarding the average number of words per story per program, the first fourth level program averaged 220 words; the fifth level program, 246 words; and the combined fifth and sixth, 242 words.

(6) Due to gross performance differences related to immediate versus delayed administration of reward points, all phases of Experiment 2 involving point rewards incorporated the immediate reward procedure as found in the instruction phases of Experiment 1.

Aside from the changes noted above, the methodology of Experiment 2 replicated that of Experiment 1.

Subjects

Participating students were selected on the bases of poverty backgrounds and depressed standardized reading test performance. The pre-instruction test scores are presented in the "Test Results" section of Experiment 2. Of the 19 students randomly assigned to the audio-visual tutoring group, 11 were black, 8 were white; 11 were female, 8 were male. There were 5 control students: 4 were black, 1 was white; 3 were female, 2 were male.

Results

Intra-program changes. Those students receiving the audio-visual tutoring showed several changes in performance across the graded programs. The larger sample size of Experiment 2 allowed for parametric statistical analyses of such changes (see Table 5).

The results of testing and the mean differences of the three significant measures in Table 5 across grade levels appears in Table 6.

While students showed an overall gain in accuracy from the fourth to the combined fifth and sixth grade levels, the greatest accuracy gains were recorded from the fourth to fifth levels with a drop between the fifth and combined fifth and sixth levels.

With regard to time taken before registering an answer, students showed a progressive rise across the three levels.

Students averaged substantially more time to answer tutored questions during the fifth level program than during the fourth. This time measure, however, dropped from the fifth to the combined fifth and sixth levels.

To further evaluate the training effect in the three comprehension categories, individual F-ratios and t values were derived for each category (see Tables 7 and 8).

As reading materials became more difficult, students showed a decrease in accuracy in interpretive comprehension. On reproductive comprehension, significant improvement occurred from the fourth to the fifth level, followed by a decrease in accuracy during exposure to combined fifth and sixth level materials. In successive comprehension, students showed progressive improvement across all three levels.

Test results. In comparing experimental students' and control students' constructed test improvement, experimental students rose from 48.8% correct on the pretest to 72% correct on the posttest ($t = 8.22$, $df = 18$, $p < .01$). Control students rose from 55.6% correct on the pretest to 69.2% on the posttest ($t = 3.57$, $df = 4$, $p < .05$). Directly comparing the gains recorded by both groups, the experimental students' improvement was significant over that of the controls (($T_u = 33$, $M = 5$, $N = 14$, $p < .05$).

Test improvement across the three types of comprehension appears in Table 9.

In one area, successive comprehension, the experimental group's improvement was significantly greater than that of the control group ($T_L = 27$, $M = 5$, $N = 14$, $p < .01$).

The two groups also showed differential improvement across the year on both untimed and timed standardized reading tests. On the untimed Form X of the California Achievement Reading Test, the A students averaged 15.6 correct answers on the 30-item pretest, whereas on the posttest they averaged a significant rise to 19.7 ($t = 3.73$, $df = 18$, $p < .01$). Control students actually registered a drop from 19.4 correct on the pretest to 18.2 on the posttest. Directly comparing the two groups, the A students' change in accuracy across tests differed significantly with the control students' ($T_L = 35.5$, $M = 5$, $N = 14$, $p < .05$).

On the timed Y form of the CAT the A students showed a 2.8 raw score gain ($t = 5.02$, $df = 18$, $p < .01$). This could be translated as a rise from grade 4.6 to 5.7. Control students showed a statistically insignificant 1.2 raw score gain, equivalent to a rise from grade 5.3 to 5.7.

Discussion

Paralleling the findings of Experiment 1, audio-visual tutored students showed greatest improvement across programs in successive comprehension and least in interpretive comprehension. Furthermore, students not only increased in accuracy as the program dvanced from the fourth to fifth grade levels, but they also increased in the time ney spent before registering first answers and in the time they spent per tutoring oplication before answering. These findings agree with those of the first experiment n suggesting that the automated tutoring directed the students toward more reflective earning tempos.

The tutored students' response to the more difficult fifth-sixth level materials f Experiment 2 was to spend more time initially reading questions and to increase rrors in all areas except successive comprehension. This pattern indicated that the rogression from the fifth to the combined fifth and sixth materials was too rapid, nd at least one additional fifth level program would be necessary in future applica-ions.

Turning to pre- - posttest changes, on the constructed reading test t' tutored tudents' statistically superior performance to that of the control students attested o the effects of the automated tutoring in helping the tutored students in reproduc-ive, interpretive, and especially in successive comprehension. Also, tutored students emonstrated substantially greater standardized reading test improvement than did ontrol students, especially on untimed tests. And even on the timed tests, tutored tudents averaged twice the gain recorded by the control group across the school ear. It is unfortunate that the randomized group selection procedures initiated t the beginning of Experiment 2 did not yield two groups having closer pretest cores.

Experiment 3

Experiment 3 concerned the application of automated tutoring in teaching disadvantaged sixth grade students a series of reading skills useful in every day living.

Objectives

Two specific aims of Experiment 3 included:

(1) assessing the role of the automated tutoring in increasing students' accuracy in 24 specific practical skills;

(2) assessing any generalization of the automated tutoring to the acquisition of practical skills not included in the instruction.

Procedures

The tutored and control students included in the study were those described in Experiment 2. The two experiments were run concurrently so that of each three sessions, students spent one participating in practical skills training and the other two in comprehension training. In the practical skills program all students first received practice in using the automated tutors. Next, they were machine tested on 51 different practical reading skills. The control students then were returned to their regular classroom activities while the experimental students received 20 sessions of automated instruction in those 24 skill areas showing the highest error densities on the test. These skills included:

I. the use of reference tools
   A. encyclopedia (4)*
   B. dictionary (3)
   C. atlas and maps (2)
   D. card catalog (2)
   E. tables and graphs (2)

II. newspaper reading
   A. weather maps (1)
   B. want ads (1)

III. reading airline schedules (1)

IV. the use of telephone directories
   A. yellow pages (2)
   B. white pages (2)

V. contents of forms and applications (e.g., catalog forms, checks, bills, job applications) (3)

VI. the use of mail order form catalogs (1)

*Number of specific behavioral objectives per skill area

In summarizing the tutors' instruction functions, the student was first presented with a projected frame including a multiple choice question pertaining to a practical skill (e.g., Above is a check written by Tom Beal. Which of the following items has he filled out incorrectly?). Dependent upon the particular skill, the source of information necessary for successful completion of each question was either presented above the question (as would be the case with the example above), or the student was informed by the tutor that he was to secure the information from a reference source close to the console (e.g., Use the atlas to help you answer this question: Which of the following countries is southeast of Germany?).

Each session consisted of 12 lessons, each representing a different practical skill. Students registered answers directly into the tutor. Correct answering yielded an advance to the next lesson plus several reward points presented on a digital counter within the student's view. Incorrect answering yielded immediate verbal tutoring, presented through each student's headset via a tape recorder. The taped voice led the student through an orderly sequence of steps relating to acquiring that particular skill. Never, however, was an answer given. During this tutoring, the original projected image did not change. Subsequent incorrect answering to the same question was followed by an imposed one minute time exposure before another answer could be made. The apparatus automatically recorded all answer inputs and advanced the student to the next lesson only after a correct answer had been given.

After the 20 instruction sessions, all experimental and control students were retested on the 51 item test. Automated testing was identical to instruction in the administration of rewards and automatic recording. Any answer registered, however, regardless of its correctness, advanced the student to his next question and, if errors were made, no audio tutoring was administered. No questions presented during instruction were included on the test.

Concerning rewards, points could be redeemed daily for low-value extrinsic rewards (e.g., candy, gum, potato chips) or could be saved over time for high-value rewards from local stores (e.g., notebooks, models, jewelry).

Results

The experimental students' improvement during instruction was obtained by comparing error frequencies in the first 24 lessons to those in the last 24. The total error average (including initial errors and errors after tutoring) dropped from 15.5 to 4.0 ($t = 6.68$, $df = 18$, $p < .01$). All but one experimental student registered a drop in total errors. With regard to initial correct answering to questions, experimental students improved from an average of 61% correct on the first 24 lessons to 89% on the last ($t = 8.70$, $df = 18$, $p < .01$). All 19 experimental students registered improvement on this measure.

Pre- - posttest gains. In viewing overall test improvement across the year, experimental students' average of 62% correct on the pretest rose to 85% on the posttest. All 19 experimental students registered gains. The control students' average of 68% correct on the pretest rose to 78%. Of the 5 control students, 4 registered gains. The 23% gain registered by the experimental students was significantly higher than the 10% gain registered by control students ($T_L = 28.5$, $M = 5$, $N = 19$, $p < .02$).

In comparing both groups on gains in the 24 skill areas trained during the program, the experimental group averaged 48% accuracy on the pretest, 81% on the posttest. All experimental students registered gains. The control students averaged 56% correct on the pretest, 70% on the posttest. Of the 5 control students, 4 showed cross-test gains. The 33% gain of the experimental students was significantly higher than the 14% gain of the control students ($T_L = 29.5$, $M = 5$, $N = 19$, $p < .02$).

Viewing the 27 skill areas not trained during the program but included during testing, the experimental group's average of 73% correct on the pretest rose significantly to 88% on the posttest. The control students' average of 78% correct on the pretest rose to 83% correct. Seventeen of the 19 experimental students registered cross-test gains; 2 of the 5 control students registered gains.

While the 15% gain of the experimental students was itself a significant gain ($t = 5.90$, $df = 18$, $p < .01$), it was not found to be significantly higher than the 5% gain averaged by the control students.

Program efficiency: time, monetary investment, and recorded gains. The experimental students averaged 10.6 hours of automated instruction to complete the 240 lessons in the 10 programs. The average pretest - posttest improvement attained by the control group served as a basis for assessing the effects of retesting and normal classroom instruction on acquisition of practical skills. Thus, to obtain a measure of the automated tutoring's effect upon practical skill acquisition, the control group's average cross-test gain was subtracted from the experimental group's cross-test gain. It was found that the 19 experimental students learned a total of 129 practical skills. All but one experimental student showed a corrected gain, and individual students gained as many as 15 new reading skills.

The average cost for the extrinsic rewards used during the tutoring amounted to $.33 per corrected acquired skill.

Discussion

The test-derived data clearly demonstrated the strong effect of the automated tutoring upon the learning of the practical reading skills. Improvement appeared not only as group statistical gains but also as individual gains exhibited by all but one experimental student. As might be expected, the experimental students showed greatest improvement over the control group in skills included in the automated instructional phase. The fact, however, that these students did average a significant improvement of 15% on test items not included in the instruction phase (compared to an insignificant 5% gain made by the control group) strongly suggests that some instructional generalization to the untutored practical reading skills did occur in the experimental group. That 17 of the 19 experimental students registered cross-test gains as high as 56% on items not included in the instructional phase further substantiates this conclusion.

Epilogue

The universality of results across the two years in demonstrating substantial positive changes in the tutored students' reading accuracy attests to the general effectiveness of the automated audio-visual tutor. In viewing the reading areas most affected by the tutor, collective results further affirm that those reading behaviors defined most precisely were most easily trained. In comprehension, the tutoring of the loosely-defined interpretive skills yielded results far inferior to the results of the tutoring of successive skills where students simply had to learn to sequence facts in order of their occurrence in the readings. Even in the practical skills area, tutoring students to find general errors in written forms was far less fruitful than training the same students to find correct volume numbers in locating articles in encyclopedias or in finding how to define the symbols on weather maps in newspapers.

The automated tutor is not an all-encompassing educational panacea that will teach all children any academic material. It is, however, a tool that can help the classroom teacher to do a better job in one area, the instruction of specific educational skills.

<u>Bibliography</u>

Atkinson, R. C., "The Computer Is a Tutor," <u>Psychology Today</u>, <u>1</u>, 1968, pp. 36-39, 57-59.

Gibson, C & Richards, I., <u>Development of Experimental Audio-visual Devices and Materials for Beginning Readers</u>, ERIC Report CRP-E-033, Harvard University, 1965.

Katz, P. & Deutsch, M., <u>Visual and Auditory Efficiency and Its Relationship to Reading in Children</u>, ERIC Report CRP-1099, New York Medical College, 1967.

Malpass, L. <u>et al</u>., <u>Programmed Reading Instruction for Culturally Deprived Slow Learners</u>, ERIC Report BR-6-8438, Tampa, Florida, 1967.

Millis Public School, <u>Electronic Program Laboratory</u> (planning), ERIC Report DPSC-66-892, Millis, Massachusetts, 1966.

Phillips, E. L., "Achievement Place: Token Reinforcement Procedures in a Home-style Rehabilitation Setting for Pre-delinquent Boys," <u>Journal of Applied Behavior Analysis</u>, <u>1</u>, 1968, pp. 213-223.

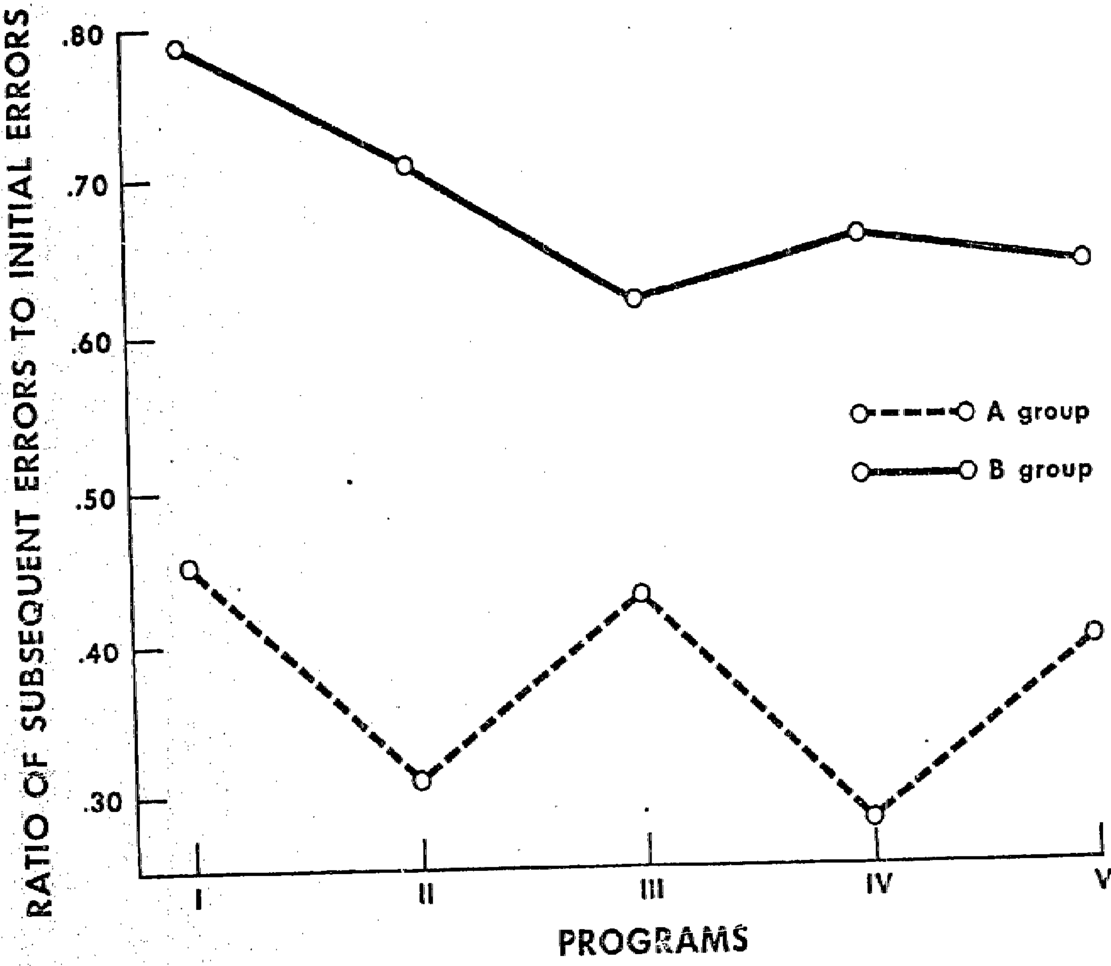
Popham, W. J., <u>Potential Uses of IOX Objectives</u>, UCLA Center for the Study of Evaluation, Los Angeles, 1970.

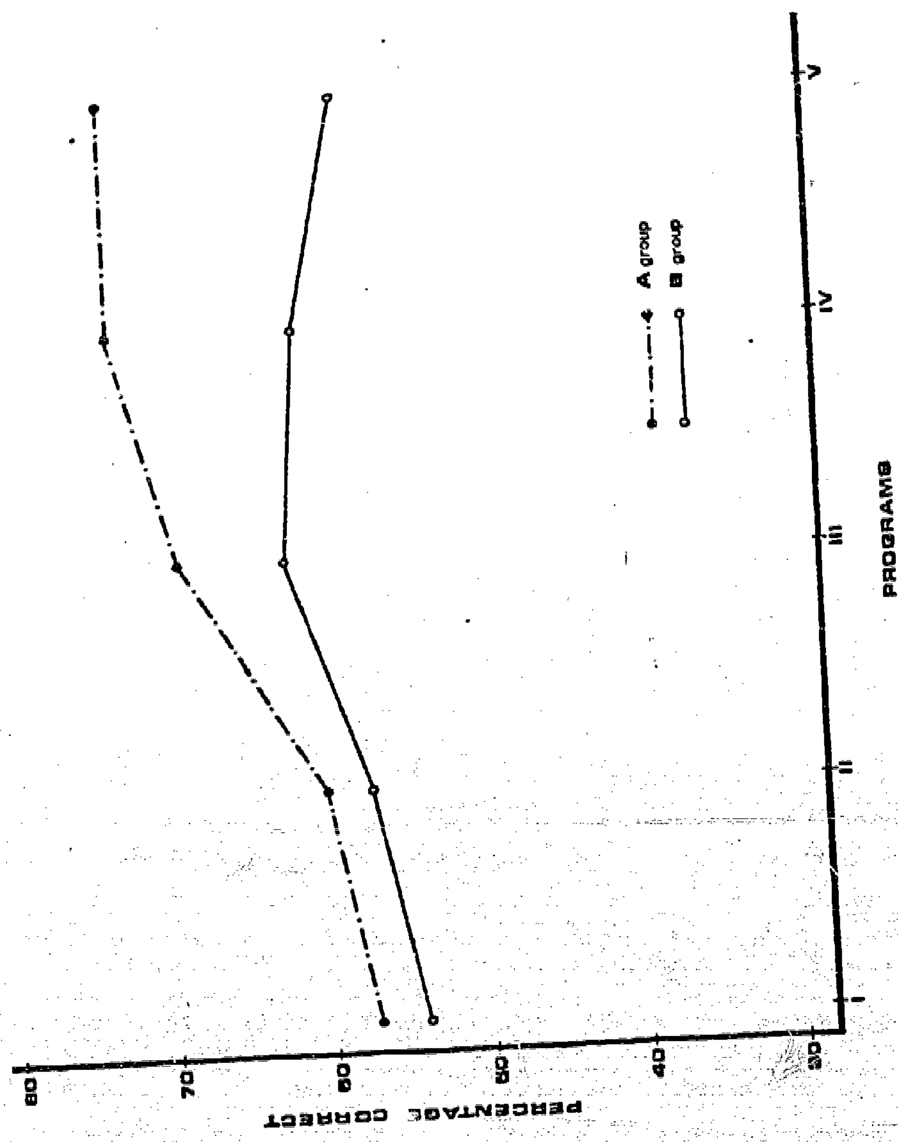
Staats, A. <u>et al</u>., "Cognitive Behavior Modification: 'Motivated Learning' Reading Treatment with Subprofessional Therapy-technicians," <u>Behaviour Research and Therapy</u>, <u>1</u>, 1967, pp. 283-299.

Strang, H. R., "Some Preliminary Findings on Token Spending Patterns in Disadvantaged Grade School Students," <u>Journal of Educational Research</u>, <u>64</u>, 1971, pp. 344, 364.

Strang, H. R. & Wolf, M. M., "Automated Reading Instruction in the Ghetto," <u>Child Study Journal</u>, 1971, (in press).

Wolf, M. M., Giles, D. K. & Hall. V., "Experiments with Token Reinforcement in a Remedial Classroom," <u>Behaviour Research and Therapy</u>, <u>6</u>, 1968, pp. 51-64.


RATIO OF SUBSEQUENT ERRORS TO INITIAL ERRORS
.80
.70
.60
.50
.40
.30
A group
B group
I
II
III
IV
V
PROGRAMS


PERCENTAGE CORRECT
80
70
60
50
40
30
I
II
III
IV
V
PROGRAMS
A group
B group

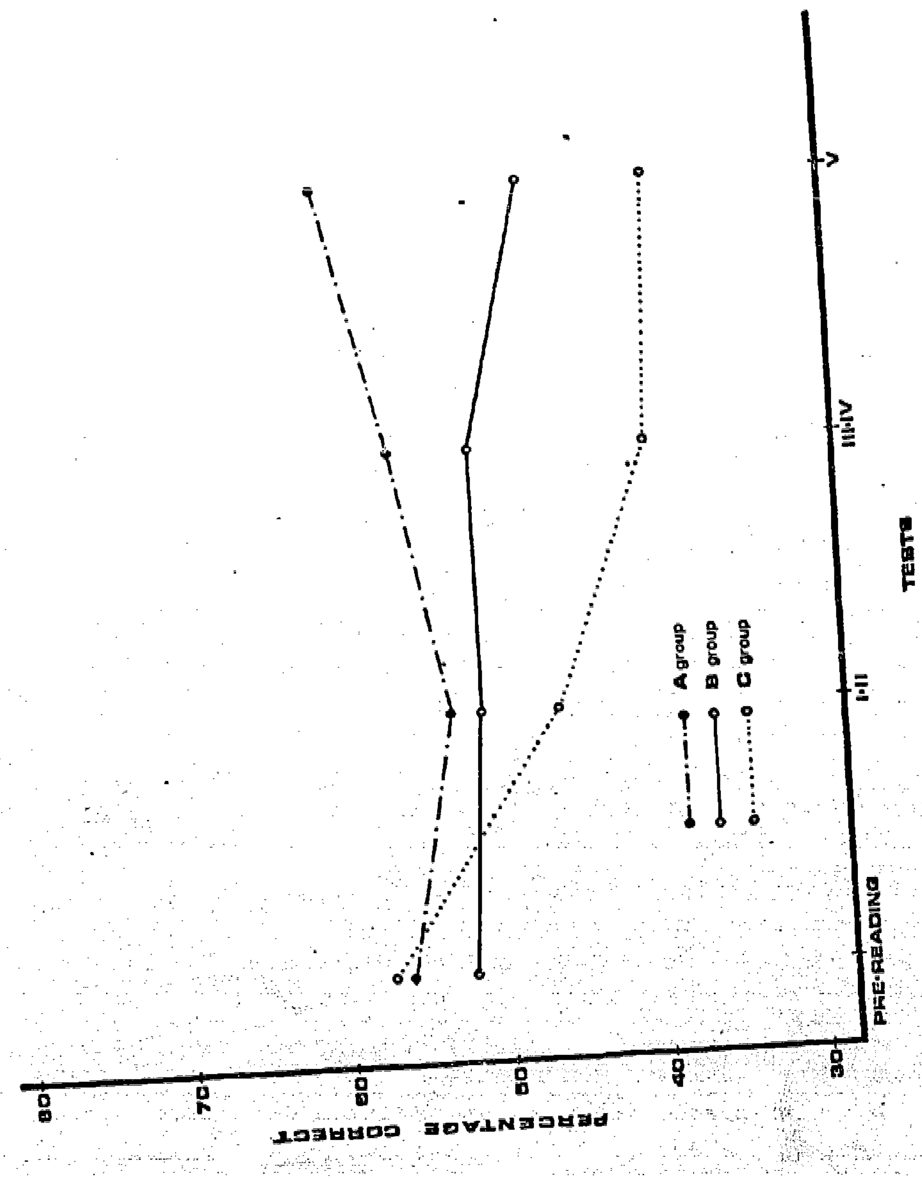

PERCENTAGE CORRECT
80
70
60
50
40
30
PRE-READING
I-II
III-IV
V
TESTS
A group
B group
C group

Table 1

Pre-instruction Verbal I.Q., Constructed Reading, and Standardized Reading Test Performance of the Three Groups

Group	Number of students	Mean Lorge-Thorndike verbal I.Q. score	% correct on constructed 4th level reading test	Mean score for timed CAT-STEP-Metropolitan standardized reading tests
A	6[a]	82	56	41[b]
B	7	81	52	42
C	6	79	59	42

[a]Although all groups were originally composed of seven students, due to public school attrition, means are given for those students who completed the public school year.

[b]Scores in months

Table 2

Improvement in the Number of Initial Correct Responses

Across Types of Comprehension from Programs I-II to IV-V

Students	Type of comprehension	Mean number of initial correct responses per program I-II	IV-V	T value	Number of students showing improvement
Group A (N=6)	Reproductive	39.3	47.3	0*	6
	Interpretive	34.5	40.3	1*	5
	Successive	18.1	26.3	0*	6
Group B (N=7)	Reproductive	36.8	39.3	8	4
	Interpretive	30.4	31.1	12.5	3
	Successive	20.6	22.3	6	4

*$p < .025$

NOTE: Each program was comprised of 60 reproductive, 60 interpretive, and 36 successive comprehension questions.

30

Table 3

Standardized Test Results

(Timed)

Group	Pretest means for timed CAT-STEP-Metropolitan standardized reading tests	Posttest means for timed CAT-STEP-Metropolitan standardized reading tests	Overall gain	Gain over controls
A	41[a]	51	+10*	+5
B	42	54	+12*	+7
C	42	47	+ 5	

[a] Scores in months

* $p < .05$

31

Table 4

California Test Results

(Raw Scores)

Group	CAT - Part G Form W administered at beginning of school year. This was timed.	CAT - Part G Form W administered at end of school year. This was timed.	CAT - Part G Form Z administered at end of school year. This was untimed.
A (audio-visual tutoring)	12.5[a]	15.2	22.2
B (trial-and-error)	12.6	16.6	20.0
C (math control)	13.5	13.8	16.8

[a] Mean number correct on the 30 item test

32

Table 5

Means, Standard Deviations, and F-ratios for Initial Correct Answers, Ratio of After-tutoring Errors to Before-tutoring Errors, Time Before Registering First Answer to Questions, and Time Spent Per Tutoring Application at 4th, 5th, and Combined 5th-6th Reading Levels

Measure	4th grade (N=19)		5th grade (N=19)		Combined 5th-6th grade (N=19)		F
	M	SD	M	SD	M	SD	
First answer errors[a]	92.53	16.71	101.58	18.48	99.47	18.83	10.05**
After-tutoring error/before-tutoring ratio	.53	.13	.48	.22	.49	.17	1.06
Time before registering first answer to question[b]	2.13	.36	2.34	.41	2.48	.49	11.27**
Time spent per tutoring application	1.77	.26	1.92	.29	1.88	.30	3.64*

[a]Total number of questions per program equal to 156

[b]Times given in minutes

*significant p∠.05

**significant p∠.01

Table 6

t Tests for Individual Mean Differences on First Answer Errors,
Time Before Registering First Answer to Questions, and
Time Spent Per Tutoring Application at 4th, 5th, and Combined 5th-6th Reading Levels

Measure	Mean difference between 4th and 5th levels	Mean difference between 5th and combined 5th-6th level	Mean difference between 4th and combined 5th-6th level
Number of 1st answer errors	-9.05**	+2.11	-6.94**
Time before registering 1st answer to question[a]	+ .21*	+ .14	+ .35**
Time spent per tutoring application	+ .15*	- .04	+ .11

[a] Time in minutes

* $p < .05$

** $p < .01$

34

Table 7

Means, Standard Deviations, and F-ratios for

Reproductive, Interpretive, and Successive Comprehension

First Answer Accuracy at 4th, 5th, and Combined 5th-6th Reading Levels

Type of comprehension	4th grade		5th grade		Combined 5th-6th grade		F
	M	SD	M	SD	M	SD	
Reproductive	39.10	10.07	42.89	7.09	40.37	7.77	3.58*
Interpretive	36.95	4.71	35.00	6.70	34.63	6.38	2.13
Successive	16.47	5.01	23.05	7.57	24.47	6.97	33.44**

*p ∠ .05

**p ∠ .01

NOTE: 60 reproductive, 60 interpretive, and 36 successive questions comprised each program.

35

Table 8

Tests for Individual Mean Differences on

Reproductive and Successive First Answer Accuracy on

4th, 5th, and Combined 5th-6th Reading Levels

Type of comprehension	Mean difference between 4th and 5th levels	Mean difference between 5th and combined 5th-6th level	Mean difference between 4th and combined 5th-6th level
Reproductive	+3.79*	-2.52	+1.27
Successive	+6.58**	+1.42**	+8.00**

*$p < .05$

**$p < .01$

36

Table 9

Tutored and Control Students' Pretest and Posttest Accuracy

Across the Three Types of Comprehension

Comprehension type	% correct in experimental condition		% correct in control condition	
	Pretest	Posttest	Pretest	Posttest
Reproductive	47.0	72.8***	63.2	80.0**
Interpretive	57.6	72.0***	52.2	66.6*
Successive	40.8	70.3***	53.6	53.4

*Change significant ∠.05

**Change significant ∠.02

***Change significant ∠.01

37