R E P O R T R E S U M E S

ED 013 539 EM 000 489

AN APPROACH TO THE USE OF COMPUTERS IN THE INSTRUCTIONAL PROCESS AND EVALUATION.

BY- SCHURDAK, JOHN J.
INTERNATIONAL BUSINESS MACHINES CORP.

REPORT NUMBER RR-RC-1432 PUB DATE 6 JUL 65
EDRS PRICE MF-$0.25 HC-$1.68 42P.

DESCRIPTORS- *PROGRAMED INSTRUCTION, CONVENTIONAL INSTRUCTION, *COLLEGE GRADUATES, *COLLEGE STUDENTS, PROGRAMING, *PROGRAMED TEXTS, LINEAR PROGRAMING, *TEACHING MACHINES, ACADEMIC ABILITY, ACADEMIC PERFORMANCE, STUDENT ATTITUDES, EXPERIMENTS, NEW YORK CITY, NEW YORK

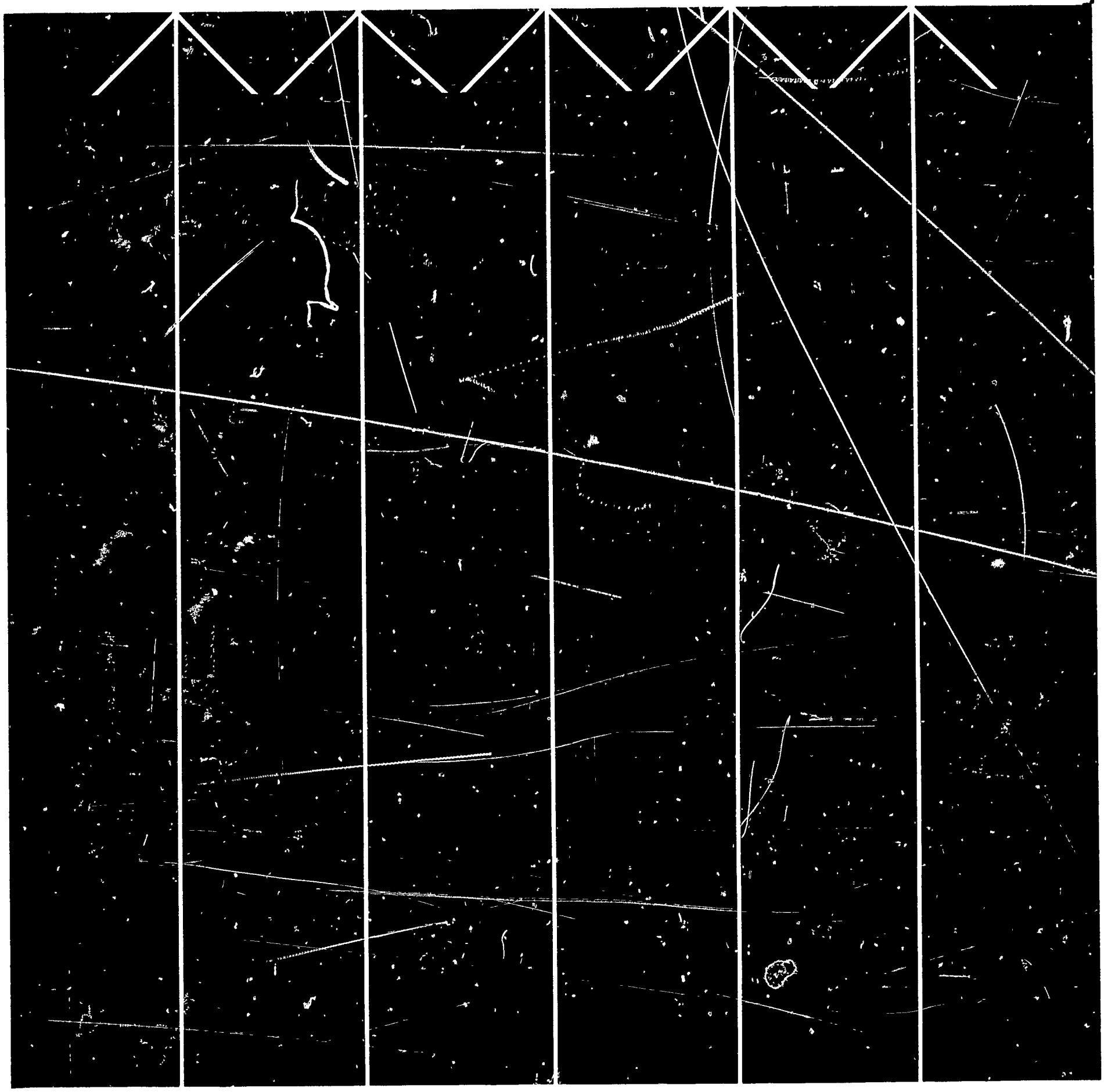
RELATIVELY ABLE, WELL-MOTIVATED COLUMBIA UNIVERSITY UNDERGRADUATES AND GRADUATES WERE ASSIGNED AT RANDOM TO 3 GROUPS AND TAUGHT FORTRAN PROGRAMING BY COMPUTER, PROGRAMED TEXT (LINEAR TYPE) AND CONVENTIONAL TEXT. GOAL OF THE COURSE WAS TO USE THE COMPUTER TO ACHIEVE INTELLIGENT INDIVIDUALIZATION OF INSTRUCTION BASED ON EACH STUDENT'S PERFORMANCE. SUBJECTS IN THE COMPUTER GROUP COMMUNICATED WITH A 1440-1448 COMPUTING SYSTEM THROUGH 2 1050 KEYBOARD-TYPE TERMINALS. KEY FEATURES OF COMPUTER INSTRUCTION WERE IMMEDIATE FEEDBACK AND CORRECTION OF ERRORS, AND PROGRESS CONTINGENT ON THE QUALITY OF EACH RESPONSE IN FREQUENT COMPREHENSIVE EXAMINATIONS. ALTERNATE FORMS OF AN ACHIEVEMENT TEST WERE ADMINISTERED ON THE DAY AFTER COMPLETION OF THE COURSE. THE COMPUTER GROUP'S MEAN SCORE WAS SIGNIFICANTLY HIGHER THAN THE MEAN SCORES OF EITHER OF THE OTHER GROUPS. THE GROUPS DID NOT DIFFER RELIABLY ON LEARNING RATE, ALTHOUGH ERROR RATE ON DIAGNOSTIC AND DRILL SECTIONS WAS RATHER HIGH. STUDENTS' ATTITUDES TOWARD THE COMPUTER AND EQUIPMENT, AND TOWARD THE PROGRAMED TEXT WERE FAVORABLE. IT WAS CONCLUDED THAT THE COMPUTER PROGRAM CAN ADJUST TO A WIDE RANGE OF INDIVIDUAL DIFFERENCES IN LEARNING AND THAT REVISION OF EXISTING CURRICULUM MATERIALS BEFORE USING COMPUTERIZED INSTRUCTION IS NOT NECESSARY. (LH)

A

IBM RESEARCH DIVISION

AN APPROACH TO THE USE OF COMPUTERS IN THE INSTRUCTIONAL PROCESS AND AN EVALUATION

John J. Schurdak



RC-1432

July 6, 1965

AN APPROACH TO THE USE OF COMPUTERS IN THE INSTRUCTIONAL PROCESS AND AN EVALUATION

by

John J. Schurdak[1]

IBM Watson Research Center
Yorktown Heights, New York

ABSTRACT: An approach to the use of computers in instruction was developed and evaluated. In a computer course students communicated with a 1440-1448 computing system through 1050 terminals. Three experimental groups were established to learn a portion of a Fortran course by three treatments--by computer, programmed text, and conventional text. There were sizeable differences among the groups on an achievement test administered the day following completion of the course. The computer groups mean score was significantly higher than the mean scores of either of the other groups. The computer program made considerable adjustments to individual differences in ability to learn the material. The students' attitude toward the computer course and equipment, and toward the programmed text was good.

Research Report
RC-1432
July 6, 1965

Reprinted 11/22/65

INTRODUCTION

The existence of the modern high speed digital computer suggests a new approach to the solution of some enduring instructional problems. Course material and course logic may be stored in computer memory. It is possible for students to communicate with such computing systems through a variety of instruments, one of the standard devices being a console with a keyboard similar to that of an electric typewriter. The computer, unlike most teaching aids, is a very flexible instrument, for its instructional logic may be shaped toward a variety of teaching objectives.

As a part of a program to evaluate the use of computers in the instructional process a portion of a computer course in FORTRAN, a programming language, was developed. Because the computer as an instructional device does not, by its inherent characteristics, rigidly structure the teaching situation, it is necessary to first discuss the instructional characteristics of this particular computer course.

1. The attempt here is to use the computer as a device for achieving an intelligent individualization of instruction based on each student's actual performance in the course. Since each student can communicate with the system independently, and since the computer is able to make logical decisions based on an analysis of incoming student data, the capability exists for intelligent adaption of instruction to each individual. In my judgment it is in its potential for precisely,

and in a sophisticated manner, adjusting the process of instruction to important individual student characteristics that most justifies the consideration of the use of computers for teaching purposes.

The FORTRAN course developed for this study provides at almost every point in the program from three to five levels on which a student may proceed and provides frequent opportunities for switching levels. The student's precise path is determined by his actual performance in the course.

2. The second instructional characteristic of the course is the use of the computer as a device for providing immediate knowledge of the correctness, partial correctness, or incorrectness of responses to each student. Though not every study has shown this to be an advantage it still seems probable that such an immediate realistic appraisal of one's knowledge is a valuable aid to effective learning.

3. A third instructional characteristic of this course is the use of the computer as a device for the immediate identification and correction of erroneous conceptions, or incomplete understanding of concepts by the student. In this course an attempt is made to prevent the learning of new material until the student demonstrates a thorough and accurate understanding of his present assignment.

4. A fourth characteristic of this course is its use of the computer to make every student response meaningful, in the sense that what

happens following each response is dependent on the quality of that response. In the computer course developed for this study a correct response enables the student to quickly move on, while incorrect responses simply bring forth more and more material on the same narrow phase of the subject, forcing the student to spend a great deal more time on small step material. For the student by far the easiest and quickest way to complete the course is to respond correctly as often as possible. It seems possible that a good deal of the contradiction and ambiguity in results in studies of instructional materials and methods is due to the extreme variability in the meaningfulness to individual students of materials, questions, responses and reinforcements.

5. A fifth characteristic of this course program is that it is designed to attempt to elicit a maximum contribution from the student in the solution of the problems presented. All students are periodically brought back to the most difficult track, and when student errors do occur the strategy is to bring the student to a correct solution to the problem with a minimum of assistance from the computer program. The student almost always finds it advantageous in terms of time and effort to attempt an answer, even though he may be basing his response on incomplete knowledge or understanding. The benefit of such an approach is unlikely to be detected in short range retention

tests, but once the rather protective school environment is left behind there is no one to carefully adjust the step size of problems to an individual's capabilities, nor is anyone ever-present to assist when an individual's solution is unsuccessful, and therefore it seems prudent to prepare students by such an approach.

It would be very time consuming to prepare the various treatments of the course material required to take advantage of a computer's decision making capabilities. Rather than engage in extensive curriculum construction activities, it was decided to use existing curriculum materials to the fullest extent possible for the FORTRAN course. The role of the computer, where such materials were employed, would be to make logical decisions as to the patterns of their use, based on each student's incoming responses. In practice, this system would appear to offer major time saving advantages where good curriculum materials exist suitable for use in this type of presentation. The computer's capability for data storage and analysis would seem to offer a major instructional advantage. Since every response of every student could easily be automatically recorded, it is possible to perform extensive statistical analyses to give the author or teacher invaluable information about the effectiveness of a course. Improvements of the course to meet learning objectives could be based on this realistic information.

COMPUTER COURSE

A flow chart of the logic of the computer course appears as Fig. 1.

Materials

The existing curriculum materials used in this course and in this study were D.C. McCracken's A Guide to FORTRAN Programming (1961) and S. C. Plumb's programmed text FORTRAN - Self Teaching (1963), third edition. McCracken's book is a conventional text, with chapters logically divided into subchapters, with headings for each of these subsections. The step size, or level of difficulty is relatively high. Questions concerning a chapter are provided at the end of that chapter, and a selected number of answers are furnished at the back of the book. The Plumb text is basically a linear program, and is characterized by a small frame presentation, a frame being followed by a question concerning that frame. The answer to each frame is always given on the page following. A small number of drill exercises and an examination are also provided by the author.

The course material and the course logic, with the exception of the McCracken text, were stored in the memory of a 1440-1448 computing system, and students communicated with the system through two 1050 terminals, which have keyboards generally like those of electric typewriters. Except for the reading of the McCracken text, all course materials, questions and instructions were given to the

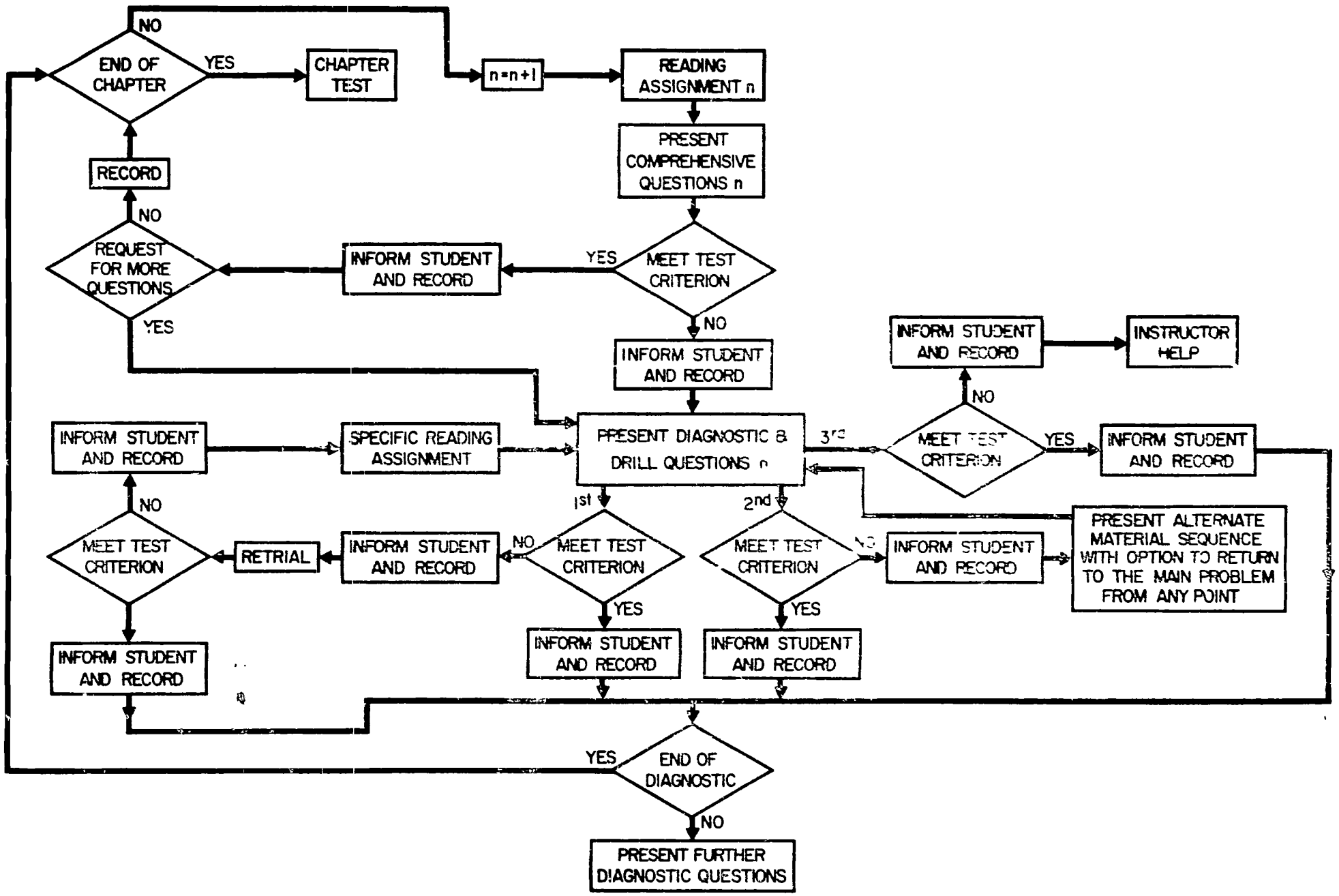


Figure 1 Computer course flow chart

students by messages typed by these 1050 terminals, and all student responses were entered through these keyboards.

Initial Presentation of Material

In the computer course portions of the McCracken text constitute in every case the initial presentation to the student. The student is instructed to read a portion of the text, a section that is considerably longer than the frame of a typical linear program, but much briefer than a full chapter in a normal text. The objective is to present the student with a logical whole, a meaningful structure, a single concept of interrelated set of concepts, yet not allow the student to proceed too far without determining the extent of his comprehension of this reading assignment. It was hoped that the relative difficulty of this material would permit a differentiation of the student population by subsequent tests of comprehension, and that this differentiation would permit the use of the computer's logical decision making capabilities to tailor the course on the basis of individual student performance.

Comprehensive Examinations

Whenever a student completes a reading assignment, "to the point where he feels he understands it," his understanding of the material is immediately tested by a set of questions that together constitute a "comprehensive examination". This examination is a set of relatively complex and difficult questions designed to determine the

student's overall grasp of the subject matter he has just studied. If the student meets the test criteria he is instructed to continue by reading the next section. If, however, after meeting the test criteria he feels uncertain of his knowledge, he has the option of receiving further questions regarding the material under study. The number of comprehensive questions for any particular assignment is small, so that the very able student is able to proceed quickly through the course, with little more time than it takes him to read the text.

Diagnostic and Drill Questions

If, however, the student fails to pass a comprehensive examination, he is presented with a series of relevant diagnostic questions --relatively narrow questions designed to identify the component causes of the student's failure to pass the more complex comprehensive examination, and to provide the student with drill problems related to these causes. If the student meets the criteria for a particular diagnostic test he is presented with further series of relevant questions, until all appropriate diagnostic questions have been exhausted.

In these diagnostic sections of the course in almost every case the student is given the opportunity to try again after making an initial error. This retrial feature was incorporated because: (a) it was desired to give the student a maximum opportunity to solve a

problem before giving him assistance, (b) it was felt that with no more additional knowledge than the fact that his original answer was incorrect, the student would frequently respond correctly on a re-trial, and (c) this feature should in effect considerably enhance the validity of the original question, in the sense that a second error would provide strong evidence that the student was genuinely unable to answer a question of that nature.

Specific Reread Instructions

If, however, a student fails to meet any of the diagnostic test criteria, he is first instructed to reread a specifically relevant portion of the text--normally a section one, two, or at the most three paragraphs in length--after which he is again presented with the diagnostic question he had just missed. If this directed rereading of a portion of the relatively large step text has accomplished its purpose, and the student answers the question correctly, he continues to receive further appropriate diagnostic questions, until all relevant questions have been presented.

Explanatory or Remedial Material

Whenever the rereading of the text proves not to be effective, the student is presented with supplementary information from computer memory which serves a remedial function. The major charac-

teristics of effective remedial material are small step size and careful sequencing of material. These are precisely the characteristics of good linear programs, and therefore the supplementary material for this course was largely taken from S. C. Plumb's programmed text FORTRAN - Self Teaching (1963).

At any point in the remedial sequence the student can decide he now understands the necessary concepts, and can depart from the sequence simply by attempting to answer the last diagnostic question he had missed. A correct answer returns him to the main stream of diagnostic questions. If the student, however, goes through an entire remedial series, he is again brought back to the last diagnostic question posed to him.

Instructor Help

In the normal operation of the course, if he again cannot respond correctly, he is directed to ask the teacher for help. The instructor thus comes in to assist the individual student when he has failed to answer appropriate diagnostic questions at least four times, even though he has reread a specifically relevant portion of the text, and worked entirely through a remedial sequence. In this study, however, the instructor help feature was not used.

The approach outlined here is an attempt to fit a computer into a normal instructional process, assigning the system specific roles in

that process, based on some of the computer's salient characteristics and capabilities.

EVALUATION PROCEDURE

To evaluate this approach three experimental groups were established to learn a portion of a FORTRAN course by three treatments--by computer, by programmed text, and by conventional text. Forty-eight undergraduate and graduate students in summer session at Columbia University were assigned, on a random basis, to the subgroups contained in the three experimental treatments. In the first treatment students were required to take the FORTRAN course described above, with the exception that, in order to provide better experimental control in this particular study, the "teacher help" and the "opportunity for review" features were not utilized. If a student came to a part of the course where he would normally receive help from the instructor, he was instead simply given the correct answer to the problem at hand.

In the second treatment the students learned the same concepts by using the programmed text written by S. C. Plumb. The subjects were instructed to answer the questions following the frames on a separate answer sheet, before noting the correct answer on the following page, and to work the small number of drill exercises and

the relevant questions on an examination provided in the text. The answers to these questions were also provided on a following page.

In the third treatment the subjects studied the identical concepts in what may be described as a "workbook" approach. They were required to read a chapter in McCracken's text to the point where they felt they understood the material, then to work the problems at the end of the chapter. They were instructed to put their responses to these problems on separate answer sheets before looking at an augmented list of correct answers contained in the back of the book.

All students worked independently. There were two 1050 teaching terminals available, so in each treatment there were two subjects in each block. All three blocks of two students each were started in their learning task at the same time on the first day of each week. All subjects worked in two-hour sessions, beginning at the same time on successive days, until they had completed the course.

During the course of this experiment it was possible to test the students voluntarily taking the regular university computing center course in FORTRAN. Though the students in this class are not a random selection from the experimental population, the data are here reported because of the paucity of information about the

effectiveness of computer assisted instruction, because the data furnish further information by which to judge the criterion test, and because many consider the only meaningful comparison is one made with a normal classroom situation. The material contained in the three treatments and in the achievement test was discussed in detail with the instructor prior to the class. The instructor was well regarded by supervisor, colleagues, and former students, a rare combination. The teacher presented this portion of the course in three 1 1/2 hour sessions, held on consecutive days, to a class of 16 students, and the students in addition received homework assignments in the McCracken text. Achievement test data were obtained from 12 students on the day following completion of these three sessions. Since these students do not constitute a random selection from the experimental population, the data from this treatment are simply reported and are not analyzed in conjunction with the other experimental results.

Subjects

Some of the students in the sample were taken from the roster of those signed for a voluntary beginning course in FORTRAN, to be presented by the Columbia University Computing Center. Others were paid subjects who received a fixed payment without regard to the length of time required to complete the material. There were 5

unpaid and 11 paid students in each group, for a total per treatment of 16. The computer and conventional text treatments contained 8 undergraduates, 6 university graduate students, and 2 Teachers College graduate students. The programmed text treatment included 7 undergraduates, 7 university graduates, and 2 Teachers College graduate students. None had any prior knowledge of any computer programming language. As a group these were relatively able, well-motivated subjects.

On the Friday prior to the start of his learning session, each subject was administered the Henmon-Nelson Tests of Mental Ability, College Level, Form A. In addition, those who were to learn via computer assisted instruction received a keyboard familiarization session of approximately 15 minutes, designed to acquaint these students with the operation of the 1050 terminal. All groups were told that they were participating in a study to evaluate several different methods of teaching programming, and no questions concerning the course material or the treatments were answered during the course of the experiment.

Criterion Tests

The day following his completion of the course, each subject was administered a comprehensive achievement test. Despite one's best efforts, it is possible, particularly after working intensely to

develop one particular treatment, to include some bias in the construction of a test. Therefore Dr. P. Kenoyer, a testing specialist, was asked to study all three treatments and to prepare an independent test that would emphasize fairness to all treatments. The questions on the two tests were interleaved on the examination so that to all appearances they appeared to be one test. The first test contained 229 points, Dr. Kenoyer's - 122, for a grand total of 351 points. Almost all of the test questions required either the translation of mathematical statements into FORTRAN or the detection and correction of erroneous FORTRAN statements from a list of program statements.

The two separate achievement tests turned out to be virt ally identical in question format and results. The two tests scores correlated at the .9 level in all three treatments. The two tests were essentially alternate forms of the same test, and therefore the total achievement test score alone is used in these analyses. After taking this examination, each student was given an attitude questionnaire, a somewhat revised version of the one originally prepared by Dr. K. H. Wodtke of Pennsylvania State University.

RESULTS

External Criteria

A summary of the average performance for each of the three treatment groups is given in Table 1. Column A of Table 1 shows that the matching of the groups was effective, the three groups are equivalent according to mean Henmon-Nelson total score. Column C of Table 1 shows that there were, however, sizeable differences among the treatment groups on the criterion test. Group C (computer group) scored eleven percentage points, over one standard deviation, higher than group P (programmed text group), which in turn performed approximately six percentage points higher than group T (textbook group). The bottom of the range for the C group is approximately at the mean for the T group. Columns D and F of Table 1 show that there was the least variation among the students in group C and the most variation in group T. The groups do not differ reliably on time to complete learning the FORTRAN materials. Although column B of Table 1 shows that group T took somewhat longer to complete the course than the other two groups, this difference was very small and not statistically significant.

Figure 2 summarizes the performance of the 16 students in each of the three experimental groups on the criterion test. The abscissa is the Henmon-Nelson total score for the students. The ordinate is

Table 1

External Criteria

	A Mean Henmon-Nelson total scores	B Mean times to complete course	C Mean scores on criterion test	D Standard deviations of criterion test scores	E Ranges of criterion test scores
C	77.1	243.2	87.4	9.5	70 - 98
P	79.0	245.4	76.2	13.0	47 - 92
T	77.4	269.1	70.6	21.2	36 - 93

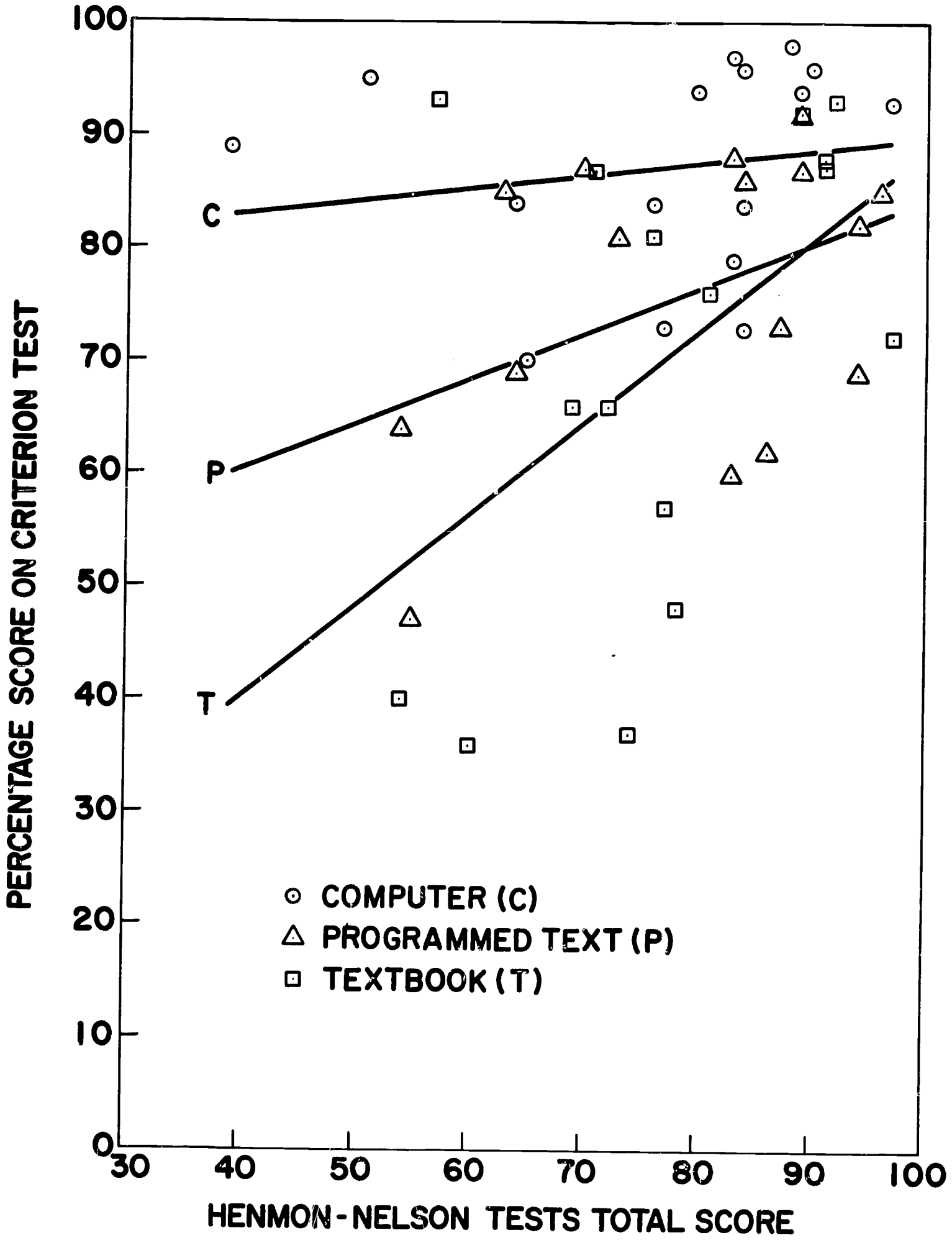


Figure 2 Regression of criterion test score on total Henmon-Nelson score for computer (C), programmed text (P), and textbook (T) groups.

the percentage correct score on the combined criterion tests. The lines drawn in Fig. 2 are the best fits for each treatment group by a least squares criterion.

An analysis of covariance was performed to determine if the differences in mean score on the criterion test among the three groups were statistically significant. Bartlett's test disclosed significant heterogeneity of variance, $\chi^2 = 9.51$, $p < .01$. In an attempt to reduce the heterogeneity an arcsin transformation was applied to the percentage correct responses on the criterion test for each student. The covariate in this analysis was the Henmon-Nelson total score. The analysis disclosed that the differences among the means for the three treatments were significant, $F = 6.50$, df 2, 44, $p < .01$. The mean test score for the students in group C was also significantly higher than the mean test score for the students in group P, $F = 7.26$, df 1, 44, $p < .01$. The difference in means between groups P and T was not statistically significant.

Using as covariates either of the two separate types of scores given by the Henmon-Nelson Test, the Verbal or the Quantitative score, the results were essentially the same. The differences in treatment means were statistically significant, and group C differed significantly from group P.

It will be noted from Fig. 2 that a considerably larger number of students in the C group received "excellent" scores on the criterion test, defined as scores above 90. Eight students in the C group, one in the P group, and three in the T group received such scores. According to Fisher's exact test for a 2x2 contingency table, the difference between the C and P groups would occur by chance only .8% of the time, and the difference between the C and T groups would occur by chance 6.8% of the time.

There is also a markedly larger number of students in the P and T groups who received "very poor" scores on the criterion test, defined as scores below 70. The numbers receiving such scores were zero for the C group, six for the P group, and seven for the T group. Fisher's exact test for a 2x2 contingency table discloses that the differences between the computer and the programmed text groups was significant, $p \leq .009$, and the difference between the computer and textbook groups was also significant, $p \leq .003$.

It is evident from Fig. 2 that students with high Henmon-Nelson scores appear to obtain approximately the same scores on the criterion test in each experimental treatment, i.e., the regression lines appear to converge for high Henmon-Nelson scores. Students with lower Henmon-Nelson scores perform quite differently on the criterion tests in the three treatment groups. Those students in the

group that learned FORTRAN by computer appear to be superior to those subjects who learned the FORTRAN by programmed instruction. The students in this latter group appear to be superior to students who learned FORTRAN from the text. The test for heterogeneity of regression discloses that the heterogeneity of regression is, however, not statistically reliable $\underline{F} = 1.95$, $\underline{df}\ 2, 42$, $\underline{p} = .17$.

One of the effects of the treatments is manifested in the differential uniformity of performance on the criterion test. Students with markedly different Henmon-Nelson scores are more similar on final test performance in group C than in groups P or T. The product moment correlation coefficients between Henmon-Nelson total scores and criterion test scores was .172 for group C, .422 for group P, and .519 for group T. These differences between correlation coefficients were not, however, statistically significant.

In the university computing center course the 6 class members who took the Henmon-Nelson test had a mean total score of 75. The average total time spent in learning the material was 6.6 hours, with homework times being reported by the students. The average score for the 12 students who took the criterion test was 82.7, between the C and P experimental groups. The range of scores for the class was 67.0 to 95.7 with a standard deviation of 9.0.

Internal Criteria

The internal results for the computer course, presented in Tables 2 and 3, give evidence of the differential treatment accorded to individual students by the computer program. In the overall course the range of student responses required to meet internal comprehensive and diagnostic test criteria was 72 to 352, approximately a five to one ratio. One student completely passed six of the eight comprehensive tests, while another was unable to meet the criteria for a single such test. The number of student responses required to meet diagnostic test criteria ranged from 22 to 291, a ratio of 13 to 1.

The error rate for the diagnostic and drill sections of the course was 23.7%, a relatively high rate, considering that some of these questions were repetitive drill material. The students' error rates for such questions ranged from 10.5% to 35.6%.

The data give evidence that the "retrial" and "reread" features of the instructional strategy were effective. Without being given any additional information after an initial error, beyond the fact that their original answer was wrong, the students were correct on 59.7% of their retrials.

After a total of 151 reread instructions, the students responded correctly 53.0% of the time. Both the retrial and reread instructions

Table 2

	Total number of responses	Comprehensive tests passed	Comprehensive tests partially passed	Number of diagnostic & drill responses	Diagnostic & drill question error rate	Percentage of correct retrials	Percentage of correct responses after reread
1	72	6	2	22	10.5	50.0	100.0
2	87	6	1	40	11.4	75.0	100.0
3	131	4	3	92	15.8	50.0	75.0
4	148	2	3	106	16.3	69.2	50.0
5	157	3	2	127	23.2	82.6	50.0
6	160	2	4	117	16.5	66.7	100.0
7	163	4	1	121	16.5	60.0	66.7
8	167	2	5	115	17.5	64.2	60.0
9	181	3	2	146	24.8	61.5	80.0
10	188	3	3	142	20.9	87.5	66.7
11	195	3	1	157	19.7	83.3	100.0
12	223	2	1	191	15.9	52.2	50.0
13	252	1	4	191	35.6	56.4	47.1
14	305	3	3	233	34.3	31.7	35.7
15	331	1	1	282	32.1	58.9	52.2
16	352	0	2	291	32.0	46.3	39.2

Table 3

Internal & External Criteria

	Total responses	Comprehensive tests passed	Comprehensive tests partially passed	Diagnostic & drill responses made	Error rate for diagnostic & drill questions	Percentage of correct retrials	Percentage of correct responses after reread	Remedial responses	Hemmon-Nelson total score	Time in minutes to complete course	Criterion test score
Ranges	72-352	0-6	1-5	22-291	10.5-35.6	31.7-87.5	35.7-100.0	0-45	39-97	155-427	70-98.
Top Quarter* Mean	109.5	4.5	2.3	65.0	14.8	67.7	70.0	1	81.8	197.5	92.5
Treatment Mean	194.5	2.8	2.4	148.3	23.6	59.7	53.0	9.5	77.1	243.2	87.4
Lowest Quarter* Mean	310.0	1.3	2.5	249.3	33.2	49.0	42.7	34.3	77.3	352.3	73.6

*Ordered according to number of responses required to meet internal criteria.

were, as indicated in Table 3, more effective with the students in the top quarter, ordered according to the number of responses required to meet internal criteria, but were also effective with the students in the lowest quarter of this treatment group.

For all problems, the combined retrial and reread strategies led the students to correct answers in 80.5% of the cases after an initial error. Since neither required much preparation time on the part of the course author, they proved to be a quite economical means of leading students to the correct answers to these problems, which test the student's understanding of the subject at hand.

Since the retrial and reread instructions were so successful, the program required but 152 responses to remedial (programmed text) material, 129 of which were made by the subjects in the lowest quarter, according to internal criteria. When the remedial material was completed, the students were again presented with the last diagnostic question they had missed. These repeat diagnostic questions were answered correctly 55.9% of the time.

Relationships between Internal and External Criteria for the Computer Treatment

The functions of the comprehensive examinations was to identify those who had learned the material well in order to quickly move them on to new subject matter, and to detect those whose knowledge

Table 4

Product Moment Correlations Between Internal & External Criteria

	Criterion Test Score	Time to Complete Course	Henmon-Nelson Total Score
Total Number of Responses	-.828**	.749**	-.228
Comprehensive Tests Passed	.747**	-.478*	.511*
Number of Diagnostic and Drill Responses	-.801**	.697**	-.303
Diagnostic Drill Question Error Rate	-.768**	.690**	-.052
Percent of Correct Responses on Retrial	.340	-.401	.048
Percent of Correct Responses After Reread	.465*	-.474*	.419

* Significant at .05, one tailed

** Significant at .01

was incomplete so as to present them with relevant diagnostic and drill material. If the comprehensive tests performed their functions effectively, student performance on these tests would correlate highly with performance on the criterion test. Table 4 presents the product moment correlations between internal and external criteria. The correlation between the number of comprehensive tests passed and criterion test scores was .747 (significant at .01).

The more the student failed to meet internal test criteria the more material and the greater the number of questions presented to him by the computer program, in an effort to immediately correct the misunderstandings detected. Thus the total number of student responses and the number of student diagnostic and drill responses constitute negative indices of student performance in the course. The correlation between total numbers of responses and criterion test scores was -.828, between the numbers of diagnostic and drill responses made and criterion test scores was -.801, and between diagnostic and drill question error rates and criterion test scores was -.768. All of these correlations are significantly different from 0 at .01. The more able students, according to internal criteria, though they received less exposure to course material, still achieved higher scores on the criterion test.

It appears that the major determinant of time to complete the

course was simply the number of student responses required to meet internal course criteria.

The Henmon-Nelson Total Scores had positive correlations with the numbers of comprehensive tests passed (.511), and the percentage of correct responses after reread instruction, (.419). The Henmon-Nelson Total Scores had, however, no correlation with either the Diagnostic and Drill Question error rates, or the percent of correct retrial responses. In this course, the Henmon-Nelson scores were reasonable predictors of student performance where a reading assignment was followed by an immediate test or question, as in the case of comprehensive tests or responses after reread instructions. The Henmon-Nelson scores were, however, completely uncorrelated with student performance on the retrial responses, or in answering the diagnostic questions. In these circumstances the student was generally responding to a series of questions, after an initial error or set of errors, with little textual or other aid, a situation different from that of the conventional classroom. The ability to perform well in this type of situation was not predictable from the Henmon-Nelson scores.

Attitude Questionnaire Results

Since student attitude is an important consideration in the evaluation of an instructional approach, the students in all three

treatments were administered an attitude questionnaire immediately following their completion of the achievement test. All were asked "How much did you enjoy this method of presentation?" Fifteen in the computer group, twelve in the programmed text group, but only seven in the textbook group, reported they enjoyed the method of presentation given them in some degree, whereas eight in the textbook group stated they did not. Apparently most students enjoyed the computer and programmed text treatments, but not the textbook presentation of this course.

The students were also asked "In terms of the amount you learned, how would you rate this course compared to usual methods of instruction?" On this, and on the questions that follow, they were given a scale ranging from extremely, quite, and slightly through neutral, slightly, quite, and extremely. The extremes of each scale are defined by bipolar adjectives such as, in this question, inferior-superior. Twelve in group C, thirteen in group P, but only four in group T rated the course given them superior in some degree. In group T nine felt it was inferior to usual methods of presentation. In the judgment of most of the students, C and P were superior presentations, but most regarded T as inferior.

All students were also asked "Did you miss opportunities for discussion of problems?" with the same scale of choices as in the

preceding question. Ten in the computer group, eleven in the programmed text group, and fifteen in the workbook group missed the opportunity for discussion to some degree. In all treatments the opportunity for discussion was missed, but this was missed by more students in the textbook treatment.

Since the equipment used in the computer course was quite unusual in instruction, the students in this treatment were given an additional questionnaire relating to this equipment. In each of these questions the students again chose from a seven point extremely through neutral to extremely scale. Most students found the computer equipment fast (10-1), interesting to use (15-1), easy to operate (11-4), and an aid to learning (9-3).

In treatment C no one rated their typing speed as above average, but seven regarded their typing speed as quite or extremely slow, out of a total of eleven who reported their speed as below average. Even the seven very slow typists found the equipment easy to use (4-2) and an aid to learning (4-1). The overall reaction to the equipment used in this course was quite good, even by those who regarded themselves as poor typists.

The students were encouraged to make their own comments about the course on the attitude questionnaire, and space was left for such comments. Concerning the computer presentation, the

most frequent positive comment was about the immediate testing characteristic. Eight students voluntarily cited this "instant testing", as one phrased it, the immediate questioning to determine the student's understanding, as valuable or desirable. Five students volunteered that the computer course led to a more comprehensive or thorough understanding of the subject matter than they believed they would otherwise have obtained. Five cited the fun of manipulating the machine.

On the negative side the most frequent comment, made by eight students, concerned the lack of opportunity for discussion of problems. It will be recalled that for better experimental control the "instructor help" feature of the course was not utilized in this study.

The general reaction to the computer course was quite good. Almost all enjoyed it, and a number of students were quite enthusiastic about this method of instruction. Even most of these who found certain aspects of the course deserving of criticism had an overall favorable attitude toward the computer presentation.

DISCUSSION

The results of this study suggest that, at least for materials and student populations of a similar nature, computers can bring

genuine advantages to the instructional process. The internal criteria demonstrate that it is possible for the computer program to adjust to a wide range of individual differences in learning, and the external criteria, and the interrelationships between internal and external criteria, suggest that this adjustment can be meaningful in terms of subsequent student performance.

The computer program used in this study was a relatively primitive, simplistic one in terms of computer capability. In view of this, of the brevity of the course, and of the small number in the treatment group, the range of adjustments to individual learning requirements is impressive. Such findings as a 13 to 1 ratio of student responses in the diagnostic and drill sections, and an overall 5 to 1 student response ratio in the course, suggest that computers, as instructional instruments, may be capable of substantially adjusting to a very broad range of inter- and intra-individual differences, leading potentially in time to the achievement of a considerable degree of ungradedness in instruction.

The efficacy of this relatively simple program in bringing about superior criterion test performance in this study suggests a potential for the improvement of instructional effectiveness. Computers give hope of actualizing, in instructional practice, some of the principles of learning or instruction that the normal teaching

situation by itself makes almost impossible.

The position of the regression lines in Figure 2 suggests that the computer course was beneficial to students at all levels of ability, but the slope of the regression lines for this treatment indicates the relative capability of the computer for assisting the less able in the group to mastery of the course material, compared to the other treatments. The differential treatment given to those whose course performance signified the need for additional assistance resulted in a substantial improvement in the performance of the students of lesser ability. A potential of this nature should not be underestimated, for in a course, or a school career, erroneous conceptions accumulate, with negative effects on both student capability for later learning, and student attitudes toward learning.

However, despite the extra assistance provided them. the less able students did not outperform the more able students, and it appears possible from the plot of the data in Fig. 2 that the full benefits of the computer course strategy to the more able students are not indicated due to a test ceiling effect. Though the overall test average of 78. 1 suggests that the test was of appropriate difficulty for this experimental population, seven of the C group with Henmon-Nelson total scores of 80 or above appear to be pushing the test ceiling with criterion test scores above 90, compared to one

person in group P, and two in group T. In essence these persons have learned this introductory material approximately as well as it is possible to do.

The effectiveness of this approach to the use of computers in instruction suggests that it may be possible to introduce computers to the instructional process with real advantage without the prior necessity for complete, and extremely time consuming curriculum reconstruction. The use of already existing curriculum materials seems quite feasible in the development of computer courses. Since the computer can store and analyze student performance data, course revision may then be undertaken from a factual base. The combined success of the retrial and reread techniques also give promise of economy in computer course preparation.

The success of the retrial feature is intriguing. With no additional assistance, students, in a majority of cases, corrected their own answers after simply knowing that their original responses were incorrect. Is this simply due to more intense student attention at these points, or to a reorganization of his thinking in the light of the knowledge that his original response was wrong? In this case the student had recently read relevant material, but to what extent, if any, would such a retrial affect the validity of questions in standard tests? Tests administered by computer could easily incorporate

such a feature.

The computer-taught students received a large step presentation, followed by questions and by immediate effective correction procedures. They did significantly better than those who received either a large or a small step presentation, followed by questions and simple immediate knowledge of the correct result. This suggests that for the learning of this type of material it is the effective correction procedure, rather than the initial step size or simple immediate feedback that is crucial.

The computer can apparently successfully perform at least partially some of the instructional functions. It can perform the function of presenting new material by providing a controlled presentation of subject matter. It can perform a testing function by continually examining the student concerning his comprehension of the material just studied. It can perform a guidance function, by analyzing such test results and then making decisions concerning future presentations to the student. It can perform a remedial function by identifying the areas of need for such special attention, and presenting the student with drill problems and appropriate small step learning materials. All of these functions may be performed on an individual basis for each student. As in other fields, the computer in instruction can partially perform functions that previously could be performed, if

at all, by human effort alone.

To the extent that computer-assisted instruction can be of aid to the teacher in the performance of these functions it could be used to make teacher time available for a more intimate interaction with the student. As in the normal strategy of this computer course, the teacher could be relieved of the major part of the functions of making group presentations of new course material, of evaluating the group's understanding of the course concepts, and of correcting the group's misconceptions. He could then spend correspondingly more time working individually with students who were in real difficulty in identified problem areas, and would have the opportunity to work on a personal basis with the able students who finished the course quickly.

An exploratory experiment of this nature can only carry one a limited distance in establishing the effectiveness of computers in instruction, or the advantage of any particular strategy of computer use. It is hoped that this study will stimulate research by others in the use of computers in the instructional process, for it is possible that such research could result in a marked difference in our ability to meet ever more difficult educational objectives.

Footnotes

1. The author wishes to gratefully acknowledge the excellent work of Mr. R. D. Hartz in entering the course material into the computer and debugging the program, and the efforts of Miss Barbara Koch in coding the computer course material.

References

1. McCracken, D. D. A guide to FORTRAN programming. New York: Wiley, 1961.

2. Plumb, S. C. FORTRAN - self teaching. (3rd ed.) Poughkeepsie, New York: IBM, 1963.

Printed in U.S.A.