DOCUMENT RESUME

ED 089 788 08 IR 000 513

AUTHOR Mitzel, Harold E.; Brandon, George L.
TITLE Experimentation with Computer-Assisted Instruction in Technical Education. Semi-Annual Progress Report.
INSTITUTION Pennsylvania State Univ., University Park. Computer-Assisted Instruction Lab.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO PSU-CAI-11
BUREAU NO BR-5-0035
PUB DATE 30 Jun 68
CONTRACT OEC-5-85-074
NOTE 62p.; For related documents see IR 000 511 and 512

EDRS PRICE MF-$0.75 HC-$3.15 PLUS POSTAGE
DESCRIPTORS Aptitude Tests; *Computer Assisted Instruction; Computer Graphics; Computer Science; *Curriculum Development; Demonstration Projects; Educational Research; Individualized Instruction; Instructional Design; On Line Systems; *Post Secondary Education; *Technical Education; *Vocational Education
IDENTIFIERS General Aptitude Test Battery; Geometric Dictionary; IBM 1500 Computer System; Pennsylvania State University; Prior Knowledge

ABSTRACT

A series of five reports is presented which describes the activities carried out by the Pennsylvania State University group engaged in research in computer-assisted instruction (CAI) in vocational-technical education. The reports cover the period January 1968-June 1968 and deal with: 1) prior knowledge and individualized instruction; 2) numerical and verbal aptitude tests administered at the CAI station; 3) an experimental procedure for course revision based on students' past performance; 4) the development of a geometric dictionary for the IBM 1500 computer system; and 5) a processor for multiple numeric entries. Also provided is a preview of activities for fiscal year 1969. Curriculum development activities are described as focusing on the development of communications skills for the vocational student and the creation of CAI physics material. Research efforts are listed as centering upon: 1) the development of a computer-based sequential intelligence test; 2) adaptation of the General Aptitude Test Battery to on-line presentation; and 3) replication of prior studies contrasting the effects of gradient and full-response feedback on immediate learning and retention. (Author/PB)

C·A·I

COMPUTER ASSISTED INSTRUCTION LABORATORY

COLLEGE OF EDUCATION · CHAMBERS BUILDING

THE PENNSYLVANIA STATE UNIVERSITY · UNIVERSITY PARK, PA.

EXPERIMENTATION WITH COMPUTER-ASSISTED INSTRUCTION IN TECHNICAL EDUCATION

SEMI-ANNUAL PROGRESS REPORT

JUNE 30, 1968

Report No. R-11

Note to accompany the Penn State Documents.

In order to have the entire collection of reports generated by the Computer Assisted Instruction Lab. at Penn State University included in the ERIC archives, the ERIC Clearinghouse on Educational Media and Technology was asked by Penn State to input the material. We are therefore including some documents which may be several years old. Also, so that our bibliographic information will conform with Penn State's, we have occasionally changed the title somewhat, or added information that may not be on the title page. Two of the documents in the CARE (Computer Assisted Remedial Education) collection were transferred to ERIC/EC to abstract. They are Report Number R-36 and Report Number R-53.

— Jaclyn Caselli / ERIC

The Pennsylvania State University

Computer Assisted Instruction Laboratory
University Park, Pennsylvania

Semi-Annual Progress Report

EXPERIMENTATION WITH COMPUTER-ASSISTED INSTRUCTION IN TECHNICAL EDUCATION

Contract No. 5-85-074
Project No. 5-0035

Principal Investigators
Harold E. Mitzel   George L. Brandon

Associate Investigators
Keith A. Hall, Laboratory Director
Betta H. Kriner, Administrative Assistant

Donald W. Johnson   Joseph L. French
Bobby R. Brown   Terry A. Bahn

Contributing Authors and Researchers
Karl G. Borman   John Tardibuono
Paul V. Peloquin

June 30, 1968

Report No. R-11

FOREWORD

The reports on the following pages comprise the sixth semi-annual description of the activities of the Penn State group engaged in research on computer-assisted instruction in vocational-technical education. To the reader who has not seen the previous five reports, this one will seem inexplicably uncoordinated because it reports on research and development carried out on two different computer configurations. In fact, the period between January 1, 1968 and June 30, 1968 has been a time of transition--in hardware, from the IBM 1410 to the IBM 1500, and in author language, from Coursewriter I (Yorktown Heights version) to Coursewriter II. Our efforts during the period have gone into the completion of experimental studies begun on the typewriter terminal, the translation of carefully selected course material from the old author language into a new and more powerful medium for reaching learners, and the development on the part of the staff of new skills and new teaching strategies for the richer learning environment provided by the 1500 system.

During the fiscal year 1969, we expect to expand our curriculum development efforts in communications skills for the vocational student. A simulated physics laboratory has been outlined and partially completed. Emphasis in this inquiry-oriented physics material is on the understanding of basic physical laws that are required by the post high school technical education student. In addition, in the curriculum development segment for the fiscal year 1969, we will expand upon the material from a pilot study in computer-assisted occcupational information completed last year with funds provided by the Pennsylvania Department of Public Instruction.

Major research efforts for 1968-69 are planned around 1) the development and evaluation of a computer-based sequential intelligence test, 2) adaptation and appraisal of the General Aptitude Test Battery to "on-line" presentation, 3) replication of prior studies contrasting the effects of gradient and full-response feedback on immediate learning and on retention.

In addition to the major curriculum development and research studies, we expect to continue dissemination about potential applications of CAI in vocational education and about computer systems innovations on the IBM 1500.

Harold E. Mitzel
University Park, Pennsylvania
July 1, 1968

CONTENTS

PRIOR KNOWLEDGE AND INDIVIDUALIZED INSTRUCTION

Bobby R. Brown and Terry A. Bahn

One of the often cited advantages of computer-assisted instruction (CAI) is the capability it affords for altering the nature and flow of instructional material presented to individual students. With CAI it is possible to adapt instruction to the needs of each individual student as reflected by measures of the student's past performance, by the student's differential aptitudes, or by some combination of both. CAI presently provides a considerable degree of individualization of instruction by allowing each student to proceed at his own rate and by providing remedial branching within "main track" programs. This individualization of instruction is achieved through the use of response sensitive programs which make little use of measures of individual differences or of the student's history prior to his beginning the program.

There are, as Cronbach (1967) has pointed out, two ways of adapting instruction to individual differences: "1) provide remedial adjuncts to fixed 'main track' instruction, and 2) teach different pupils by different methods." A search of the literature reveals a general absence of prescriptive statements concerning how to "teach different students by different methods." A comparison of Smallwood's assessment of the situation in 1962, with conclusions drawn by Gilman and Gargula five years later, is indicative of the relative lack of progress in this area.

> At this point we must rely on our intuition to convince ourselves that a system capable of making systematic changes in its presentation of material and in its internal decision process must be a potentially better teaching device than one without these advantages. The ultimate answer lies at the bottom of a mass of as yet uncollected data. (Smallwood, 1962, p. 108.)

> The results of this study indicate that if branching is to be used to advantage in computer-assisted instruction, there must be a thorough investigation of those situations where it facilitates learning. Also, research needs to be implemented to determine the criteria for branching decisions. (Gilman and Gargula, 1967.)

Cronbach (1967) has suggested that "we ought to take a differential variable we think promising and design alternative treatments to interact with that variable." In line with this suggestion, and in an effort to achieve the as yet idealized goal of "optimal learning" for each student, more research in CAI needs to be directed toward the accomplishment of three highly interrelated tasks:

1. The delineation of valid and potent individual differences among students, on the basis of which differential treatments will be of consequence in affecting student performance.
2. The design of a repertoire of instructional strategies by which material can be presented.
3. The formulation of models or decision rules for pairing student and strategy such that each student receives the instruction best suited to him.

This paper reports the results of an experiment designed to investigate a possible technique for adapting to individual differences, specifically for adapting instructional methods to the extent of prior knowledge.

One of the most obvious ways in which students differ is in the extent of their prior knowledge of the subject matter to be taught (Glaser, 1968). The extent of prior knowledge has received a considerable amount of attention from experimenters, not, however, as an individual difference variable but as a troublesome source of experimental error. In their efforts to control this source of experimental error, experimenters have employed pretests as screening devices, or have used esoteric or contrived subject matter about which subjects are unlikely to have any prior knowledge (or interest). An example of one ingenious and effective means of assuring the naivete of subjects appears in a study by Merrill and Stolurow (1966) in which students were taught an "imaginary" science.

While the experimenter who is not interested in studying this variable is justified and perfectly correct in controlling it in his experiment, extra-experimental applications of CAI make its control impractical if not impossible. So despite the fact that most experimenters have chosen to consider extent of prior knowledge a source of experimental error, either to be ignored or controlled, the fact remains that, wi' . long-range routine use of CAI, provision will have to be made for teaching students who differ in extent of prior knowledge of the subject matter to be taught. Because of different past educational experiences and differential forgetting, students will likely vary not only in amount but also in specific areas of prior knowledge.

Programs which have been developed using only naive students may not be at all satisfactory for use with students who have various amounts of prior knowledge of the subject being taught. In evaluating a course with all naive subjects two measures are commonly employed: time on program and amount learned. These two are often combined into a single derived score of the form, "amount learned/time = efficiency score." However, if one is concerned with achieving criterion performance, that is some fixed, predetermined level of proficiency, the only measure of concern should be "time to criterion." An efficiency score established with naive students may be totally misleading, and the program may be very inefficient for teaching students with prior knowledge of the subject matter, unless provision is made for skipping past the material which the students have already mastered. In addition to being inefficient, the program may have detrimental effects; for example, students who are required to proceed frame by frame through material which they have already mastered may become bored and disinterested with the entire program.

If a program is to be responsive to the needs of students who have various amounts of prior knowledge, it should possess the following:

1. a means of assessing each student's knowledge of each concept or sub-concept prior to instruction;
2. a means of skipping past material which the student has already mastered;
3. a means of providing a rapid review of the material about which the student has some prior knowledge;
4. a means of providing instruction on the material for which the student has little or no prior knowledge.

An effort was made to provide these four capabilities in a CAI program.

Three separate techniques were employed in an effort to prepare a program which would be adaptive to the needs of students possessing various amounts of prior knowledge. The reasoning behind the selection of each of these techniques is covered in the next section.

Rationale

In the research literature on size of step as a program variable, the most consistent finding has been that students proceed more rapidly through large step programs. (Coulson and Silberman, 1959, 1960; Evans, Glaser, and Homme, 1959; Shay, 1961; Hamilton and Porteus, 1965.)

It was decided therefore to use large step programing as a means of providing a rapid overview or review of the material to be taught.

Evaluation of the student's knowledge is bound up with two desirable but mutually exclusive characteristics. One would like to have detailed and reliable information concerning the student's prior knowledge. Also, since the information is to be used to save instruction time, it is necessary that the evaluation of the student's knowledge not be so time consuming that saving in instructional time is negative. If the repetitive evaluation of the student's knowledge is considered within the framework of information exchange, an alternative to making the evaluation time longer is to enrich the flow of information within a given interval. Other things being equal, multiple choice questions take less time than do constructed response questions, but multiple choice questions may not provide as much information as do constructed response questions. However, multiple choice questions with carefully constructed choices, coupled with a modified form of responding, may provide a rich flow of information without sacrificing the time-saving characteristic of multiple choice tests.

Consider what a student does when confronted with a multiple choice question. He considers the choices given with the knowledge that one of the choices is the correct answer. This automatically rules out many incorrect answers the student might have given to a constructed response form of the same question. The student may next rule out one or more of the choices as definitely incorrect. From the remaining choices, the student selects the most likely answer. Any uncertainty on the part of the student as to which answer is correct is lost in this form of responding. All the information is binary, the student is either right or wrong. However, if instead of choosing one correct answer to a multiple choice question the student is instructed to state his subjective probability (De Finetti, 1965) or degree of certainty for each choice, the amount of information which can be obtained from a multiple choice item can be increased. (See Shuford, Albert, and Massengill [1966] on this mode of response.) For example, suppose a student is presented with a multiple choice item for which he is certain that choices C and D are incorrect, and he is 70% sure choice A is correct, yet he has a 30% degree of belief that choice B may be correct. In the normal mode of responding the student would choose answer A and the fact that he had some belief in answer B would be lost. However, if the student were responding by stating his subjective probability for each response, this uncertainty could be expressed. This form of responding which allows the S to express his present state of knowledge for a test item was employed in embedded tests of the Ss' prior knowledge of the concepts on which instruction was to be presented. In order to process responses in this form on the IBM 1410 system it was necessary to write a rather lengthy subroutine. For a description of this subroutine, see pages 55, 56, and 57 of this report.

Typical small step, high response rate programing was employed to instruct students whose performance on the embedded tests indicated they were in need of instruction or remediation on some concept or subconcept.

A description of the integration of the three techniques in the experimental program is presented in the section on materials and is shown in Figure 1 of this report.

Method

Subjects

Sixty-five volunteers from an introductory educational psychology course at The Pennsylvania State University served as subjects.

Materials

Two CAI programs were used. One program consisted of a section of a modern mathematics (MMS) course developed at the Computer Assisted Instruction Laboratory of The Pennsylvania State University.[1] The section selected

[1]The writers would like to thank Professor C. Alan Riedesel and Professor Marilyn Suydam, who developed the modern mathematics program.

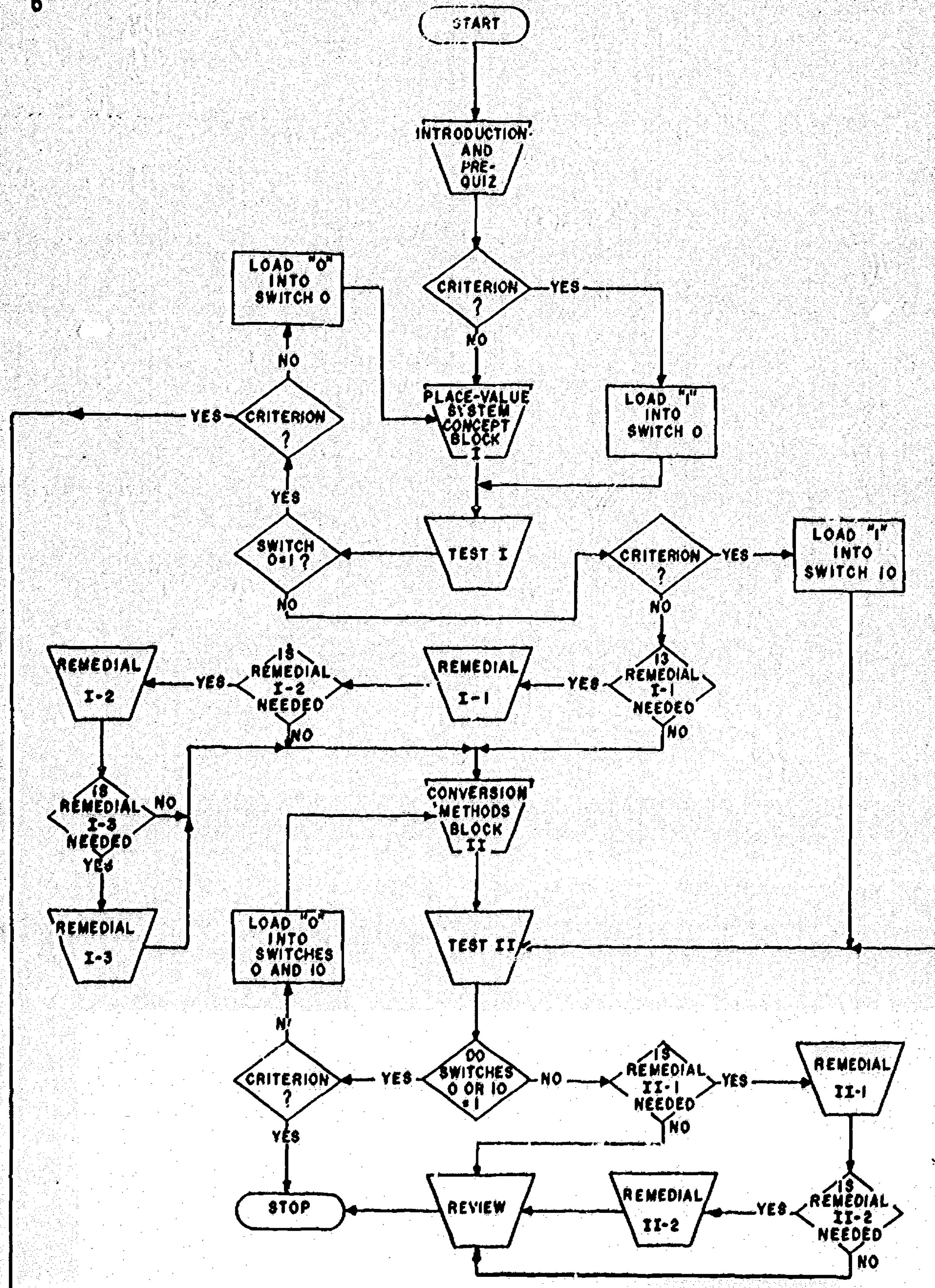


Fig. 1. A description of the integration of the three techniques in the experimental program.

consisted of instruction on base conversion of number systems with bases other than base ten. This program is essentially linear with remedial loops. The course from which this section was taken has been rather extensively evaluated and revised on the basis of student performance records, and there is evidence that it produces satisfactory criterion performance. The minimum number of responses possible to complete this program segment was 61, the maximum number of frames was 107 with repeated responding possible on some frames.

The second program (EXS) consisted of instruction on the same concepts presented in the first program. In designing this program an effort was made to implement a strategy which would be responsive to individual differences in amount of prior knowledge of the subject matter. Specifically, the program had the following characteristics: 1) large steps, 2) embedded tests of student performance, and 3) "skip ahead" and remedial branching. The minimum number of responses possible to complete this program was 19, the maximum number of frames was 81.

There were two (large step) blocks of instructional material in the program, one dealing with the place value system and the other with algorithms for converting from one number base to another (see Figure 1).

Three tests were embedded in the program. The purposes of the tests were to assess the student's performance, diagnose specific problems the student might be having and thereby provide information on the basis of which the student was either branched ahead or branched to remedial material. These tests consisted of multiple choice questions to which the student responded by stating his subjective probability or degree of belief for each of the choices.

Procedure

Subjects were randomly assigned to each of the two programs, 33 Ss were assigned to EXS and 32 to MMS. Subjects were run individually on IBM 1050 terminals. Each subject was administered a 22-item pretest immediately before going on the program, a 22-item posttest, and an 11-item transfer test immediately following his completion of the program, and a 22-item retention test one week later. The pre-, post-, and retention tests specifically included conversions from one base to another, which were taught in the program. The transfer test required the Ss to perform addition and subtraction in bases other than base ten. All four tests required constructed responses. The reliability estimates for the respective tests by Kuder-Richardson Formula 20 are as follows: pretest, .93 (from a prior study; Wodtke et al., 1967); posttest, .93; transfer test, .93; retention test, .96.

Results

Students who received non-zero scores on the pretest or who had received prior instruction on numbers systems other than base ten were categorized as having prior knowledge (PK) of the content. Students who scored zero on the pretest and had received no prior instruction on numbers systems other than base ten were categorized as having no prior knowledge (NPK) of the content. Performance data, consisting of posttest, transfer test, and retention test scores along with instructional time, were analyzed within a 2 x 2 analysis of variance design. One factor consisted of program (MMS or LXS), the other of extent of prior knowledge (PK or NPK).

The results of the analysis of variance of posttest scores for the PK and NPK groups by program are shown in Table 1. Neither of the main efforts were significant; however, there was a marginally significant interaction between extent of prior knowledge and program taken ($P < .10$). The means, standard errors, and n for this analysis are given in Table 2. The marginally significant interaction is shown in Figure 2.

Table 1

Analysis of Variance of Posttest Scores for PK and NPK Groups by Program

Source	d.f.	M.S.	F	P
Program	1	23.01	<1.0	
Prior Knowledge	1	72.38	1.84	>.10
Interaction	1	126.11	3.21	<.10
Residual	61	39.25		

Table 2

Means, Standard Errors, and n for the Posttest by Program and Extent of Prior Knowledge

Group	EXS	MMS
PK	$\bar{X} = 17.38$ $\sigma_{\bar{X}} = 1.74$ n = 13	$\bar{X} = 15.73$ $\sigma_{\bar{X}} = 1.89$ n = 11
NPK	$\bar{X} = 12.30$ $\sigma_{\bar{X}} = 1.40$ n = 20	$\bar{X} = 16.43$ $\sigma_{\bar{X}} = 1.37$ n = 21

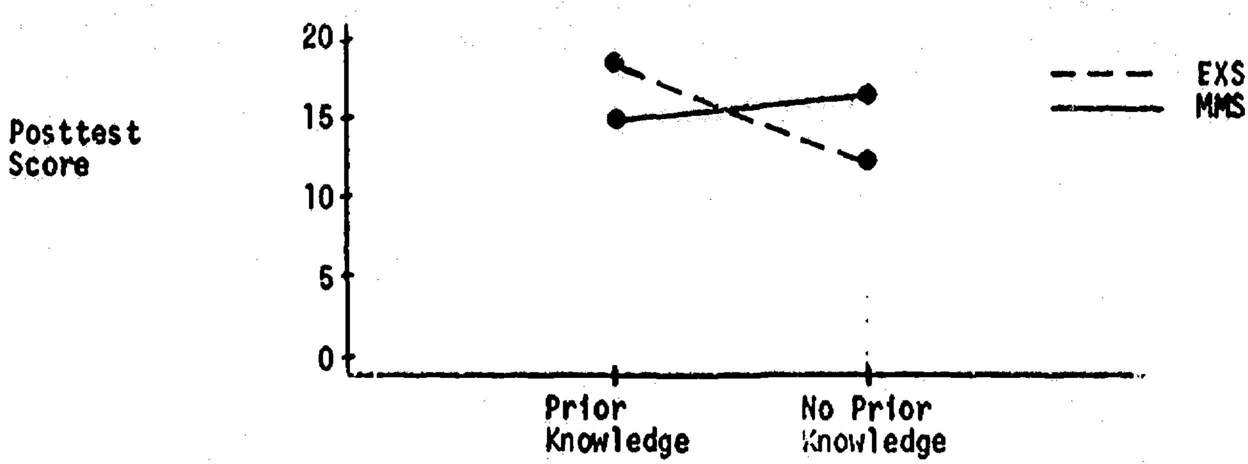


Fig. 2. Plot of the means of the posttest scores for PK and NPK Ss on MMS and EXS.

The results of the analysis of variance of retention test scores for the PK and NPK groups by program are shown in Table 3. The results of this analysis are very similar to the results of the analysis of posttest performance. The main effect for programs was not significant. The main effect for extent of prior knowledge was significant (P <.05); however, as on the posttest, there was a marginally significant interaction between extent of prior knowledge and program taken (P <.10). The means, standard errors, and n for this analysis are given in Table 4. The interaction is plotted in Figure 3.

The results of the analysis of variance for the transfer test scores are shown in Table 5. There were no significant differences in retention test performance.

Table 3

Analysis of Variance of Retention Test Scores for PK and NPK Groups by Program

Source	d.f.	M.S.	F	P
Program	1	41.80	1.01	>.10
Prior Knowledge	1	234.64	5.69	<.05
Interaction	1	134.51	3.26	<.10
Residual	58	41.25		

Table 4

Means, Standard Errors, and n for the Retention Test by Program and Extent of Prior Knowledge

Groups	EXS	
PK	$\bar{x} = 19.17$ $\sigma_{\bar{x}} = 1.85$ n = 12	$\bar{x} = 17.80$ $\sigma_{\bar{x}} = 2.03$ n = 10
NPK	$\bar{x} = 12.00$ $\sigma_{\bar{x}} = 1.47$ n = 19	$\bar{x} = 16.81$ $\sigma_{\bar{x}} = 1.40$ n = 21

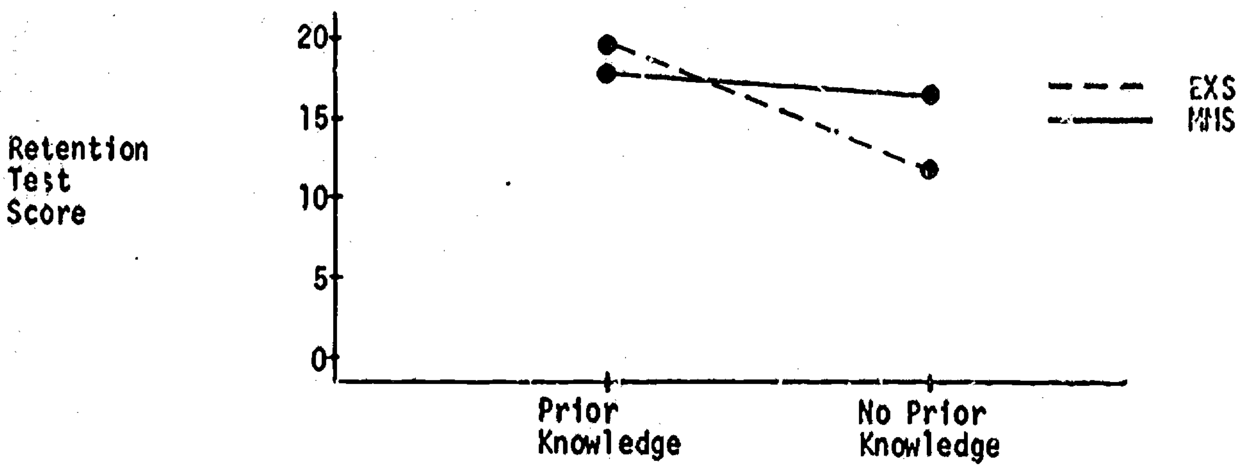


Fig. 3. Plot of the means of the retention score for PK and NPK Ss of MMS and EXS.

Table 5

Analysis of Variance of Transfer Test Scores for PK and NPK Groups by Program

Source	d.f.	M.S.	F	P
Program	1	4.23	<1.0	
Prior Knowledge	1	25.49	1.67	>.10
Interaction	1	2.24	<1.0	
Residual	57	15.23		

The analysis of variance results for instructional time are presented in Table 6. There was a significant main effect for programs (P <.05). The main effect for extent of prior knowledge as well as the interaction was non-significant. The means, standard errors, and n for this analysis are given in Table 7. The means are plotted in Figure 4.

Table 6

Analysis of Variance of Instructional Time for PK and NPK Groups by Program

Source	d.f.	M.S.	F	P
Program	1	4018.89	4.22	<.05
Prior Knowledge	1	2800.07	2.94	>.10
Interaction	1	982.20	1.03	>.10
Residual	61	952.99		

Table 7

Mean Instructional Time in Minutes, Standard Error, and n by Program and Extent of Prior Knowledge

Group	EXS		MMS	
PK	$\bar{x} = 69.97$		$\bar{x} = 94.38$	
	$\sigma_{\bar{x}} = 8.56$	n = 13	$\sigma_{\bar{x}} = 9.31$	n = 10
NPK	$\bar{x} = 91.68$		$\bar{x} = 99.94$	
	$\sigma_{\bar{x}} = 6.90$	n = 20	$\sigma_{\bar{x}} = 6.74$	n = 21

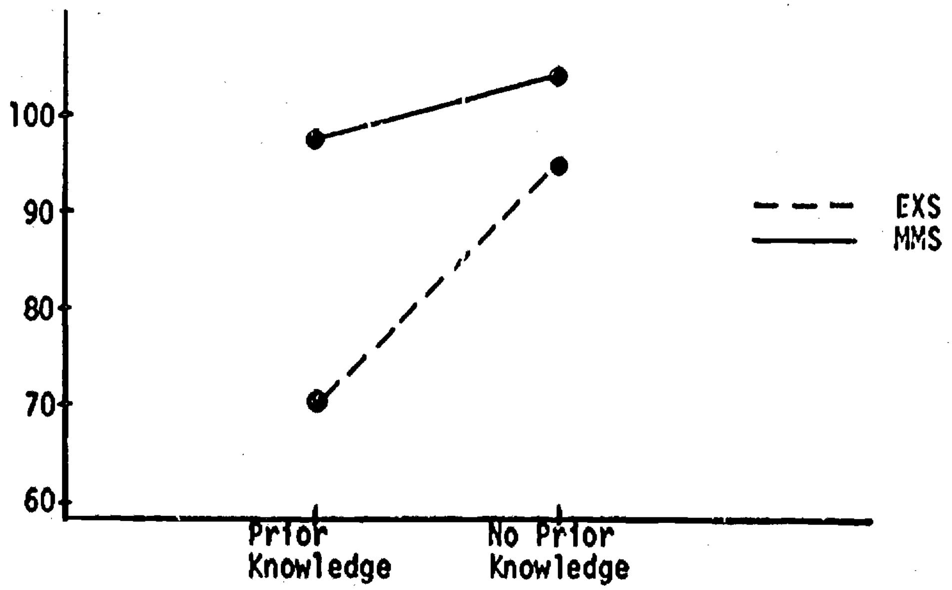


Fig. 4. Plot of mean instructional time for PK and NPK Ss on MMS and EXS.

Because of the marginally significant interaction between program taken and extent of prior knowledge for posttest and retention test data, the simple main effects were calculated. From analysis of the simple main effects, the following pattern emerges. PK Ss on EXS do significantly better than NPK Ss ($P < .05$) for the posttest, and ($P < .01$) for the retention test. Ss on MMS having PK do not do significantly better than Ss having NPK ($F < 1.0$).

The simple main effects of PK - NPK across MMS - EXS are as follows: for Ss having PK there were no significant differences in performance on posttest or retention test attributable to the instructional programs ($Fs < 1.0$). Ss having NPK did significantly better on MMS on both posttest ($P < 1.0$) and retention test ($P < .10$).

The EXS program seems to have capitalized on the knowledge which Ss had prior to instruction. The finding of no significant differences between MMS and EXS on posttest, transfer test, and retention test, coupled with the time saving for EXS, suggests that students with prior knowledge would benefit by having instruction on EXS rather than MMS. For NPK Ss the lower posttest and retention test scores on EXS seem to call for MMS for these Ss in spite of the time saving on EXS.

Discussion and Conclusions

The procedure reported here seems to provide a means of adapting to extent of prior knowledge which results in considerable time saving with no decrease in criterion performance. The results of this study also suggest that neither of the programs could be recommended for all students if they vary widely in extent of prior knowledge. Perhaps parallel programing employing the formats of both the programs with a branching procedure for switching students from one program to the other may provide the benefits of both.

The procedure employed in EXS may have somewhat limited application in terms of the instructional content. The content in this study was such that a student's prior knowledge of a concept could be evaluated. For content consisting of more or less discrete units of information, evaluation of prior knowledge by the method employed here may not be feasible.

References

Coulson, J. E. and Silverman, H. F. Effects of three variables on a teaching machine. Journal of Educational Psychology, 1960, 51, 135-143.

Cronbach, L. J. How can instruction be adapted to individual differences? In R. M. Gagne (Ed.), Learning and Individual Differences, Columbus, Ohio: Charles E. Merrill, 1967, pp. 23-39.

de Finetti, B. Methods for discriminating levels of partial knowledge concerning a test item. British Journal of Mathematics and Statistical Psychology, 1965, 18, 87-123.

Evans, J. L., Glaser, R., and Homme, L. E. A preliminary investigation of variation in the properties of verbal sequence of the "Teaching Machine" type. Teaching Machines and Programmed Learning, Washington, D. C.: National Education Association of U. S., 1960, 446-451.

Gilman, D. A. and Gargula, C. Remedial and review branching in computer-assisted instruction. In Semi-Annual Progress Report Experimentation with Computer-Assisted Instruction in Technical Education, Project No. 5-85-074 prepared by Harold E. Mitzel, et al., 1967.

Glaser, R. Adapting the elementary school curriculum to individual performance. Proceedings of the 1967 Invitational Conference on Testing Problems, Princeton, New Jersey: Educational Testing Service, 1960, 3-36.

Hamilton, N. R. and Porteus, B. D. Increasing Long-Term Retention of Knowledge. Experiment II: Instruction From Small Step versus Large Step Programs with Immediate versus Delayed Review Procedures for Students of Different Ability Levels. Palo Alto, California: American Institutes for Research, 1965, AIR-D49-7/65-FR(11).

Merrill, M. D. and Stolurow, L. M. Hierarchical preview versus problem oriented review in learning an imaginary science. American Education Research Journal, 1966, 3, 251-261.

Shay, C. B. Relationship of intelligence of step size on a teaching machine program. Journal of Educational Psychology, 1961, 52, 98-103.

Shuford, E. H., Albert, A., and Massengill, H. E. Admissible probability measurement procedures. Psychometrika, 1966, 31, 125-145.

Smallwood, R. D. A Decision Structure for Teaching Machines. Cambridge, Massachusetts: The M. I. T. Press, 1962.

Wodtke, K. H., et al. Scrambled versus Ordered Sequencing in Computer-Assisted Instruction. University Park, Pennsylvania: Computer Assisted Instruction Laboratory, 1967.

NUMERICAL AND VERBAL APTITUDE TESTS ADMINISTERED AT THE CAI STUDENT STATION

Joseph L. French and John Tardibuono[1]

Each year the youth of our country spend more time in psycho-educational testing than the previous year. Not only are more and more tests being used, but the length of testing time has increased as it has become apparent that short conventional tests are less reliable than long ones. Probably no single school-related activity is more distasteful to the average student than the experience of being tested. Even though a major part of this negative attitude is the result of anxiety aroused by the importance of the test, there remains a substantial portion which becomes associated with the experience itself. The picture of the anxiety-aroused student dropping papers and pencils, overlooking a page of items, looking at another student's paper, miscalculating his time, and answering numerous questions which are either too easy or too difficult, is a familiar one to most people. It is obvious that more of the above-mentioned problems can be attributed to both the physical aspects of the test environment and to the properties of the test itself.

The utilization of computer-assisted instruction (CAI) equipment for testing can do much to alleviate many of these problems. First of all, the CAI terminal provides a well-controlled environment with minimal distraction. Secondly, the branching capabilities of the computer appear well suited to the construction of a test with a format in which 1) each student attempts only those items which are most appropriate to his particular ability level, and 2) the time permitted for each item is the same for everyone. It was hypothesized that a program could be developed for such a test which would take the subject in relatively large steps to the area of his threshold of understanding, and in that area move the subject in small steps to the point where he is unable to answer items correctly. By minimizing the very easy and very difficult items and maximizing the number of items at one's thresh-hold of understanding, a highly reliable and valid test should result. In

[1]Special acknowledgement is due Mrs. Jacqueline Sallade, graduate assistant, and Mrs. Diane Knull, programer.

order for the procedure to be made operational it was necessary to assume that 1) items could be arranged in increasing order of difficulty and 2) the subject's score on such a test would define a point below which all eliminated items would be passed and above which all items would be failed.

The present project was an attempt to produce a test which, when compared with traditional tests, will 1) have fewer items for each examinee, 2) require less examiner time, 3) be more reliable, 4) be just as valid, and 5) be viewed with favor by examinees.

Because the building of a test suitable for administration via computer introduces many problems which are conceptually different from those involved in traditional testing, it seems more reasonable, at this time, to specify goals or objectives rather than formal hypotheses. In constructing a Computer Assisted Test (COMPAT) four goals were formulated. The purpose of this study was to arrange tests using multiple choice, numerical and verbal items, in a program so that:

(1) the average number of items attempted by each student will be significantly less than the number of items traditionally required by a test of similar content (in this case, the Henmon-Nelson Tests of Mental Ability);[1]

(2) the test will take significantly less time to administer than the traditional test;

(3) indices of reliability obtained by means of internal consistency formulas (KR 20) will be higher for COMPAT than for those reported in the manual for the Henmon-Nelson tests; and

(4) there will be a significantly high correlation between each COMPAT and the Henmon-Nelson Test, and between each COMPAT and student performance in other academic activities.

The various steps involved in this report will be described as follows: 1) description of the equipment, 2) selection of items and final format of COMPAT, 3) format of the COMPAT program, 4) sample characteristics, 5) procedure, results, and graphic description, 6) conversion to the IBM 1500 system, and 7) plans for future research.

---

[1]Used with permission of Houghton Mifflin Company.

Equipment

The basic equipment consisted of two units: a 1410 IBM computer located a few blocks from the CAI Laboratory and an IBM 1050 terminal composed of a standard Selectric typewriter keyboard and, under program control, an 80 slide random-access carousel projector with a rear image screen. Four terminals were available for subjects

Development of the Tests

The process of item selection necessary to implement the above-mentioned testing conditions included several stages. In order to minimize the initial problems of item validity and level of item difficulty, permission was obtained from the Houghton Mifflin Company, publishers of the Henmon-Nelson Tests of Mental Ability, to use many of the Henmon-Nelson items for this study. The Henmon-Nelson items are predominantly verbal and numerical and appear in conventional spiral-omnibus format in two parallel forms at each academic level. To obtain a large pool of items for separate numerical and verbal tests suitable for a Vocational Technical school population,items were selected from the High School Form A, High School Form B, College Level Form A, and College Level Form B tests. The result was a pool of more than 200 verbal items and about 100 numerical items. Although each of the four tests presented the items in order of increasing difficulty, the amount of overlap between the high school and college levels was not known. The four forms were administered to approximately 100 high school students from several communities to further determine the difficulty level of each item.

The next step involved a cross check of the order of difficulty. The verbal items were rearranged in agreement with the obtained difficulty levels and administered to 53 Penn State undergraduates. The same procedures were followed with the numerical questions. Some additional re ordering of items followed.

Utilization of the IBM 1410 computer system imposed a limitation of 80 items for presentation via a carousel slide projector; a limitation necessitated by virtue of the study which preempted the loading of additional trays. The complexity of the branching program served to limit the number of items as well.

From the above-mentioned Henmon-Nelson items, seven tests were generated. They were as follows:

1. Pretest

   Approximately every eighth item from the pool of 229 verbal items was extracted and used to construct a 25-item pretest. It was orginally planned to administer this test via the computer, but preliminary investigations indicated that such a procedure involved inefficient use of machine time. This test was subsequently administered in paper and pencil form to determine which of tests Verbal A, B, C, or D should be administered.

2. Verbal A-B-C-D

   From the 204 verbal items remaining after 25 had been used in the pretest, four 80-item tests were formulated: Test A contained the easiest 80 items from the ordered pool; Test B included the last 38 items from Test A plus 41 of the next easiest items; Test C was composed of the last 38 questions from Test B plus the next 42 items; Test D consisted of the last 39 items from Test C and the rest of the 204 questions. Which 80-item test was administered depended on the score obtained on the pretest.

3. Verbal 80 (V-80)

   In order to test the possibility that fewer items, with greater difficulty intervals would yield results equivalent to use of the above tests, a third verbal test was developed. In this case, every fourth item (in difficulty level) was used to construct an 80-item verbal test. Items in the V-80 test also appeared on the four tests described in 2 above.

4. Numerical 80 (N-80)

   From the pool of numerical items, 80 of the most difficult ones were used to form the numerical test.

Format of the COMPAT Program

The program used to present COMPAT was written in Coursewriter author language. The complexity of the program is reflected in the fact that it includes about 4425 statements and required 11 hours to compile. However, the branching strategies within a single program accommodated six different tests and required merely the selective use of a specific tray. The examiner installed the tray with one of the six tests and then signed the subject onto the system with a pre-registered student number. In order to change tests it was necessary both to sign the student onto the system with a new student number, and to change slide trays. The sequences of item presentation were as follows:

Each subject started with the eighth item and was presented every eighth item up to question 80, provided no mistakes were made, at which point he was assigned a score of 80 and signed off. However, most students are not able to answer all questions correctly. When an error is made, the branching capabilities of the computer are utilized. As soon as the first error was made, the system backed up five items and if the student correctly answered that item, the program moved ahead two items at a time; if a second error was made, the system again backed up five items and then proceeded forward, one question at a time. The system continued in that fashion until the subject missed four out of the last seven items presented. A score was then assigned based on the highest item number answered correctly minus the number of items attempted but missed below that point. If at any point during the test, the sequence reached an item which had been previously presented, the adjacent item was presented, depending on the direction in which the program had been moving orginally. For example, if a subject had failed one or more items, and if the program was now moving forward and reached item 54 which had already been presented, item 55 would be given. If the subject misses 59, the next item should be 54; but if 54 has been attempted, then item 53 or the highest item not attempted below 54 would be given. From item 54, the system would back up one item at a time until an unattempted item was found. Each item not answered in 60 seconds was considered incorrect, but all the items remained on the screen until an answer was given. The decision to use a 60-second time limit was based on pilot studies. (Further research is needed on time limits.) Examples of actual patterns of response can be found on page 26.

The Sample

One of the practical limitations encountered in the first phase of this project was the availability of subjects of the appropriate age and educational level willing to take time for research projects. Since this project focuses on the feasibility of developing short, highly reliable tests, vocational-technical school subjects were not sought. The 73 subjects used in this reporting period included high school juniors and seniors, housewives, and undergraduate and graduate students from Penn State. It should be noted that 80 subjects took the various tests. However, systems difficulties and scheduling problems for subjects reduced the number of subjects for various computations and analysis. Good data were available for all tests for 67 subjects.

In Table 1 can be found data from the standardization of the College Level forms of the Henmon-Nelson and for the subjects used in COMPAT studies responding to Form B. It becomes apparent from Table 1 that in each instance the sample for this project obtained higher scores and exhibited greater variability than the comparison samples. As can be seen in Table 2, the correlations between the numerical and verbal and total test scores are remarkably similar for a random sample of 200 subjects from the Henmon-Nelson College Level standardization population and for COMPAT when used with the sample described.

Table 1

Comparison of Verbal, Quantitative and Total Scores from Henmon-Nelson Standardization Data with Scores From the Project Sample on College Level Form B

	Henmon-Nelson Sample of College Freshman[a]				Project Sample	
	N = 148		N = 95		N = 73	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Q	19.62	5.89	19.36	6.34	24.08	8.65
V	28.86	10.47	27.57	10.17	38.59	12.74
T	48.48	13.60	46.93	14.24	62.60	19.39

[a]The Henmon-Nelson Tests of Mental Ability Examiner's Manual; 1961, p.15.

Table 2

Interrelations Between Quantitative and Verbal and Total Raw Scores

	Henmon-Nelson N = 100		COMPAT N = 73	
	V	T	V	T
Q	.587	.827	.66	.87
		.941		.94

Procedure

Each subject took five tests: 1) the Henmon-Nelson College Level, Form B, 2) Verbal 80, 3) Numerical 80, 4) Pretest, and 5) Verbal A, B, C, or D. The directions for the Henmon-Nelson are contained in the test booklet which each subject read himself. For the COMPAT tests, each S was shown how to operate the terminal, how to respond, and how to correct a mistake. He was instructed as follows:

> You will be shown questions, either verbal or mathematical on the screen at your left. As soon as you have the answer, type the number indicating that answer, then press EOB; the next question will then be shown. If you wish to change an answer before you press EOB, follow the procedure previously described. Remember once you press EOB you will not be able to change your response. There is a time limit for each question; but if you work steadily, you need not rush. I will be in the next room if any complications arise. Any questions? You are now ready to begin.

Due to the length of time necessary to complete all five tests, most of the subjects returned for a second session. The availability of computer time as well as consideration for the subject's schedule made exact intervals between sessions impractical. Each subject was paid a flat rate of $3, unless system failures necessitated returning for a third session in which case payment was made at the rate of $1.25 per hour.

As each S entered the terminal room, he was given either the Henmon-Nelson College Level, Form B, the pretest, or assigned to one of the computer-administered tests. Due to the varying lengths of time subjects spent on each of the tests, it was not practical to randomize the order of the tests.

Post-facto examination revealed the order of administration described in Table 3. The effect of order was not determined in the present investigation. (Further studies will be designed to account for the effect of order.)

Table 3

Number of Subjects and Order of Test Administration

	First	Second	Third	Fourth	Fifth
Verbal 80	18	41	4	4	9
Numerical 80	16	11	40	3	6
Pretest	4	6	13	53	0
Verbal A-B-C-D	0	4	9	7	56
Henmon-Nelson B	38	14	10	9	5

It should be noted that the method of item selection, involving replication of certain items across the different tests, complicates the interpretation of results. The responses to questions which appeared twice were relatively stable, as can be seen in Table 4. Mention should be made of the relatively high percentage (22) of items answered correctly on the Henmon-Nelson Form B yet missed on the COMPAT version. One reason for this occurrence may be that many of those items were scored as incorrect because the 60-second time limit had been exceeded.

The above observation becomes crucial in view of the actual influence of the time factor in the present test. Inspection of the data revealed that on the Verbal test, 71 of the 117 overtime responses were correct; furthermore, on the Numerical test, out of 333 "timed-out" responses, 235 were actually correct.

Table 4

Relationship Between Responses to Items Which Appeared on Both the Henmon-Nelson College Level, Form B, and Two COMPAT Tests

Test	No. of items appearing twice	No. Responses made to those items	Per cent Correct on both	Per cent Wrong on both	Per cent Correct on HN-B but Wrong on CAI	Per cent Wrong on HN-B but Correct on CAI	Total Per cent Match
N-80	18	383	47	21	22	8	68
V-80	18	478	62	17	7	21	79

Results

To achieve goals 1 and 2 required the administration of relatively few items in a relatively short period of time when compared with the administration of the conventional format of the Henmon-Nelson. Table 5 reflects the data relating to these two goals. In conventional format the Henmon-Nelson test provides 100 items and allows no more than 40 minutes or working time. While not all subjects attempt the 100 items, it is assumed that as subjects run into difficult items they scan the remaining items in hopes that they can find some which they can answer. With the COMPAT procedures, the mean number of items attempted varied from 23 to 30 and the mean number of minutes varied from 21 to 31. Relatively large standard deviations can be observed for both number of items attempted and time spent on the test. It can be seen in Table 5 that the objectives of constructing a test with fewer administered items than on the criterion test was realized. When the numerical test and verbal test were added together, a mean of 57 items was obtained.

The time indicated in Table 5 was obtained by subtracting "sign on" time from "sign off" time. The time figures include system delays and program malfunctions. By analysis of the response latencies of a random sample of 25 subjects, it was learned that the time to complete the test is decreased significantly when only response latencies to each item are considered. (Translation of Compat to the IBM 1500 system, as described later in the report, is expected to decrease testing time due to system delay.)

Table 5

Means and Standard Deviations of Items Attempted on Compat Tests: Compared to Number of Items and Time Required for Henmon-Nelson, College Level, Form B

	HN-B	Verbal 80	Numerical	Verbal C	Verbal D
Mean No. of Items Attempted	100	29.5	26.5	22.9	24.5
S.D.	-	20.8	18.6	19.2	15.5
Mean Time on Test (in Minutes)	40	30.9	35.7	20.5	24.9
S.D.	-	24.3	24.9	14.8	18.5
N	-	67	67	17	41

Goal 3 pertains to reliability. At this stage of development with these tests, test-retest procedures over short periods of time did not seem appropriate. Statistical formulas requiring that the same items be presented to each subject were inappropriate also since the same pattern of items was presented to few subjects. Since great care was exercised in the selection of items arranged in order of difficulty, it was assumed that all items which were not administered and which were below the score would have been passed. It was further assumed that all items which were not administered and which were above the subject's score would have been failed. Since these assumptions are implicit in the computation of the score, it is reasonable to assume them also in the computation of reliability coefficients. (Such procedure is followed in the administration of such tests as the Stanford-Binet.) A demonstration of the patterns of actual response can be found in Figure 1.

Item No.	COMPAT Verbal S35	[illegible]	S69	COMPAT Numerical S10	S13	S80
1				+		
2				+		
3				+		
4				+		
5				+		
6				+		
7				+		
8	+	+	+	+	+	+
9				+		
10				(+)		
11				0		
12				+		
13				+		
14				0		
15				+		
16	+	+	+	0	+	+
17				0		
18				0		
19		+		+		
20			+	+		
21		+				
22			+	+		
23		+	+	+		
24	+	0	+	0	+	+
25		+	+	(+)		
26		+	+	(+)		
27		+	+	(+)		
28		0	+			
29		+	+			
30		+	+			
31		+	+			
32	+	+	+		+	+
33	0	+	+			
34		+	0			
35		+	0			
36		+	+			
37		0	+			
38	0	0	+			
39		0	+			
40	+	0	0		+	+
41			+			
42			+			
43	0		+			
44			+			+
45			+			+
46			+			+
47			+			+
48	0		+		+	+
49			0			+
50			+			+
51			+			+
52			+			+
53			+			+
54			+			+
55			+			+
56			+		+	(+)
57			+		+	+
58			+		0	0
59			+			+
60			+			(+)
61			+			0
62			0		(+)	+
63			+			+
64			+		+	+
65			+			0
66			+			+
67			+		(+)	+
68			+			+
69			+			(+)
70			+			+
71			0			+
72			+		0	+
73			0			+
74			+			(+)
75			0			+
76			0			0
77			0			0
78			+			0
79			+			0
80			0			0
Scores:	38	34	69	17	62	68

\+ correct response within time limit
(+) correct response in excess of time limit
0 incorrect response

Fig. 1. A demonstration of the patterns of actual response.

The patterns of response for six subjects (in Figure 1) were selected to show a variety of responses rather than to depict group tendencies. Subjects 29, 10, and 80 provide illustrations of the system working as designed. Of these, Subject 10 appears to be weakest in measured ability, and Subject 80 seems to be strongest. The influence of time in determining a score is indicated by plus signs in parentheses. It can be noted that Subject 10 and 80 correctly responded to four items for which they did not receive credit because they exceeded the time limit in making their response.

Subject 69 answered items 18, 16, 24, and 32 correctly, then failed items 40 and 35. Item 30 was answered correctly. Eventually, the subject tried all items between 20 and 80 before the test was discontinued. In this instance, little time saving was achieved.

Subject 35 and 13 present somewhat different problems. Subject 35 responded to items 8, 16, 24, 32, and 40 correctly. Then items 48, 43, 38, and 33 were failed, and the test was discontinued. With this pattern of response, one is not sure whether the score of 38 is a reliable one or not. It would have been desirable to see more of the student's ability on items 32 through 48.

A similar situation can be observed in following the pattern of response for Subject 13. The subject was correct on the first 8 items but failed item 72. Item 67 was then answered correctly but in more than 60 seconds. Item 62 was answered correctly,but in more than 60 seconds. Although item 57 was answered correctly item 58 was failed and the program was discontinued. The program is based on the assumption that the individual will, at some point, begin to fail items and will then pass and fail a block of items. A discontinuation after 4 of 7 items have been failed was selected because in many instances the individual will try a difficult item and fail it, an easy item and pass it, a difficult item and fail, an easy item and pass it, etc. The patterns of Subject 29 and 10 are typical. Confidence can be placed in the scores for Subjects 69 and 80 also.

In Table 7 can be found Kuder-Richardson Formula 20 reliability coefficients for the four COMPAT tests used with enough subjects to warrant computation. It should be noted that Kuder-Richardson formulas yield slightly higher correlations than other methods of measuring reliability and that the technique of assigning a minus or a plus to unanswered items will tend to inflate the obtained correlations. The Kuder-Richardson coefficients ranging from .977 to .988 compare favorably with the odd-even reliability coefficients of .94 and .95 reported in the technical manual for total scores of the college level Henmon-Nelson tests. The technical manual reports alternate form reliability coefficients of .84, .876, and .887 for Q, V, and total scores. Thus, it is concluded from the first phase of the reliability study that a highly reliable test can be adapted for presentation by a computer. Future studies will investigate other aspects of reliability.

Table 7

Means, Standard Deviations and KR20 Reliability Coefficients for Computerized Tests of Mental Ability

	Verbal 80	Numerical	Verbal D	Verbal C	Verbal B	Verbal A
Mean	44.66	35.56	54.67	39.52	40.5	16.3
S.D.	22.75	18.14	22.28	25.81	*	*
KR20 r	.986	.977	.985	.988	*	*
N	69	71	40	18	7	3

*not computed

The final area of investigation to be reported here involves a comparison between COMPAT and the conventional Henmon-Nelson test. These data can be found in Table 8. The relationship between COMPAT and the conventionally administered Henmon-Nelson test are not as substantial as those reported in the Examiner's Manual for alternate forms of the Henmon-Nelson. Future studies will deal with some of the issues raised by these figures and bear heavily on goal 4.

It is evident from the data in Table 8 that the normative data for the Henmon-Nelson can not be used to interpret COMPAT scores. Norms will need to be developed if COMPAT scores for individuals are to be interpreted.

Table 8

Correlations Between Scores Obtained on the Henmon-Nelson Test, College Level, Form B and Three Computerized Tests

	Henmon-Nelson				COMPAT	
	Total*	Verbal*	Quantiative*	Numerical*	Verbal C**	Verbal D***
Compat Verbal 80	.701	.760	.463	.352	.577	.751
H-N Total		.947	.861	.558	.762	.574
Verbal			.659	.419	.758	.648
Quantitative				.662	.548	.406
Compat Numerical					.415	.286

*N=67, **N=17, ***N=41

It should be noted that the number of subjects who took COMPAT Verbal-C or Verbal-D was dependent on their score on the pretest. Since COMPAT Verbal-A and Verbal-B were administered to only 10 subjects, correlational data for those tests were not computed.

Conversion to the 1500 System

Shortly after the present project became operational for experimental research, the IBM 1500 system became available. The 1500 provides several distinct advantages over the system used for the collection of data reported here: 1) presentation of items via a cathode ray tube (CRT) thus eliminating the necessity of changing slide trays and the limitation of 80 questions, 2) the CRT will black out when time limits have been exceeded, 3) greater flexibility in branching is available in the author language, 4) more terminals are available to permit a greater number of subjects to be tested simultaneously, and 5) the students will respond by using a light pen (LP) which will eliminate some typographical errors.

Another distinct advantage is that the processing capabilities of the IBM 1500 system is faster and less time is needed to search for items when the program is branching because the efficiency of the program is increased by using a Buffer to Return Register function which allows a counter to be loaded into a Buffer and a Buffer into a Return Register. This procedure is being used in conjunction with the Move function to branch to the label located in the Return Register. The programing described in this paragraph replaces the use of 80 conditional branch statements with branching on the basis of a counter to the next question.

Also used in the 1500 system is the Buffer to Counter function which converts the Buffer into Counters to store the question number if the subjects respond with a wrong answer or "time out"; but only if that question is numerically higher than the highest one they have answered correctly at that point in time does this event occur. When the student has missed four out of the last seven questions, this Buffer will be checked to see if the number stored in the counter within the Buffer is less than the highest one correct. With this procedure it is possible to know the number of questions missed below the highest one answered correctly. The COMPAT programs have been converted for use on the 1500 during this reporting period.

Future Research Direction

Now that the programs have been transferred to the 1500 system, it is more feasible to test the level of difficulty of each item on-line. A program is being written in which each item can be presented in such a way that the current order can be used without branching. It is important to have such a check since level of difficulty often changes with mode of presentation. To date, level of difficulty has been obtained by off-line testing. The next phase of the project will allow for a study of time spent per item. The 60-second time limit may be adjusted after reviewing these data.

Furthermore, during the next phase of the project, validity studies will be undertaken. Correlational data will be developed for the various COMPAT programs and grade-point average and other aptitude tests available for subjects. During this phase of the study it will be possible to ascertain the feasibility of retaining the long verbal test. So far it appears as if the 80-item verbal test has adequate reliability. If it is equal to the longer verbal test in concurrent validity, the longer test could be abandoned.

Another study will involve the development of test-retest reliability coefficients.

A study of reading ability on the CRT is under way. Each student participating in this study will respond to COMPAT V and COMPAT N. The relationship of reading ability to the test scores will be obtained.

An Experimental Procedure for Course Revision Based on Students' Past Performance

Karl G. Borman and Donald W. Johnson

In a previous report (Johnson and Borman, 1967) the authors reported a study in which they investigated the time required to complete a program under various modes of stimulus presentation (audio static, display, typewriter, and slide). No significant differences were demonstrated between the various modes of presentation because of the large variances obtained in the dependent variable for the various modes. At that time, it was the intent of the authors to analyze each subject's performance on each frame in order to determine whether or not there were certain frames in the program that contributed a large proportion of variance to the total within variance group. Table 1 contains the means and standard deviations of program completion times from data obtained and included in a previous report. (Semi-Annual Progress Report, December 31, 1967, Report No. R-9, Experimentation with Computer-Assisted Instruction in Technical Education).

Table 1

Mean and Standard Deviations of Program Completion Times

Group	n	Mean (in seconds)	S.D. (in seconds)
Audio	20	3319.60	1147.65
Loose-leaf Chart Static Display	23	2674.04	565.64
Typewriter	24	3003.79	1011.68
Slide	20	2885.60	782.52

A Student Records Program (Bahn, 1966) provided the authors with a listing of each subject's performance on each frame in the program. (Because of a programing error, these data were available for only three of the four groups;

data were not available for the group which received the stimulus material from the typewriter. The error was corrected in a later experiment.) The listing provided the authors with the response entered by the student, the number of attempts on each question, and the response latency for each response. The mean number of attempts per student and the mean response latency per student were calculated. From these data, the mean number of attempts and the mean response latency per group was calculated and plotted on a graph. A sample of these graphs is shown in figures 1 and 2.

Examination of these graphs resulted in the following conclusions:

1. In general, the group which received instruction in the audio mode of presentation had a tendency to have a higher mean response latency as well as a higher number of attempts per frame.

2. In general, the chart display mode of presentation produced the best results; that is, tended to have the lowest mean response latencies and the lowest mean number of attempts per frame, (although the differences were non-significant).

3. On certain frames, the pattern was reversed, i.e., the group receiving instruction in the audio mode had the lowest mean response latency while the group which received the instruction in the display mode had the highest mean response latency suggesting a frame-by-chart-mode-of-presentation interaction.

4. Certain frames showed a high mean response latency and a high mean number of attempts for all groups which seems to indicate a difficult question regardless of mode of presentation.

Based on the above information, the authors examined the frames and the student responses in order to determine what was causing the high mean response latencies and the high mean number of attempts. In some cases it was found that the student knew the correct answer and entered a form of the correct answer, but the computer would not accept the response because of the author's failure to anticipate a particular correct answer in programing. An effort was made to correct the programing wherever possible.

In other cases it was found that some frames were too long, especially for the group receiving the material from audio tape. Information presented during the early part of the frame was forgotten by the end of the frame when the subject was required to use the information. An effort was made to shorten the frame, in some cases cutting the length of the frame in half; for example, the frame:

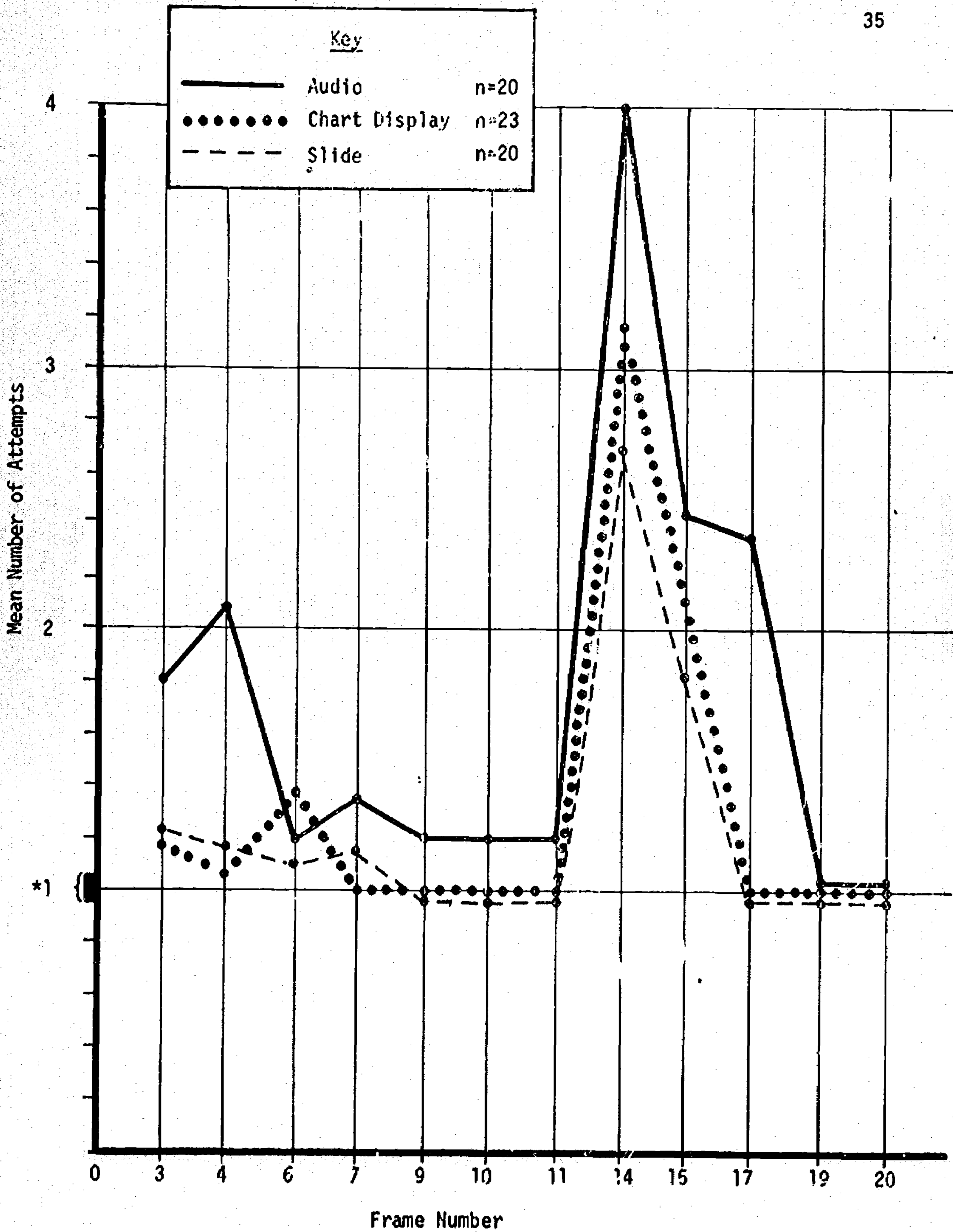


Fig. 1. Mean number of attempts per group on each frame.

* All points within range of rectangle have a value of one.

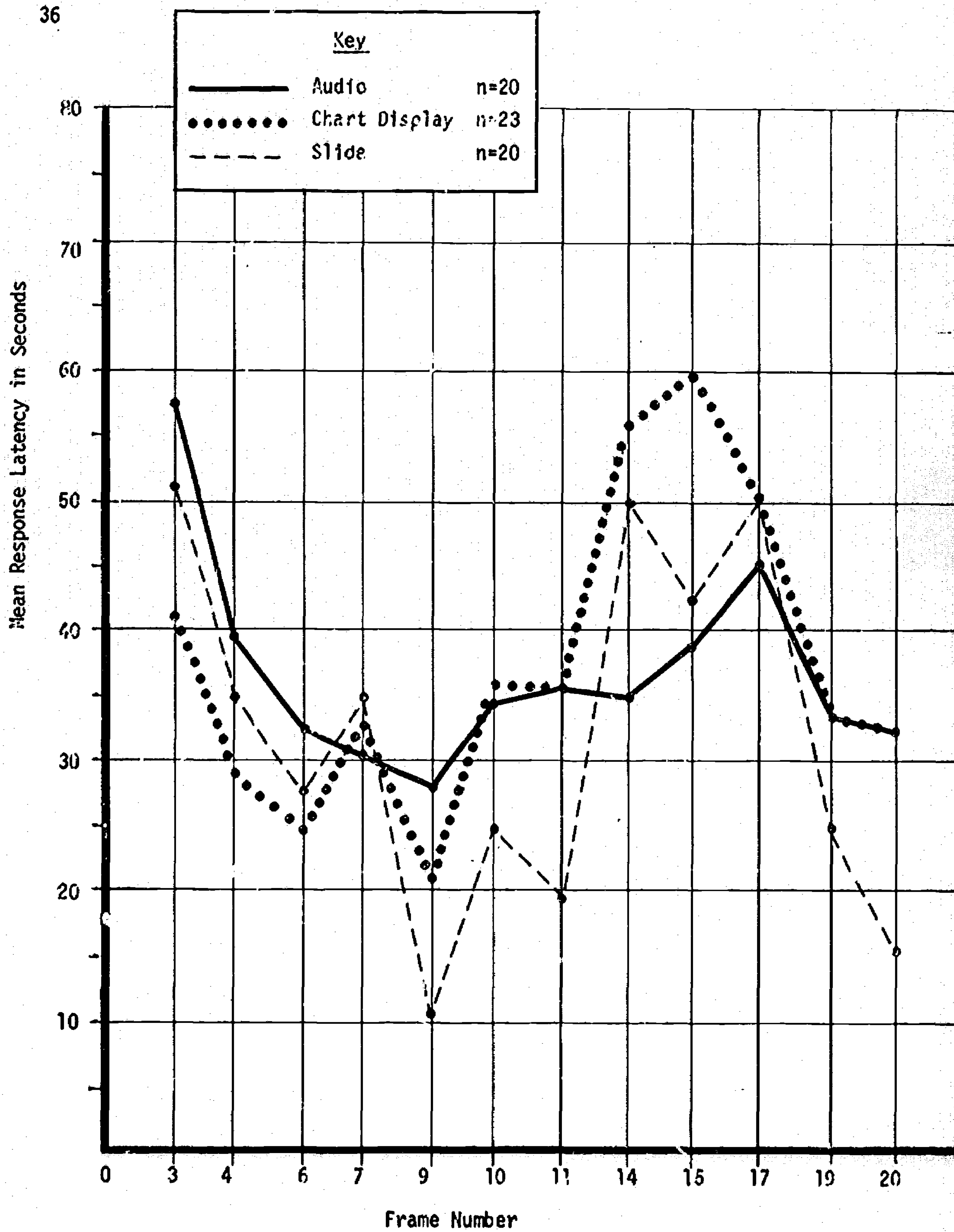


Fig. 2. Mean response latency per group on each frame.

By using this method of working with units, you should always be able to calculate units which measure the quantity you are trying to find. If you are trying to find a distance and obtain a unit of seconds, your answer is wrong. If you calculate the proper units, this is a clue that you are working the problem correctly.

You will have better success in learning how to use units if you actually write each problem on a paper and perform the computation. Be sure to cancel units where it is possible.

was split into two frames, each one containing a single paragraph.

It was found that certain questions were too long to remember, especially for a naive subject in the given subject matter. For example, the following frame has a high mean number of attempts.

Each of the following quantities contains a number and a unit. Type the words which are the units in each of the quantities.

5 miles 2 hours 3 cm 6 grams

To compensate for this, the last part of the question (6 grams) was dropped in the course revision.

Further analysis of mean response latencies and mean number of attempts showed that students were not asked to respond to material presented in the course until 4 or 5 frames had passed. In the revised course, an attempt was made to examine the student on each piece of new information immediately after presentation. In some cases additional drill material was added to the course to insure that the student mastered the concept before proceeding to new material based on the concept.

In restructuring the material the concept was introduced, an application presented, and students were then tested on the application. For example, the following frame:

The electric company charges you on the basis of how many kw-hr (kilowatt-hours) of electricity you use.
When units of kilowatts are multiplied by units of hours, the unit obtained is kw-hr.
How much electrical energy do you pay for if you used a 75 kw electrical motor for 4 hours?

was changed to:

> When units of kilowatts are multiplied by units of hours, the unit obtained is kw-hr.
> The electric company charges you on the basis of how many kw-hr (kilowatt-hours) of electricity you use.
> How much electrical energy do you pay for if you used a 75 kw electrical motor for 4 hours?

Upon completion of the course revision, the course was tested on a small number of students in order to determine the value of the revisions.

Method

Materials

Upon completion of the revisions, as described, the course, orginally titled "Working With Units," contained 47 frames including 6 slides common to all modes of presentation. The sequence was designed so that all Ss received four warm-up frames, each frame presented in a different mode of presentation, the main purpose of which was to acquaint each Ss with the correct method of terminal operation. Following the introductory material, 43 frames of material were presented for which data were collected and analyzed.

In order to provide for a variation in stimulus modes, four versions of the course were created. The material from the 37 frames (not including slide material) was presented four ways. One group received this material on audio tape to provide the audio mode of presentation. One group received this material printed on charts and put together in a booklet, each page of which contained the material from one frame of the program. The program instructed the Ss to read a given page. This mode of presentation provided the chart display mode of presentation. In a third version of the course, the type mode, the material was typed to the student on the typewriter associated with the IBM 1050 computer terminal. The fourth group received the instructional material on 2 x 2-inch photographic slides; the material was identical to that contained on the audio tape, chart display group, and typewriter output. All groups received identical versions of the course and all groups were required to answer the questions by typing their answers on the typewriter keyboard at the terminal.

A 20-item constructed-response test was created. The test was designed to measure factual material as presented in the program as well as a subject's ability to transfer what he learned to similar problems. For example, in addition to dividing meters by seconds, a concept taught in the course, the Ss also had to divide fictitious units such as dividing "yens by fuds." Since the program was not designed to teach computational skill, it was decided to score the test only on the basis of whether or not the S had the correct units, not whether or not the S had the right numerical answer. The Kuder-Richardson Formula 20 reliability of this test was .862.

Subjects

The Ss consisted of 33 volunteer upperclassmen majoring in education and taking Instructional Media 435 at The Pennsylvania State University during the Fall Term, 1967. The Ss were randomly assigned to one of the four experimental treatments. The Ss did not have a mathematics or physics background. In addition another group consisting of 90 upperclassmen majoring in education and taking Instructional Media 435 at The Pennsylvania State University during the Fall Term, 1967, served as a naive control group and took only the posttest.

Procedures

Each S signed on the course and was presented with the instructional material and questions based on the material. The S responded to the questions by typing his answer on the typewriter keyboard. Feedback material was presented by the computer to all Ss via the typewriter. Upon completion of the course, Ss were administered a 20 question constructed-response test off-line. Total for each question, the response latency for each response, and the number of correct responses on the posttest were collected for each subject.

Findings

Table 2 shows the means and standard deviations of the completion times (in seconds). Table 3 is an analysis of variance summary table for the completion times. The F-ratio was not significant. Table 4 shows the means and standard deviations of posttest scores for all groups. An analysis of variance procedure (Table 5) produced an F-ratio significant beyond the .001 level of confidence. Sheffe's procedure showed that significant differences

Table 2

Means and Standard Deviations Of Completion Times

Group	n	Mean (in seconds)	S.D. (in seconds)
Audio	11	3056.00	692.75
Chart Display	10	2988.70	166.11
Typewriter	6	3078.17	740.76
Slide	6	3062.50	995.44

Table 3

Analysis of Variance For Completion Times (In Seconds)

Source	d.f.	Sum of Squares	Mean Square	F
Treatment	3	40,888.48	13,629.49	0.03
Error	29	12,745,466.43	439,498.84	
TOTAL	86	12,786,354.94		

Table 4

Means and Standard Deviations of Posttest
Raw Scores (Maximum Possible Score: 20 Points)

Mode	n	Mean	S.D.
Audio	11	15.45	1.44
Chart Display	10	14.70	3.71
Typewriter	6	13.00	3.58
Slide	6	16.17	3.06
Control	90	9.69	4.56

Table 5

Analysis of Variance For Posttest Scores

Source	d.f.	Sums of Squares	Mean Square	F-Ratio
Treatment	4	693.07	173.27	9.71 (P<.001)
Error	118	2104.95	17.84	
TOTAL	122	2798.02		

existed between the audio and control groups, the display and control groups, and the slide and control groups. The difference in means between the type and control groups was not significant.

Figures 3 and 4 show a segment of the graphs obtained from the revised course. Because of revisions, frame numbers in figures 3 and 4 are not comparable with figures 1 and 2. A general inspection of the graphs shows that the majority of the extreme peaks have been tempered, indicating that the frames in the course are approaching equal difficulty. It is also apparent that the type group now has the shortest mean response latencies. This is most probably due to the fact that the Ss could read the material as it was being typed to them and be ready to answer as soon as the computer gave them control of the keyboard. The other groups were forced to read the material during the time allotted for responding. Thus, the major differences in mean response latency are interpreted as a result of the programing technique, and not of the mode of presentation. This interpretation is also reinforced by the lack of significance in the obtained analysis of variance on the total time to completion data.

The graph for mean number of attempts per group on each frame also indicates that many of the high peaks were tempered; however, there are 1 or 2 new peaks to be examined, as well as some smaller peaks.

Implications

The main purpose of this experiment was to test a procedure for course revision based on past performance. Graphs were prepared to indicate weak sections or frames in the course based on high mean response latencies and/or high mean numbers of attempts. The graphs also indicated frame-by-mode-of-presentation interactions, i.e., indicated frames where one mode of presentation was superior or inferior to the other modes of presentation. It is felt that by examining these interactions, it may be possible to indicate which mode of presentation should be used for which purposes as well as how to best use a medium for a specific purpose. For example, it was found that, for the given course material, it was necessary to keep the audio messages relatively short.

In revising the course, care was taken to improve those frames which contributed large proportions of variance to large mean response latencies and

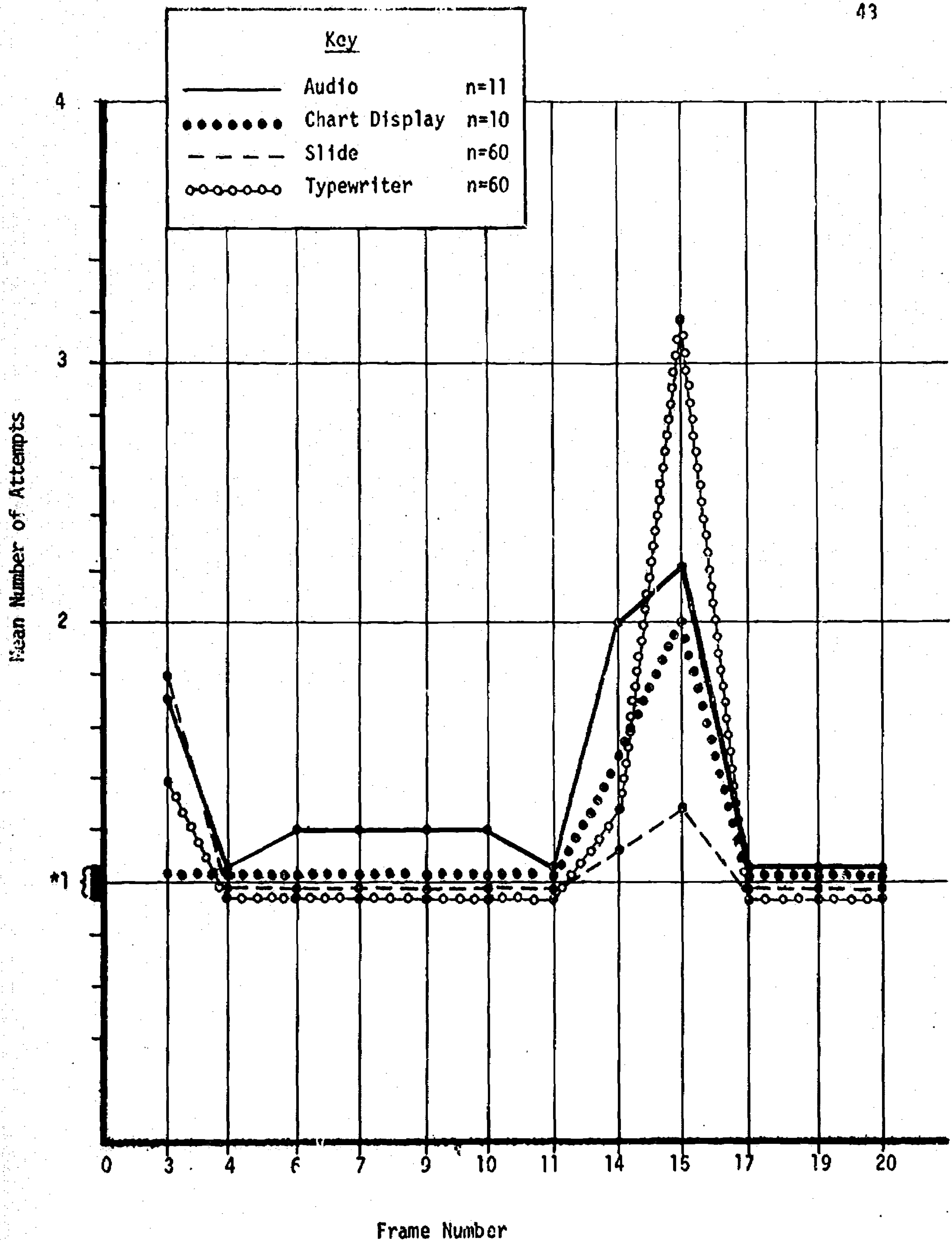


Fig. 3. Mean number of attempts per group on each frame.

* All points within range of rectangle have a value of one.

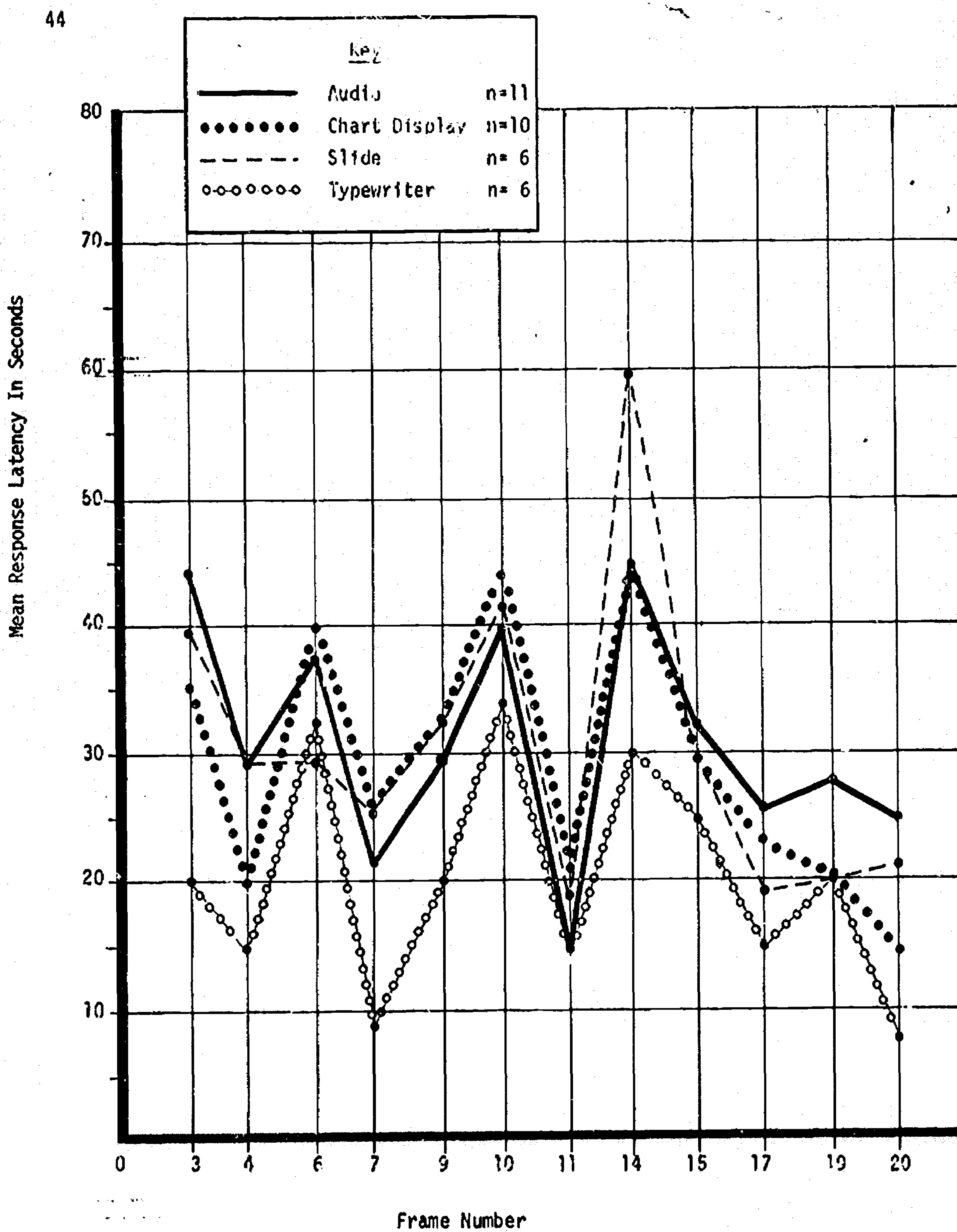


Fig. 4. Mean response latency per group on each frame

the high mean numbers of attempts. The revisions would tend to make the course more uniform and easier for all groups, since all versions of the course were improved, a condition reflected by the minute mean differences for total time to complete the program. A possibility for future research would be to revise only the frame for the modality which experienced difficulty, leaving the other frames unchanged. Diagnostic revisions may serve to lower within group variance and capitalize on the differences that are inherent in the various modes of stimulus presentation available with CAI.

The course material used in this experiment is the product of a number of revisions after it had been carefully written by a subject matter and programing expert. Even now, it is far from being optimally efficient and effective stimulus material to promote student learning. There are still many peaks and valleys that must be accounted and compensated for through revisions or branches. The most significant finding to date is that course development is a complex, time-consuming process which must be carried out in a context where student performance data are continually used as a basis for subsequent revisions.

References

Bahn, T. A. Student performance summaries for CAI courses. Experimentation with Computer-Assisted Instruction in Technical Education. Semi-Annual Progress Report, (by H. E. Mitzel and others), University Park: Computer Assisted Instruction Laboratory, The Pennsylvania State University, December 1966.

Johnson, D. W. and Borman, K. G. Relative effectiveness of various modes of stimulus presentation through computer-assisted instruction. Experimentation with Computer-Assisted Instruction in Technical Education. Semi-Annual Progress Report, (by H. E. Mitzel and others), University Park: Computer Assisted Instruction Laboratory, The Pennsylvania State University, June 1967.

Sparks, J. N. Expository notes on the problem of making multiple comparisons in a completely randomized design. Journal of Experimental Education, 1963, 31, 343-349.

1500 SYSTEM GEOMETRIC DICTIONARY

Paul V. Peloquin

Introduction

This is a general description of the geometric dictionary used by the Computer Assisted Instruction Laboratory at The Pennsylvania State University. It is hoped that this introduction will enable readers to duplicate and use this dictionary on any cathode ray tube (CRT) terminal of the IBM 1500 system. Only an elementary knowledge of certain Coursewriter II instructions will be assumed throughout this paper. (The reader is referred to the following sections of the Coursewriter II Manual [IBM, 1967]: Part I, pages 2-12; Part II, pages 7-15, 58-66.)

There are at least four advantages in using this geometric dictionary. First, the geometric dictionary uses less core storage than a graphic set, yet there is no limit to the number of line drawings which can be made. Because the components of a dictionary are small, they become more general and may be used in many different combinations. By analogy, a graphic set may be equated to a vocabulary of 64 words, while the geometric dictionary may be equated to an alphabet of 128 letters. The geometric dictionary may be equated and used within various courses thereby conserving core storage. Second, the necessity of keypunching each and every one of the line drawings dot by dot is eliminated. Third, the geometric dictionary allows the author to construct, on-line, the graphics for his course. This ability allows him to instantly see the line drawing as it is entered and make necessary changes or corrections. Fourth, words and line drawings can be combined without the one column gap necessary when using a graphic set.

There are, of course, some disadvantages as well. The geometric dictionary has been designed for producing moderate and large-sized line drawings. Except for some gross shading, such as "blackened" areas and hachures, the figures produced with this dictionary have been only line drawings. Some constraints are placed upon the drawing by the availability and the nature of the line segments. This restriction necessitates careful planning, but with the help of the IBM 1500 Instructional Display Planning Guide sheets it does not constitute a serious problem. Usually the addition of an extra character into the dictionary or a bit of programing ingenuity will overcome problems in producing still or dynamic line drawings.

General Description

The geometric dictionary consists of line segments entered as dictionary characters which may be manipulated as such. Throughout this description the characters appearing on the terminal keyboard and on the CRT under the system dictionary, whether letters, symbols, or numbers, will be referred to as "associated keyboard characters." The characters containing the line segments of the geometric dictionary will be referred to as geometric characters, and the line parts contained within each of the geometric characters will be referred to as line segments, whether straight line segments, arcs, or special characters. Since we are dealing with dictionary characters, once the geometric dictionary has been called by a dictionary change the geometric dictionary has all the operating characteristics and functions of the system dictionary. A particular line segment is called and displayed on the screen by entering the associated keyboard character in a display text (DT) or display text insert (DTI) instruction. For example:

```
DT  12,10///*laaaaa*b*b78 a a a a a 78*b*baaaaa*e
```

After a dictionary change (denoted by *l) is made, this instruction will display the "a" as a horizontal straight line and the numbers "78" as the left and right half of a small circle respectively. The backspace function (*b) has been used to superimpose lines.

Procedure

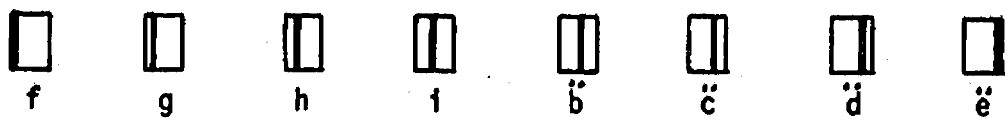
Sketching

The procedure involved in the construction of a line drawing requires three steps: sketching, coding, and entering. First, a sketch of the line drawing is made on an Instructional Display Planning Guide. The sketches should be rather simple and should be constructed of those line segments that the programer knows are available in the dictionary. On the three pages following this description of the first step are reference pages indicating the orientations of the line segments that are presently available to the programer. It is recommended that these reference pages be reproduced in a transparent form in order that the programer may overlay the available line segments on his sketch for comparison. The reader is reminded that each geometric character may be used independently. Thus if the third (and middle) character of a 30° line is needed, it may be used independently of the other four geometric characters which make up the completed line. If the programer cannot match his sketch with an available line, he should select the closest approximation and revise his sketch accordingly. A limited number of special characters may be inserted into the geometric dictionary if the programer finds that their omission seriously handicaps him. Additional flexibility can be gained by the use of the keyboard functions such as space, backspace, index, reverse index, and dictionary change. Superimposition, offsetting by a half-line, the display of text and geometric figures in juxtaposition or superimposition, the display of only half of a geometric character, and shading are some facilities gained by the use of standard keyboard functions. With a little practice, the programer's familiarity with the lines available in the dictionary should grow to a point where he will be able to produce sketches which require no lines that are not already in the dictionary. In anticipation of the second step, the programer may wish to make mental or written notes on the line segments he intends to use in constructing the line drawing.

REFERENCE SHEET FOR STRAIGHT LINE SEGMENTS

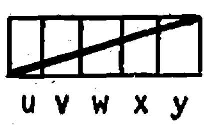
Horizontals

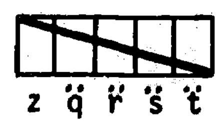
Verticals



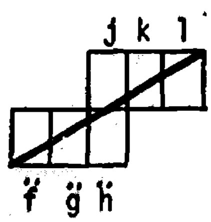
Positive | Negative

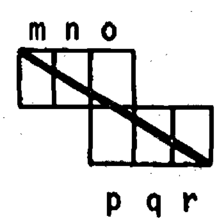
15° Lines



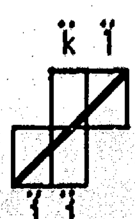


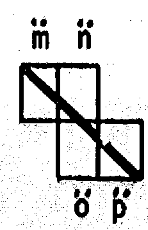
30° Lines





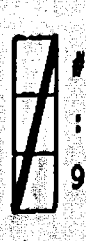
45° Lines





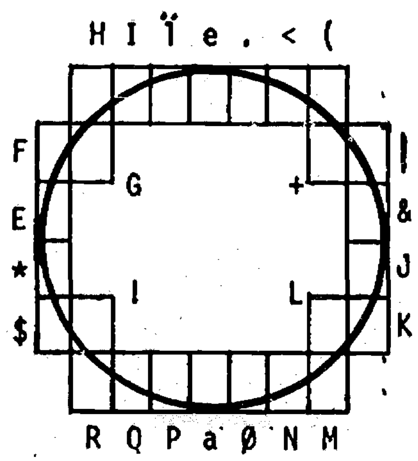
60° Lines

75° Lines

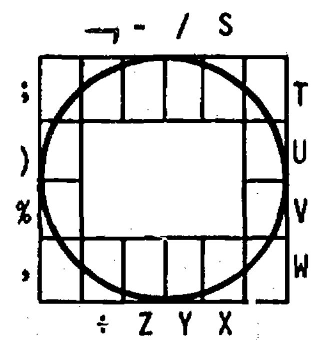


REFERENCE SHEET FOR CURVED LINE SEGMENTS

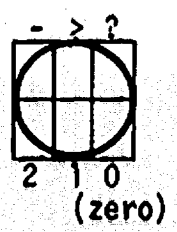
6 Characters Tall



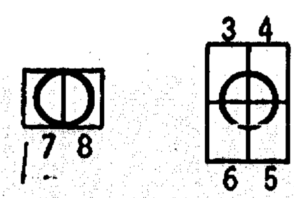
4 Characters Tall



2 Characters



1 Character



1/2 Character

REFERENCE SHEET FOR SPECIAL AND BLANK CHARACTERS

Half Horizontals

ü v̈ 0̈ (zero) X (times)

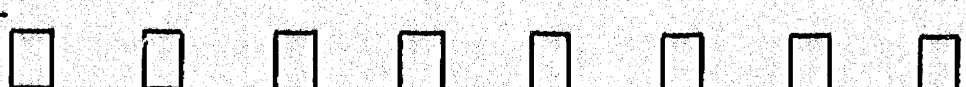
Large Dot

"

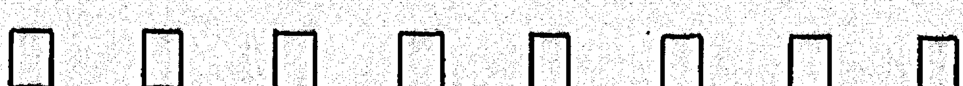
Cursor

;̈

Blanks



3̈ 4̈ 5̈ 6̈ 7̈ 8̈ 9̈ Ẅ



Ẍ ÿ z̈ ;̈ /̈ A B C

Coding

The second step in constructing a figure with the geometric dictionary may take place once the sketch in the Instructional Display Planning Guide contains no lines which are not in the geometric dictionary.

Short Form. The three reference pages preceding this section are used in the "short form" of encoding. The reader should note that the associated keyboard characters are written beside, above, or below the geometric characters to which they refer. One simply chooses the line segment and geometric character he wishes, then encodes the associated keyboard character.

Long Form. The use of the long form is not described here since simple line drawings can easily be constructed with the use of the short form. More complex figures requiring knowledge of every lighted dot within the character would necessitate the use of the long form.

The purpose of this abbreviated report is to introduce readers to the 1500 system geometric dictionary being developed at the Penn State Computer Assisted Instruction Laboratory. Further information enabling reproduction and use of the dictionary on other IBM 1500 instructional systems will be made available upon final completion and refinement.

A PROCESSOR FOR MULTIPLE NUMERIC ENTRIES

Terry A. Bahn and Bobby R. Brown

For the sequential testing program (see Prior Knowledge and Individualized Instruction in this report) it was desired that a student be allowed to respond to multiple choice questions in the embedded tests by stating his subjective probability or degree of belief for each of the choices presented. It was further desired that the mode of responding not be unduly time consuming or unnecessarily restricted in the range of format variations accepted. Due to the time factor involved following each student response on the 1410 system, it was not desirable to have the student enter his subjective probability for each choice separately. If the student was to be allowed to enter his subjective probabilities for all choices in a single response, there were two possible ways of processing the response in which the information concerning the subjective probability for each choice would be preserved. A series or "stack" of possible answers could have been provided in the program against which to compare the student's response. However, because of the number of combinations of subjective probabilities possible and the permutations possible for each combination, it was not practical to employ this procedure.

The alternative procedure entailed evaluation of each subjective probability in the student's response when the individual probabilities were themselves components in a string of probabilities which made up a single response. No such capability exists in 1410 Coursewriter. The subroutine described here provided this capability and was employed in the previously referenced instructional program.

The algorithm for the multiple entry subroutine is as follows:

1. The student's response is entered in the form:

   xx, xx
   xx, xx, xx, xx
   or
   xx, xx, xx, xx, xx

where xx is any two-digit number within the range 00 to 99 or the three-digit number 100.

2. An edit function deletes extraneous spaces, letters, and special characters.

3. A series of edit functions rounds the numbers in the student's response to the nearest ten and converts the number 100 to 99.

4. The response is now in the form:

yy, yy
yy, yy, yy, yy
or
yy, yy, yy, yy, yy

where yy is a member of the set of numbers (00, 10, 20, 30, 40, 50, 60, 70, 80, 90, 99).

5. Response processing now enters a series of twenty-two (22) to fifty-five (55) initial character function (ic fn) calls depending upon the number of entries (2, 4, or 5) and the values of the entries. The initial character function allows one to compare n initial characters of a response and to include "don't care" characters (in this case $) which will match any single character in the string of characters which is to be matched.

6. The first set of initial character functions compares the first three (3) characters of the student's response with answers of the form:

yy

7. The second set compares the first seven (7) characters of the response with answers of the form:

$$, yy

8. If only two (2) entries were required, processing passes to step 13.

9. The third set of functions compares the first eleven (11) characters with answers of the form:

$$, $$, yy

10. The fourth set compares the first fifteen (15) characters with answers of the form:

$$, $$, $$, yy

11. If only four (4) entries were required, processing passes to step 13.

12. The fifth set of functions compares the first ninteen (19) characters with answers of the form:

$$, $$, $$, $$, yy

13. Each time an entry was matched, its value (yy) was placed in a counter corresponding to its original position in the total response. These counters are now added together to see if their total is one hundred (90 to 110 to allow for rounding error).

14. The entry with the highest value is loaded into counter six (c6) and a switch is set to indicate the original position of this value.

15. Control is returned to the main program.

With the advent of the IBM 1500 instructional system and the extract integer function (er fn) this same procedure can be implemented with fewer statements and greater accuracy. A macro has been written to accomplish this task. The coding for this macro (spb 666) is contained in Appendix A.

Appendix A

```
 1  aa  *≯sp*e
 2  1d  0≯c1*e
 3  1d  0≯c2*e
 4  1d  0≯c3*e
 5  1d  0≯c4*e
 6  1d  0≯c5*e
 7  1d  0≯s1*e
 8  1d  0≯s2*e
 9  1d  0≯s3*e
10  1d  0≯s4*e
11  1d  0≯s5*e
12  fn  e1≯b0≯1≯c1*e
13  fn  e1≯b0≯2≯c2*e
14  fn  e1≯b0≯3≯c3*e
15  fn  e1≯b0≯4≯c4*e
16  fn  e1≯b0≯5≯c5*e
17  1d  c1≯c6*e
18  ad  c2≯c6*e
19  ad  c3≯c6*e
20  ad  c4≯c6*e
21  ad  c5≯c6*e
22  br  #01##3≯c6≯1≯100*e
23  br  #02##3≯c6≯g≯100*e
24  1d  c1≯c6*e
25  br  #03##3≯c2≯1e≯c6*e
26  1d  c2≯c6*e
#03##3*e
 1  br  #04##4≯c3≯1e≯c6*e
 2  1d  c3≯c6*e
#04##3*e
 1  br  #05##3≯c4≯1e≯c6*e
 2  1d  c4≯c6*e
#05##3*e
 1  br  #06##3≯c5≯1e≯c6*e
 2  1d  c5≯c6*e
#06##3*e
 1  br  #07##3≯ck≯ne≯c6*e
 2  1d  1≯s1*e
#07##3*e
 1  br  #08##3≯c2≯ne≯c6*e
 2  1d  1≯s2*e
#08##3*e
 1  br  #09##3≯c3≯ne≯c6*e
 2  1f  1≯s3*e
#09##3*e
 1  br  #10##3≯c4≯ne≯c6*e
 2  1d  1≯s4*e
```

```
#10##3*e
  1  br  #11##3/c5/ne/c6*e
  2  ld  1/s5*e
  3  br  #11##3*e
#01##3*e
  1  dt  28,5/4,28/40,0/The sum of your answers should not*c*ibe LESS than 100.
 Try again.*e
#02##3*e
  1  dt  28,5/4,28/40,0/The sum of your answers should not*c*ibe MORE than 100.
 Try again.*e
  2  br  re*e
#11##3*e
  1  en  *e
```