DOCUMENT RESUME

ED 029 082 VT 005 978

By-Mitzel, Harold E.; And Others
Experimentation With Computer-Assisted Instruction In Technical Education. Semi-Annual Progress Report. Report No. R-9.
Pennsylvania State Univ., University Park. Computer Assisted Instruction Lab.
Report No-Project 5-85-074
Pub Date 31 Dec 67
Note-82p.
EDRS Price MF-$0.50 HC-$4.20
Descriptors-*Aptitude Tests, *Computer Assisted Instruction, *Feedback, Programed Instruction, Research, *Research Projects, Research Reviews (Publications), *Spelling Instruction, Testing, Verbal Communication

A paper and four research studies involving computer assisted instruction (CAI) are reported. "On Narrowing the Credibility Gap for Computer-Assisted Instruction" by Harold E. Mitzel reviews the cost, shortage of programs, and lack of detailed educational plans as major reasons for the nonacceptance of computer-assisted instruction. "A Comparison of the Effectivess of Five Feedback Modes in a Computer Assisted Adjunct Auto Instruction Program" reports a study involving five feedback modes and 75 college students. "Numerical and Verbal Aptitude Tests Administered at the Student Station" reports an experimental attempt to teach remedial spelling to college students. "Effects of Reducing Verbal Content in Computer Assisted Instruction Programs" deals with correlation between intelligence and learning after studying a low verbal content program, a decrease in instructional time, and no learning reduction due to content reduction. VT 005 977 is a similar report. (EM)

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

DECEMBER 31, 1967

Report No. R-9

The Pennsylvania State University

Computer Assisted Instruction Lab~~oratory~~
University Park, Pennsylvania

Semi-Annual Progress Report

EXPERIMENTATION WITH COMPUTER-ASSISTED INSTRUCTION IN TECHNICAL EDUCATION.

Project No. 5-85-074

Principal Investigators
Harold E. Mitzel And Others George L. Brandon

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

Helen L. K. Farr Joseph L. French
David A. Gilman

Contributing Authors and Researchers
Nancy Harvilchuck Harriett A. Hogan
John Tardibuono

December 31, 1967
Report No. R-9

FOREWORD

Periodically, the staff of the CAI Laboratory at Penn State reports to its monitors in the U. S. Office of Education on the ways in which University and Federal resources have been employed to shed new light on the subject of computer-assisted instruction in technical education. The experiments, descriptive studies, and talks on the following pages represent activities completed during the period July 1, 1967 to December 31, 1967.

The research and development progress in CAI completed up to December 31, 1967, has been accomplished on a hardware configuration organized around an IBM 1410 computer, as a central processor, located in Penn State's Computation Center. Connected to the central processor by telephone lines, the CAI Laboratory has maintained student stations consisting of eight IBM 1050 typewriters, each augmented by a random access tape recorder and a random access slide projector. This modified business application equipment has provided yeoman service for two and one-half years, but was replaced late in December, 1967, with an eight-terminal IBM 1500 instructional system. The new system designed for instructional purposes offers numerous advantages over the former. Displays of materials are achieved almost instantaneously on the cathode-ray tube terminal instead of the tedious typewriter type-out. Student's answer processing is vastly improved. Last, but not least, the cost per student terminal hour is about one-fourth that of the former 1410/1050 system.

During the first half of 1968, the staff on the project will be primarily engaged in the translation of experimental course materials from Coursewriter I, the author language of the IBM 1410 system, to Coursewriter II, the author language of the IBM 1500 system. In addition, we will be teaching ourselves the intricacies of the new hardware/software configuration and developing new dictionaries and macros. We hope to initiate studies during the period which build upon what we have already learned about response modes, sequencing, and student attitudes.

Computer-assisted instruction, of all the new educational technologies, is in flux and cannot be stabilized with respect to hardware, author languages and teaching strategies for a few more years. The strident demands of some school administrators and military training officials for a premature evaluation of CAI as a generic element in schooling offers a clear and present danger to the orderly development of CAI. Universities, like Penn State, engaged in carefully planned research activities need time and funds to investigate and consolidate what they are learning about CAI before being forced to evaluate its ultimate application in education.

Harold E. Mitzel
University Park, Pennsylvania
February 15, 1968

CONTENTS

ON NARROWING THE CREDIBILITY GAP FOR COMPUTER-ASSISTED INSTRUCTION[1]

Harold E. Mitzel

I'm happy to be here today, a consequence of your program committee's collective short memory, since I had the pleasure of being on this platform three years ago. It is easy for me to be nostalgic about ERANYS because I remember the days of the middle '50's when the active membership gathered for lunch around two small tables in a Syracuse hotel. That was quite a contrast to this large gathering today

We have been hearing a great deal about "credibility" on the political scene these last few months, so I have taken it as my theme for today's discussion. First, let us begin with the candid premise that computer-assisted instruction (or CAI) lacks credibility as a feasible operational tool in the minds of an overwhelming majority of educators. I know that this unbelievability for CAI is characteristic of my colleagues in the elementary and secondary schools and, to a large extent, in colleges and universities. I assume that it is also the predominant attitude among your colleagues in a variety of responsible positions. Although the educational applications for this new technology cover the gamut of teaching and learning situations, I hope that you will bear with me if I talk about some possible ways of narrowing the credibility gap for CAI in the lower schools.

There are currently a host of ways in which modern digital computers can be used to facilitate education and training. All of these uses can be called computer-assisted instruction in a generic sense, but I would feel on firmer ground if we could review briefly some distinctions within CAI. I think that these distinctions are best understood in relation to the function of the computer in executing a variety of educational tasks.

First, a computer's remote typewriter terminal may be located in a classroom and used as a <u>laboratory computing device</u>, or scientific calculation aide.

---

[1]Remarks prepared for New York State Research Convocation, Albany, New York, presented November 13, 1967, at the Annual Meeting of the Educational Research Association of New York State.

This use has been exploited in several high schools in the Boston area, with tele-typewriters connected by wire to a computer facility located at Bolt, Beranek and Newman, Inc., a defense contracting agency at Cambridge. Staff members from this project have reported quantum jumps in the motivation of high school pupils who vie with one another to "play with" the computer terminal. At Altoona High School in Pennsylvania, a similar use for remote tele-typewriters in the teaching of computer programing and in mathematical computing has been developed. The prevailing pattern is to locate a single terminal in a classroom and the students of the class take turns on the system. The students become quite expert at devising and executing their own programs; in fact, they often excel their teachers in this capacity. It is this potential for awakening the minds of a vast number of secondary school pupils that has prompted the U. S. Office of Education to purchase feasibility and cost studies on large multi-programed computer systems. The National Science Foundation has recently contracted with Dartmouth for a demonstration of computer usage, in high school classes, by means of a sizeable number of remote terminals.

In colleges and universities, this instructional use of computers will probably come on strong as new third generation computers are provided with operating systems which allow for a non-exclusive conversational mode between typewriter or cathode-ray tube terminal and the computer. At the college level, the University of California at Irvine has pioneered in this auxiliary use of the computer in instruction. Other time-shared systems with this application include the University of Texas, the University of Pittsburgh and Systems Development Corporation at Santa Monica, California.

A second CAI definition involves the use of the computer as a <u>record keeper and retriever</u> of student's biographical and achievement data. This use involves batch processing of data cards by staff members instead of use of the remote on-line terminal by the learner which is essential to a rapid-fire inter-active process. In a sense, a high school or college which does its course and room scheduling, or stores and prints its test results for the guidance department on a computer, is engaged in <u>computer-assisted</u> instruction. This application is becoming more and more common as

the availability of large computers increases, estimated by the President's Science Advisory Committee to be 2500 in number in the country by 1970. Major colleges and universities and large progressive high schools currently employ this application routinely, frequently buying computer time from industry or another educational system.

Related to the record-keeping and information-retrieving role of the computer is the concept of "computer-managed instruction." Under this conception, the computer never interacts directly with the pupils, but provides a student progress or status report for the teacher. Pupils take short segments of conventional programed text material which are graded and the results put into a gigantic computer file containing information about every learner. On command, the computer provides the teacher with diagnostic progress reports, achievement summaries, groupings of pupils, and suggestions for drill and practice exercises. Flanagan (1967) has mounted a major program in a dozen widely scattered school systems to develop this computer application.

Third, CAI can be defined in terms of simulation problems with the computer responding adaptively to learner input. Swets and Feurzeig (1965), in an article in Science, have described an interesting medical diagnosis sequence involving an "on-line" interaction between the computer and the medical school trainee. At the University of Illinois Medical School in Chicago, a patient management exercise has been developed which provides a learner with an intriguing set of simulated word situations. Some "inquiry" oriented materials have been prepared by Suchman and others for the University of Illinois PLATO system. At Penn State, Igo is preparing a simulated physics laboratory sequence for students in technical education programs. And, at IBM's Yorktown Heights facility, simulation models illustrating physical and chemical laws have been stored in a computer. The student manipulates the several variables of a model at the computer terminal in a few minutes to see the effects of one variable on another instead of manipulating complex laboratory equipment over a long period of time. These and other efforts underway in a variety of locations suggest a bright future for this type of computer application which I will refer to collectively as "inquiry methodology."

The fourth definition of CAI involves the computer in the role of tutor. Although a little anthropomorphic in concept, this use of the computer in instruction can be usefully characterized as TUTORIAL CAI. I am indebted to Dr. Thomas F. Hartman of IBM's Watson Research Center for the following three-pronged breakdown of the tutorial CAI approach. First, the most common sub-definition is programed interaction. Short bursts of instructional material are interleaved with computer interrogation of the learner, followed by appropriate author-anticipated feedback. With this pattern in mind, the course author expects the CAI system to carry the major curriculum load and only small enrichment or depth experiences in the subject are expected to take place away from the computer. The second sub-definition of tutorial CAI can be termed the practice and remediation model. In using this model, initial presentations of information are generally made via inexpensive large scale methods such as a lecture, a television presentation, or the reading of a textbook. Following an initial exposure to the material, the student practices partially learned skills and receives feedback at the computer station. Such an application has a number of advantages. It requires minimal amounts of computer and terminal time for each student. The disadvantages include considerable monitoring by the teacher and a lack of a general applicability to many different kinds of subject matter. At Penn State, we have developed for our CAI system a standard set of operational commands called a "macro" which can be used to quiz students who have been presented with small conventional segments of instruction, either text material or by closed circuit video tape lecture. The macro makes it possible to provide a variety of feedbacks to students and an error bucket for storing and representing missed questions. The student has no prior knowledge of the questions that are going to be asked him. The third aspect of tutorial CAI is the recitation and remediation model which differs only slightly, but importantly, from the practice and remediation set. With recitation and remediation, the student receives the conventional lecture or information input, but in addition he prepares as homework the answers to illustrative problems or questions. The student inputs his previously determined responses at the computer terminal and receives confirmation on correct solutions and tutoring

on those incorrectly solved. Thus the CAI tutorial mode partakes of three sub-definitions: programed interaction, practice and remediation, and recitation and remediation.

I have tried to identify four distinct definitions for computer-assisted instruction in which the computer terminal is used as a laboratory calculating device, as a record keeper and retriever of student information, as a problem-solving or model simulation device, and as a tutor. For the balance of my remarks, I'll be employing either the tutorial or the simulation definition, both of which involve a sophisticated dialogue between the student and the teacher's program stored in the computer.

Let us turn now to some of the sources of the CAI credibility gap and an examination of how this unbelievability might be lessened.

First, it is often alleged that CAI is too expensive for implementation in public elementary and secondary schools. This allegation is usually accompanied by figures which show that America spends about fifty-seven cents per hour per pupil (shakily undergirded by dozens of tenuous assumptions) in order to educate its children, and that the most realistic figures for present CAI applications of an operational nature are in the neighborhood of two to four dollars per student terminal hour. I suppose that no one here believes that automobiles ought to operate on the same hourly budgets that horse-drawn buggies do, but it might be instructive to see what are some of the implicit assumptions which usually underlie this negative reaction to CAI costs versus current average costs for traditional instruction. The fifty-seven cents per hour per student figure (really 33 cents, according to Kopstein and Seidel, 1967) for American education is incidentally about one-half the cost of a good reliable baby sitter, but more importantly, it represents an average of educational costs from kindergarten through twelfth grade in all curricula, in rich and poor schools. Every cost-conscious school executive knows that there is a wide variation in instruction costs just within a single school district. We know that it costs more on the average to teach chemistry than it does to teach history, but just how much more is difficult to say since there are almost no published data on course-specific costs. In South Orange, New Jersey, the assistant superintendent

tells me that it costs five dollars per hour per student to teach the home-bound. In almost any modern school, it costs several dollars per hour per student to instruct the mentally and physically handicapped because of extraordinary equipment and staff requirements. Mr. Freeman, Assistant Superintendent of Memphis Schools, writes me that their high school office machines course is more than nine times as costly per hour as physical education for the same pupils, although I'm sure that he hasn't taken space into his estimates since gymnasiums cost more than business education practice rooms. I don't have reliable figures on the per pupil-hour costs for an individualized remedial reading program, but they must be very high. So it seems to me that an average per pupil-hour cost of instruction is not particularly relevant to a specific decision regarding the use of CAI because of the extreme variation in these costs among different programs of instruction.

A second implicit assumption, related to the first, is usually made in examining the feasibility of CAI for schools. This is the assumption that a computer system can by some feat of magic be made simultaneously operable in all subjects taught to all ages of children; that on some given morning, early in September, a school can be converted from its present teacher-mediated program to CAI. We might call it the myth of instant CAI. A strong component of this myth is the notion that tutorial CAI can replace or supplant all instructional personnel, and that the current costs of instruction, primarily salaries, may be diverted to equipment rental or purchase.

I believe that every progressive school should first divest itself of myths and oversimplifying statistics in relation to the instructional costs of CAI. The school staff should then do an exhaustive self-study of its program in order to decide on an area of the curriculum that needs improvement. Many schools under this constraint would find that they were woefully weak in foreign language instruction. Others, particularly small rural schools, would sometimes find themselves deficient in advanced mathematics and science. Impelled by the example of need, the decision, as far as costs are concerned, then becomes one of deciding whether or not a potential improvement in quality in one or two crucial subjects is worth an increase in instruction costs in order to achieve that particular goal.

Such a decision ought not commit any innovator to a district-wide, school-wide, grade-wide, or subject-wide implementation of CAI. Thus the implementation of CAI into any school ought to be deliberately done on a limited and carefully evaluated basis. This innovation implementation effort requires a staff with special training. Unlike industry, public schools have not traditionally had the benefit of a resident research and development staff. Some schools under the impact of Title III of the Elementary and Secondary Education Act are moving in this direction by appointing federal funds coordinators. Staffing the schools with such people is, it seems to me, an absolute necessity if we are to cut the innovation lag and make a majority of schools reasonably modern.

One way of lowering some of the high equipment costs connected with CAI is to plan both day-school and after-school instruction involving pupils during the day and adults (perhaps teachers) in after-school and evening use of CAI computers and terminals. Preliminary experience suggests that roughly a one-third increase in program benefit can be achieved by extending hardware utilization into after-school and evening hours.

But, perhaps the greatest advance can be made out of the high cost dilemma by the simple expedient of moving computer equipment from the category of current operating expense to capital outlay. The citizens in many communities have, since the end of World War II, approved the erection of handsome edifices for housing school activities. School bonds to cover buildings generally have been much sought after by bankers and investors. Why not consider a modern teaching computer as being in the same category as a modern teaching building? Perhaps the useful life of a computer is only ten to fifteen years compared to a building's forty to fifty years, but the shorter term could result in lower interest charges. To be sure for this strategy to be effective, we must have a great deal more information than we now possess on hardware reliability, the general effectiveness of CAI, and on the long-term availability of good programs of instruction.

The latter point brings me to the second factor which contributes to CAI's current lack of credibility. This is the dearth of good course content material for computer-based presentation. Apart from Stanford University's

reading and arithmetic material for young children, Penn State's modern mathematics course at the college level, and Irvine's introduction to computer programing, relatively little CAI course material exists with a tested history in an operational education setting. Most educators remember the teaching machine fiasco of the early sixties, where there were at one time more different types of machines than there were good programs to run on them. The implication of this experience is that no educator wants to be in the position of having to defend the presence of a warehouse full of useless equipment because there are no teaching programs or "software" to fit the hardware.

The "lack-of-software" portion of the credibility gap, in my opinion, can be narrowed most effectively by the organization of user groups, the members of the group having common needs and making mutually supportive contributions to the larger goal. User groups are commonplace among universities that employ similar computer hardware configurations for scientific computing purposes, and it is true that computing equipment manufacturers make significant contributions to the operation and success of these groups. I strongly advocate similar kinds of consortia among public schools, state education departments and higher education institutions for the development of CAI course material. Emphasis within such a framework would be placed on the compatibility of the jointly developed course material for the computer and terminal hardware possessed by the consortia members without concern about copyright infringements and proprietary interests. Hopefully, industry would follow its example with user's groups and support these new-style CAI consortia. Failure on the part of education institutions to grasp the initiative in the determination of what is to be taught with the assistance of the computer will allow the choice to go by default to non-educators.

A third factor contributing to the CAI credibility gap is that educators and training specialists can't quite visualize how this new tool can be integrated successfully into their on-going programs of instruction. This disbelief seems to me to be essentially a problem in concept development and staff education.

In-school educational television, for example, is a medium that has not been fully exploited in the public schools because its utilization has not been integrated with on-going programs of instruction. Typically, the classroom teacher turns on the set at the appointed hour and then files her nails while the children watch the tube. As soon as the program is finished, the teacher says, "Now let's see. What were we doing before we were interrupted?" The Midwest Airborne Television Instruction program sponsored in five states by the Ford Foundation provides another example of a failure to determine the importance of utilization patterns in advance of the introduction of an educational innovation. If a classroom teacher wanted to use the Airborne Television presentation for teaching fourth grade arithmetic, she had to turn on the tube at ten o'clock for the fourth grade lesson, ready or not. These kinds of rigidities can sabotage the best technology. Technological innovations like ETV and CAI, which are purported to be capable of carrying a major portion of the curriculum load, cannot have their utilization patterns ignored by development teams if the innovations are to become creditable.

Perhaps the worst way to go about implementing CAI into an operational setting is to get a computer and then try to decide what to do with it. A better approach for a school or a user group is to survey its educational needs and then pick out one or two significant programs where massive improvement is needed. If CAI can provide an optimistic approach to the solution of a particular problem, then it ought to be tried and carefully evaluated. The major problem of integrating CAI into an on-going school program can then, be solved on a limited trial basis, with maximum involvement on the part of teachers, supervisors, and administrators.

I have reviewed three factors that contribute to the CAI credibility gap: 1) the cost of instruction, 2) the shortage of good course content for CAI presentation, and 3) the lack of a detailed plan for utilization of CAI within the context of an on-going instructional program. You can, I'm sure, think of others, but I hope that I may have helped you narrow your own CAI credibility gap a little.

References

Flanagan, J. Phi Delta Kappan, 49. Sept. 1967, pp. 27-31.

Kopstein, F. & Seidel, R. Computer administered instruction versus traditionally administered instruction: economics. Human Resources Research Office, Geo. Washington U., Washington, D. C. April 1967, (mimeo), 43 pp.

Swets, J., & Feurzeig, W. Science, 150. 1965, pp. 572-76.

A COMPARISON OF THE EFFECTIVENESS OF FIVE FEEDBACK MODES IN A COMPUTER-ASSISTED ADJUNCT AUTO-INSTRUCTION PROGRAM

David Alan Gilman[1]

The success of computer-assisted instruction as an instructional device depends to a great degree on how well it can meet the needs of the individual student. Computer-assisted instruction (CAI) differs from programed texts in that the student's responses are evaluated against anticipated answers stored in the memory of the computer. The computer is programed to react in specific ways when the student matches an anticipated response and to provide assistance to the student in the event that the student gives an unanticipated response. The student's responses are evaluated by the computer and he is provided with feedback appropriate to his responses. The mode of feedback and prompting can identify and correct specific student errors and thus may be an important advantage for CAI over other types of instruction.

With computer-assisted instruction, it is possible to tailor the feedback to the student's unique misunderstandings. The importance of this potential has been suggested by Uttal (1961).

> The success of teaching machines will depend largely, I believe on the degree to which they provide feedback to the student and are responsive to the student's needs. The desired relationship between the student and the teaching machine may be termed "conversational interaction" by analogy with the relationship between a student and a human teacher.

Glaser (1966) concludes that there have been few studies dealing with "corrective feedback" in verbal learning. One of the important reasons for the lack of investigation in the area of corrective feedback has been a rigid adherence to the axiom of linear programers (Skinner, 1954; Klaus, 1965; Holland, 1965) which states that the student should complete the program with few or no errors.

---

[1]Currently at Indiana State University, Terre Haute, Indiana.

Another reason for the paucity of studies in error correction is an ethical consideration. Programed instruction researchers are very reluctant to teach learners inaccurate information so that they can study how to correct errors. It has been difficult to design controlled experiments that could investigate the correction of learners' errors without first teaching the subjects erroneous material.

However, the correction of errors is one of the goals of adjunct auto-instruction techniques (Pressey, 1963). In adjunct auto-instruction, the learner is first exposed to a substantial and organized unit of instruction. Following the presentation, there are series of questions designed to enhance the clarity and stability of cognitive structure by correcting misconceptions and deferring the instruction of new material until there has been clarification and elucidation (Pressey, 1962).

This study investigated the use of feedback in a computer-assisted adjunct auto-instruction program. The types of feedback compared were (a) no feedback, (b) knowledge of results, (c) knowledge of correct response, (d) response contingent feedback, and (e) feedback consisting of a combination of knowledge of results, knowledge of correct response, and response contingent feedback.

Rationale

The use of knowledge of results as a mode of feedback has its basis in the principle that reinforcement of correct responses enhances learning. The principle of reinforcement is based on results obtained from many experimental studies of the law of effect. The law, as formulated by Thorndike, (1927), states

> . . . responses accompanied or followed by certain events (called reinforcers) are more likely to occur on subsequent occasions, whereas responses not followed by this class of events subsequently show a lessened probability of occurrence.

Thorndike (1927) systematically investigated the law of effect in a classical study. Seven blindfolded subjects were asked to draw a line four inches long. After 400 trials with no knowledge of results, their errors averaged practically the same as in the beginning. They were given 25 more

trials and after each trial they were permitted to open their eyes and check the line they had drawn. After only four trials, they were able to reduce their average error to 3/16 of an inch.

English (1942) invented a device which helped train soldiers to squeeze a rifle trigger. It provided visual feedback through the use of a manometer, which revealed to the soldier a change in the height of a liquid column. If he squeezed the trigger smoothly or spasmodically, the mercury column would rise correspondingly and provide visual feedback. Better performance was reported for those soldiers who used the device.

The value of knowledge of results in programed instruction has been demonstrated in several studies. Pressey (1926), Jensen (1949), Smith (1964), and Paige (1966) demonstrated that test devices which allow the subject to find the correct answer at the time of taking the test contribute to increased learning. Knowledge of results may be beneficial even if it takes a form as simple as confirmation of a correct response. Angell (1949) and Kaess and Zeaman (1960) found significant advantage for confirmation as compared with no confirmation. Kaess and Zeaman found that a group which had no confirmation on the first trial did as poorly on the second trial as the confirmation group did on its first trial and strikingly worse than the confirmation group had done on its second trial.

However, Klaus (1965) describes a point of view of those programers using the knowledge of results technique which may make that technique ineffective for correcting student errors. In describing this point of view, Klaus concludes that many programers using the knowledge of results technique believe that its sole value is in its reinforcement quality and that reinforcement occurs only when the student's response is correct. Klaus contrasts this point of view with that of other programers who use feedback as a means of providing information to correct the student's misunderstanding. If there were no purpose to feedback other than to provide the student with reinforcement, then statements such as "you are correct" should prove equally effective as a statement of the correct answer.

However, Glaser (1966) cites evidence that providing the correct answer following an incorrect response is a reinforcing event in the same way that

confirmation after a correct response is a reinforcing event. These findings are in agreement with those of Holland (1965) who concluded after analyzing several studies that there were no advantages for prompting a student to give the correct answer. Holland concluded that if a student does not know the correct answer, he might as well be told it, this occurs when the student is provided with knowledge of the correct response feedback.

Klaus (1966) describes response contingent feedback as a process whereby differentially applied reinforcement improves the quality of a response by shaping it to the desired degree of proficiency. Crowder's (1962) definition of programable material requires that the material be adaptable to contingent feedback.

> If we can say to the student (a) your answer is wrong, (b) this is what is wrong with your answer, (c) this is the feature of your answer that is wrong, (d) this is how you go about figuring out the correct answer, and (e) now try again; then we are dealing with programable material.

A study by Bryan and Rigney (1956) demonstrated the benefit of providing students with response contingent feedback, as compared with providing the student with knowledge of results. Two matched groups were provided with adjunct auto-instruction by means of a tab test. Students in one of two matched groups were told only whether they were right or wrong when they lifted the tab. The students in the other group were given a short verbal explanation of why any alternative selected was correct or incorrect. When compared on a multiple-choice test administered one week after the training, the explanation of choice group was significantly superior to the non-explanation group.

There may be some advantage in providing Ss with a combination of feedback modes in order to take advantage of reinforcement and at the same time provide the student with information. However, Swets and his co-workers (1962) [sic] found that "fairly extensive feedback may be detrimental to learning." Extensive feedback may be inefficient since providing the student with lengthy feedback messages will require the expenditure of a greater amount of instructional time.

Intuitively, it seems probable that the correction of errors in a program should be beneficial to the student and that an efficient mode of feedback for correcting errors should be developed. However, Glaser (1965) concluded:

> The use of errors and the possible value of incorrect responses may not have been as widely nor systematically investigated as other response contingencies in studies of learning related to the educational process.

The present study used materials designed to teach commonly misunderstood general science concepts. Errors made by the subjects occurred as a result of misconceptions they had acquired in previous conventional instruction. It was thus possible to correct Ss' errors without teaching them erroneous material and without intentionally tricking them to commit errors.

Method

Subjects

Subjects for the study were 75 students in teacher preparation curricula (science teaching excluded) at The Pennsylvania State University. Ss were students in audio-visual classes and had no previous experience with computer-assisted instruction.

Materials

An adjunct auto-instruction program was prepared to teach commonly misunderstood general science concepts. The frames of the program were multiple-choice items. One response to each item was a correct response, one response to each item was a common misunderstanding of the concept, and the other two responses were plausible distractors.

The program caused all of the items to be presented on the first iteration and all items missed on the first interation to be repeated on the second iteration; all items missed on the second iteration to be repeated on the third iteration, until the subject had answered all of the items correctly. Criterion for the program was a correct response to each of the thirty items.

Equipment

The programs were administered under the control of computer-assisted instruction. The student interfaces were four IBM 1050 electric typewriter terminals equipped with a random-access Kodak Carousel slide projector and a pseudo-random access Uher Universal tape recorder. Instruction was tele-processed from an IBM 1410 computer located approximately one-half mile from the student terminals.

Items were presented under computer control to the student from the slide projector. The typewriter provided a space for each response and a space for the S's estimate of his degree of certainty for each response. Feedback for the four experimental groups was provided through messages typed by the typewriter terminal.

The Pretest

The first iteration of the 30-item adjunct auto-instruction program served as the pre-test and also provided instruction.

The Posttest

A paper and pencil posttest of 30 items similar to those of the adjunct auto-instruction was constructed. The items were similar to those of the adjunct auto-instruction program, but the items in the posttest were stated in different contexts than those of the program.

Scholastic Aptitude Measure

Four part scores from The Pennsylvania State University Scholastic Aptitude Examination were obtained from Ss' entrance records.

Procedure

The 75 Ss were assigned to 15 strata on the basis of Scholastic Aptitude Examination scores. The five Ss in each strata were randomly assigned to one of the five treatment groups. The five treatment groups were (Group A), no feedback; (Group B), knowledge of results feedback; (Group C),

knowledge of correct response feedback; (Group D), response contingent feedback; and (Group E), a combination of the feedback modes of Groups B, C, and D. Appendix A shows an example of interaction between the student and the terminal for the five treatment groups.

The posttest was completed by the Ss immediately following the instruction.

Analysis of Data

Scores and data from the adjunct auto-instruction were recorded and stored in the computer memory. The following criteria were measured:

a. Time required to reach criterion
b. Number of correct responses
c. Accumulated response latency

In addition, the following criteria were also employed:

a. Number of iterations of the program to attain criterion
b. Posttest score
c. Scholastic aptitude test score

Analysis of variance were computed for the feedback treatment group x scholastic aptitude levels design. If significant differences due to treatment effects were obtained from an analysis of variance procedure, differences between pairs of means were analyzed by Tukey's W-Procedure (Tukey, 1953). When significant differences found during the analysis of variance procedure were attributable to level effects, the means of the 15 strata were examined to determine whether the significance was indicative of any observable trend in the data.

Tables summarizing analysis of data for each of the variables are included in the appendices of this report. In each of the tables, Part A contains the means of each group and the grand mean. Part B of each table includes data from the analysis of variance. In cases where treatment effects are significant, Part C of the table contains the differences between all pairs of means and a summary of Tukey's W-Procedure for analysis of differences between pairs of means.

Results

The Pennsylvania State University Scholastic Aptitude Examination. Analysis of variance was used to ascertain if the means of the treatment groups differed with respect to the independent variable, scholastic aptitude, as measured by The Pennsylvania State University Scholastic Aptitude Examination (Marks, 1964). A summary of this analysis is found in Table 1.

Table 1

Analysis of The Pennsylvania State University Scholastic Aptitude Examination by Treatment Groups

A. Group Means

	Group A (n = 15)	Group B (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
Means	139.53	141.47	140.93	141.07	141.27	140.85

B. Analysis of Variance

Source of Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	35.10	8.78	<1.00	n.s.
Levels	14	26,578.60	1898.47	154.09	($p < .01$)
Treatments X Levels	56	689.70	12.32		
Total	74	27,303.40			

The test of the treatment effect which was not significant at the .05 level ($F<1.00$) forms a kind of guarantee that the randomization procedure was not in error. The treatment groups were not significantly different with respect to scholastic aptitude as would be expected in a treatment x levels design. The test of the effect of levels of scholastic aptitude yielded a high F-ratio ($F = 154.09$) which exceeds the critical value at the .01 level. The differences in the levels are of course, to be expected in the treatment x levels design.

Correct Responses

First iteration of the program (pretest). Analysis of variance was performed to ascertain if the treatment groups differed in the number of correct responses during the first iteration of the program. There was no instruction prior to the program, so the first iteration served as a pretest, as well as an instructional program.

Table G1 in the Appendix contains the analysis of data for the number of correct responses on the first iteration. Both the F-ratio for the treatment effects (F = 2.24) and the F-ratio for level effects (F = 1.37) were not significant at the .05 level. The mean for groups A to E respectively were 7.67, 9.47, 8.60, 9.60, 11.07, and the grand mean was 9.28.

Second iteration of the program. The analysis of data obtained for the number of correct responses from the beginning of the program to the end of the second iteration is found in Table G2 in the Appendix. Analysis of variance indicated that the F-ratio for treatment effects (F = 199.42) exceeded the critical value for a .01 level test. Thus, the experimental data indicate a statistically significant difference between means for the correct responses to the end of the second iteration of the program. The means for groups A to E respectively were 16.53, 17.40, 28.60, 28.67, 29.13, and the grand mean was 24.07. The F-ratio for effects due to levels was not significant at the .05 level (F = 1.61).

Since a significant F-ratio was found for the treatment effects, it was appropriate to employ Tukey's W-Procedure to determine whether each mean was significantly different from the others.

Results from the W-Procedure indicate that Group C, Group D, and Group E each scored significantly higher ($p < .01$) on the number of correct responses than did either Group A or Group B.

Posttest. The analysis of the number of correct responses on the posttest is presented in Table G3 in the Appendix. Analysis of variance results show an F-ratio for treatment effects (F = 3.97) statistically significant at the .01 level. The means for groups A to E respectively were 25.87, 25.73, 25.80, 27.60, 28.67, and the grand mean was 26.73. The F-ratio for level effects was not statistically significant at the .05 level (F = 1.00).

The Tukey W-Procedure showed significant differences ($p < .01$) between the mean scores of Group E and those of groups A, B, and C. The significance level is .05. Although the mean of Group D was higher than the means of groups A, B, and C, there was no significant difference between any two of these means.

Amount of Instruction

Number of responses to criterion. The analysis of variance for the number of responses required for Ss to reach criterion is found in Table G4 in the Appendix. The F-ratio for treatment effects is clearly statistically significant ($F = 65.83$) at the .01 level. The means for groups A to E respectively were 74.53, 71.93, 54.20, 54.00, 50.47, and the grand mean was 61.03. The F-ratio for level effects was not significant at the .05 level.

Tukey's W-Procedure was used to determine whether there were statistically significant differences between pairs of means. Again, the results of the Tukey W-Procedure indicated that the means of groups C, D, and E were each significantly different from those of groups A and B.

Number of iterations of program to criterion. The range of iterations of the program required by an S to reach criterion were from two iterations for several Ss in groups D and E to seven iterations for one S in Group A.

The data from the number of iterations to criterion are presented in Table G5 in the Appendix. The F-ratio obtained from the analysis of variance conducted on the number of iterations to criterion scores ($F = 37.44$) showed a clearly significant difference at the .01 level. The means for groups A to E respectively were 4.67, 4.60, 2.73, 2.87, 2.53, with a grand mean of 3.48. The F-ratio for treatment effects ($F < 1.00$) was not statistically significant at the .05 level.

The results from the Tukey W-Procedure found in Part C of Table G7 again show statistically significant differences between each of the means of groups C, D, and E, and those of groups A and B.

Time Required to Complete Instruction

Because of the relatively slow (about 100 words per minute allowing for frequent carriage returns) typing rate of the IBM 1050 terminal, those groups which received longer feedback messages (groups D and E) naturally required longer to complete the instruction. The data from time required for Ss to complete the first iteration of the program is found in Table G6 in the Appendix. Analysis of variance yielded a high F-ratio for treatment effects ($F = 32.70$) which was statistically significant at the .01 level. The means of groups A to E respectively were 26.36, 27.80, 29.58, 47.10, 44.44, with a grand mean of 35.0. The F-ratio for the level effects ($F < 1.00$) was not significant at the .05 level.

Tukey's W-Procedure showed that those treatment groups which received long typed feedback messages (groups D and E) required significantly longer to complete the program than those groups which received short feedback messages (groups B and C) and Group A which received no typed feedback message as was expected. In each case, the significance was at the .01 level. Differences between means of other groups were not significant at the .05 level.

Second iteration. The results of the analysis of the amount of time required for Ss to complete the second iteration of the program are contained in Table G7 in the Appendix. The means of groups A to E respectively were 15.92, 15.94, 13.80, 17.62, 18.59, and the grand mean was 16.26. The mean times of the groups receiving long feedback messages (groups D and E) are slightly longer than those of other groups. The F-ratio for treatment effects ($F = 4.62$) was significant at the .01 level. The F-ratio for level effects ($F = 2.25$) was also significant, but at the .05 level.

The Tukey W-Procedure showed that the only significant differences between pairs of means was between that of Group C and those of Group D ($p < .01$) and Group E ($p < .05$).

The level effects from the analysis of variance showed that differences in time required for the second iteration existed between the strata. The means for the time required for the second iteration are found in Table 2. An apparent negative relationship exists between academic aptitude and length of time required to complete the second iteration of the program.

Table 2

Means of Strata for Time Required for Second Iteration of the Program

Ability Strata	Means (in minutes)	N
1 (Highest)	11.52	N=5
2	14.13	N=5
3	11.30	N=5
4	16.56	N=5
5	17.41	N=5
6	14.57	N=5
7	17.25	N=5
8	16.32	N=5
9	13.60	N=5
10	15.21	N=5
11	19.81	N=5
12	15.95	N=5
13	17.30	N=5
14	19.05	N=5
15 (Lowest)	18.41	N=5

Time to criterion The amount of time required for Ss to attain criterion, analyzed in Table G8 in the Appendix, was significantly lower ($p < .01$) for Group C than for the other treatment groups and significantly longer ($p < .01$) for Group D than for any other group except Group E. Differences between pairs of means of all other treatment groups were not significant at the .05 level of significance. The means for groups A to E respectively were 57.21, 56.12, 44.70, 69.00, 64.02, and a grand mean of 58.21.

Accumulated Response Latency

The measure of the time for the S to give a response to an item was the S's response latency for that item. The instruction was programed so that the accumulated response latencies of the S was a measure of the time required for instruction minus the time the terminal was typing messages. Accumulated response latency was a measure of the time the S spent reacting to the stimuli presented by the computer-controlled terminal.

First iteration of the program The data from accumulated response latencies during the first interation of the program are presented in Table G9 in the Appendix. The means of the treatment groups indicate that Ss in Group D acquired longer response latencies during the first iteration of the program than did any of the other four treatment groups. The means for groups A to E respectively were 14.86, 14.55, 14.72, 20.82, 15.41, and the grand mean was 16.08 minutes. An analysis of variance was performed and the F-ratio for treatment effects ($F = 4.09$) was statistically significant at the .01 level. The F-ratio for level effects ($F = 1.11$) was not significant at the .05 level.

The statistically significant differences found ($p < .05$) between pairs of means through Tukey's W-Procedure indicate that the mean accumulated response latency for Group D was longer than those for any other groups. Thus, the statistical significance obtained through analysis of variance represents a difference between the mean of Group D and the means of each of the other treatment groups.

Second iteration of the program. The response latencies accumulated during the second iteration of the program are presented in Table G10 in the Appendix. The mean accumulated response latency time for Group A and Group B are longer than those of the other treatment groups. The means of groups A to E respectively were 7.71, 6.28, 3.89, 4.10, 3.00, and the grand mean was 5.00 minutes. The high F-ratio for treatment effects (F = 19.98) was statistically significant at the .01 level. The F-ratio for level effects (F <1.00) was not significant at the .05 level.

The analysis of possible differences between pairs of means showed that the means of the mean accumulated response latencies for groups A and B were significantly higher than those of groups C, D, and E at the .01 level. The poorer performance on the part of the Ss in groups A and B is similar to the results for several other variables cited earlier in this report.

Response latencies accumulated to criterion. The data obtained from Ss' response latencies accumulated to criterion are found in Table G11 in the Appendix. Analysis of variance shows that the F-ratio for treatment effects (F = 6.56) was statistically significant at the .01 level. The means of groups A to E respectively were 27.11, 24.73, 18.84, 25.18, 18.44, and the grand mean was 22.85 minutes. The effects due to scholastic aptitude levels (F = 1.33) were not significant at the .05 level.

Tukey's W-Procedure identified several pairs of means which were significantly different at the .01 and .05 level. From the differences between means contained in Part C of Table G11, there is a rather clear trend toward differences between means between groups C and E and the other three treatment groups. Some differences were significant at the .01 or .05 levels.

Discussion, Conclusions, and Recommendations

Independent Variables

Analysis of variance performed on data obtained from the scores of Ss on The Pennsylvania State University Scholastic Aptitude Examination (obtained from University files prior to the study) showed no differences between treatment groups at the .05 level of significance. Analysis was

made of the correct response scores attained by Ss immediately following the first iteration of the 30-item program (the pretest). No significant differences were found among treatment groups.

From this evidence, it may be concluded that there were no significant differences among the teatment groups with respect to scholastic aptitude, or to prior knowledge of the concepts.

Rate of learning. In terms of the results obtained during the second iteration and during the time spent by Ss to reach criterion, there were strong indications that Ss who received feedback guiding them to the correct response were learning more effectively and performed better than did those who were forced to "discover" the correct response. The means of groups C, D, and E are significantly better at the .01 level of significance than those of groups A and B on the following criteria:

Number of correct responses to second iteration of program
Number of responses required to attain criterion
Number of iterations of program required to attain criterion
Accumulated response latencies on second iteration of program

These results and their level of significance clearly indicate some of the advantages to be gained by instructing students with a feedback mode that guides them to the correct response.

The results of these comparisons indicate the value of providing information to students during a programed instruction sequence. The findings are in agreement with those of Holland (1966) who concluded, after analyzing several studies, that if a student does not know the correct answer, he might as well be told it.

However, Klaus (1965) in describing the point of view of those programers using the knowledge of results technique stated that they found no advantage in showing the correct answer to learners who provide incorrect responses. Klaus states, "Simple substitutes, such as the statement, 'you are correct' should prove equally effective as a confirmation of the correct answer." In other words, Klaus holds that the appearance of a correct answer serves as reinforcement only when the response is correct; otherwise, the response is wasted.

The poor results demonstrated by the knowledge of results feedback group (Group B) in the present study raise questions as to whether this mode of feedback is adequate for an adjunct auto-instructional program. Most of the studies involving adjunct auto-instruction (Pressey, 1926, 1950: Kaess and Zeaman, 1960; Smith, 1964) have utilized knowledge of results feedback and have only informed the S whether his response was correct or wrong. This type of feedback has been utilized in many types of teaching machines. Data from the present study, however, indicate that providing a student with a statement of which response was correct, or providing him with a statement of why the correct response was correct would be of more value than merely telling him "correct" or "wrong."

From the analysis of the means of all of the variables in this study, it is interesting to note that there was little difference between the means of the knowledge-of-results feedback group and the no feedback group. In none of the variables analyzed was there a significant difference at the .05 level between means of groups A and B.

In the comparisons cited in the first part of this section as being indicative of the advantage of using a feedback mode which guides the S to the correct response, there were no significant differences between groups C, D, and E. Apparently the factor which accelerated the learning of Ss was "being informed as to which response was the correct one." In all of these comparisons, however, the mean of Group E, the combination of feedback modes group, was only slightly, but not significantly better than the means of groups C and D, and in all cases, significantly better than the means of groups A and B. This finding is contrary to those of Swets and his co-workers (1962) who found that "fairly extensive feedback may be detrimental." However, Swets et al. used a small step programmed instructional sequence which resulted in few response errors.

The findings of the present study are in agreement with those of the previously cited study by Bryan and Rigney (1965). Although Bryan and Rigney

found response contingent feedback to be superior to knowledge-of-results feedback, they made no comparison of these feedback modes with knowledge of correct response feedback.

The results of the present study clearly demonstrate some of the inefficiencies of using a "discovery" approach in teaching facts or concepts by an auto-instruction program. Those Ss who were required to discover the correct response demonstrated poorer performance than did those Ss who were merely told the correct answer.

Time required for instruction. The results from the time required to complete the first interation showed clearly that those treatment groups which received long feedback messages (groups D and E) required significantly more time to complete the thirty items in the first iteration than did groups A, B, or C. The time to criterion means show that Group C required significantly less time than did groups D or E, and required the least time of the five treatment groups.

The time required for a student to receive instruction by CAI is a function of the number of instructional frames he completes and also is a function of the amount of time the terminal spends typing messages. Several studies (Gilman, 1967; Wodtke and Gilman, 1966; and Wodtke et al., 1966) have demonstrated that the operating speed of the IBM 1050 terminal is slower than would be ideal for an interface between student and computer. The longer feedback messages require much more time because of the slow (100 words per minute) typing rate of the terminal. However, the new interfaces using cathode ray tube display devices reported by Wodtke (1967) display verbal and graphic material much more rapidly than does the typewriter terminal. Therefore, the additional time required by Ss in groups D and E should be interpreted with caution, since better equipment may soon eliminate these observed differences in instructional time.

The analysis of the data from the first and second iteration of the program and during the entire program indicate that the principal difference between the treatment groups is in rate of learning. Rate of learning may be considered in terms of the amount of instruction that must be presented or in terms of the amount of time required to complete the instruction. When

rate of learning is considered in terms of amount of instruction presented, then a feedback method which guides the student to the correct response is clearly superior to a feedback method which requires the student to discover the correct response. When learning rate is measured in terms of the amount of time required for instruction, a feedback method utilizing short messages requires less instructional time per frame than does one utilizing long feedback messages. However, this difference may be eliminated as better interfaces between computer and student are designed.

Retention. The analysis of variance on posttest scores indicated that the combination of feedback modes group (Group E) was superior to other feedback and no feedback groups. Apparently the amount of information the S derives from the feedback is important in affecting retention.

Because many of the programs used in previous studies have been of the linear low-error-rate variety, little work has been done to ascertain how to deal with errors committed by the student during a program. Glaser (1965) concludes that there have been few studies dealing with "corrective" feedback in verbal learning. Glaser cites one study by Bower which found that providing the correct answer following an incorrect response is a reinforcing event in the same way that confirmation after a correct response is a reinforcing event. The results of the present study indicate the advantages for learning attained by providing the correct response when the S makes an error and also show the retention advantages of providing the S with as much information as possible in the feedback messages. These findings disagree with those of Swets et al. (1962) [sic] who found that extensive feedback may be detrimental.

Results obtained from the posttest also indicate some differences favoring the response contingent feedback groups. On the posttest, the response contingent feedback group (Group D) received the second highest scores and scored higher than any other treatment group receiving a single feedback treatment.

It is interesting to note that Group D accumulated significantly higher response latencies during the first iteration of the program and during their performance to criterion. Apparently the Ss receiving response contingent feedback were contemplating their previous feedback messages during the time period that they might have been responding.

Relationship between analyzed variables and scholastic aptitude. Analysis of variance showed only one significant difference for level effects - instructional time for the second iteration of the program. There was no apparent pattern in the means of strata for the admissible probability score on the first iteration of the program, but examination of the time to criterion means for the 15 strata reveals a negative relation between scholastic aptitude and time for the second iteration.

A low correlation between Ss' rate of learning and academic ability is one of the desirable characteristics of computer-assisted instruction as expressed by Mitzel (1966). Mitzel hypothesizes that computer-assisted instruction

> . . . at its best should offer a distinctly individualized course of instruction in which gaps in the learner's knowledge are filled by means of diagnostic and remedial sequence steps. Thus, it seems to be theoretically appropriate to ask the typical CAI learner to achieve mastery of the content as long as we allow him a reasonable amount of time.

Mitzel concluded that if examining is done at appropriate intervals throughout the program, then every learner should have achieved mastery of the content up to the limits of his capacity.

Futher research is necessary to determine the effects of using various modes of feedback to correct errors. Many forms of programed instruction require the student to reveal, by making some sort of error, the kind of instruction he should receive next. However, most typical programed instruction studies have been conducted with relatively error-free programs and little is presently known about correcting student errors in programed instruction.

The present study should be repeated using a student terminal capable of faster communication and response time than the 1050 terminal.

Also, the present study should be repeated using a delayed retention measure in addition to the immediate retention measure.

The high posttest scores achieved by all groups, including the control group, demonstrated the value of reiterating the program items until the student had answered all items correctly. Several studies (Angell, 1949; Bryan and Rigney, 1956; Paige, 1966) have demonstrated the value of providing feedback on test items. The results of the present study indicate that an additional advantage can result from having an S repeat all unanswered questions until he has correctly responded to each one. This procedure provides a "drill" type of exercise in a test situation in addition to providing feedback.

References

Angell, G. W. The effect of immediate knowledge of quiz results and final examination scores in freshman chemistry. Journal of Educational Research, January, 1949, 42, 391-394.

Bryan, G. L. and Rigney, J. G. An evaluation of a method for shipboard training in operations knowledge. University of Southern California Report No. 18, September, 1956.

Crowder, N. A. Automatic tutoring and intrinsic programming. In A. A. Lumsdaine and R. Glaser (Ed.), Teaching Machines and Programmed Learning. Washington, D. C.: National Education Association, 1960, pp. 286-298.

English, H. B. How psychology can facilitate military training: a concrete example. Journal of Applied Psychology, February, 1942.

Gilman, David A. Feedback, prompting, and overt correction procedures in non-branching computer-assisted instruction programs. Journal of Educational Research, June, 1967.

Gilman, David A. A comparison of the effectiveness of feedback modes for teaching science concepts by means of a computer assisted adjunct auto-instruction program. Unpublished Doctoral Dissertation, The Pennsylvania State University, University Park, Pennsylvania, 1967.

Glaser, Robert. Toward a behavioral science base for instructional design. In R. Glaser (Ed.), Teaching Machines and Programmed Learning II. Washington, D. C.: National Education Association, 1965, pp. 771, 809.

Holland, J. G. Teaching machines: an application of principles from the laboratory. In R. Glaser (Ed.), Teaching Machines and Programmed Learning II. Washington, D. C.: National Education Association, 1965, pp. 66-117.

Jensen, B. T. An independent study laboratory using self-scoring tests. Journal of Educational Research, 1949, 43, 134-137.

Kaess, W. and Zeaman, D. Positive and negative knowledge of results on a Pressey-type punchboard. Journal of Experimental Psychology, July, 1960, 60, 12-17.

Klaus, David J. An analysis of programming techniques. In R. Glaser (Ed.), Teaching Machines and Programmed Learning II. Washington, D. C.: National Education Association, 1965, pp. 118-161.

Marks, E. Evaluation of The Pennsylvania State University Academic Aptitude Examination. University Park: The Pennsylvania State University Report 64-1, April, 1964.

Mitzel, H. E. Five barriers to 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., University Park: The Pennsylvania State University, December 31, 1966.

Paige, D. D. Learning while testing. Journal of Educational Research, 1966, 59, 276-277.

Pressey, S. L. A simple apparatus which gives tests and scores - and teaches. In A. A. Lumsdaine and R. Glaser (Ed.), Teaching Machines and Programmed Learning. Washington, D. C.: National Education Association, 1960, pp. 33-41. Reprinted from School and Society, 1926, Vol. 23 (a).

Pressey, S. L. Basic unresolved teaching machine problems. Theory into Practice, 1962, 1, 30-37.

Pressey, S. L. Teaching machine (and learning theory) crisis. Journal of Applied Psychology, 1963, 47, 1-6.

Skinner, B. F. The science of learning and the art of teaching. Harvard Educational Review, 1954, 24, 86-97.

Skinner, B. F. Reflections on a decade of teaching machines. In R. Glaser (Ed.), Teaching Machines and Programmed Learning II. Washington, D. C.: National Education Association, 1965, 5, 19.

Smith, Philip D. Knowledge of results and continuity of various techniques in presenting a filmstrip as factors in immediate learning and retention. Unpublished doctoral thesis, The Pennsylvania State University, 1964.

Thorndike, E. L. The law of effect. American Journal of Psychology, 1927, 39, 212-222.

Tukey, J. W. The Problem of Multiple Comparisons. Princeton: Princeton University Press, 1953.

Uttal, W. R. On conversational interaction. In J. E. Coulson (Ed.), Programmed Learning and Computer-Based Instruction. New York: John Wiley and Sons, 1961, pp. 171-190.

Wodtke, K. H. Educational requirements for a student-subject matter interface. University Park: Computer-Assisted Instruction Laboratory, The Pennsylvania State University, March, 1967.

Wodtke, K. H. and Gilman, D. A. Some comments on the efficiency of the typewriter interface in computer-assisted instruction at the high school and college level. Experimentation with Computer-Assisted Instruction in Technical Education. Semi-Annual Progress Report prepared by Harold E. Mitzel et al., University Park: Computer-Assisted Instruction Laboratory, The Pennsylvania State University, June, 1966.

Wodtke, K. H., Gilman, D. A. and Logan T. Supplementary data on the deficits in instructional time resulting from the typewriter interface. Experimentation with Computer-Assisted Instruction in Technical Education. Semi-Annual Progress Report prepared by Harold E. Mitzel et al., University Park: Computer-Assisted Instruction Laboratory, The Pennsylvania State University, June, 1966.

APPENDICES

Page No.

Appendix A

Appendix B

Appendix B

(Continued)

Appendix A

Computer-Student Interaction
Sample Program Items

QUESTION (on slide)

01. Which of the following statements describes the relationship between insects and insecticides?

    a) species of insects develop immunity to insecticides through natural selection
    b) individual insects can build up a resistance to an insecticide through repeated contact
    c) insecticides no longer have any effect on insects
    d) insecticides have little effect on insects

COMPUTER-STUDENT INTERACTION (on electric typewriter)

Group A: No Feedback

Correct Response				Incorrect Response			
Computer:	01.	____	____%	Computer:	01.	____	____%
Student:		a	75	Student:		b	50
Computer:	02.	____	____%	Computer:	02.	____	____%

Group B: Knowledge of Results Feedback

Correct Response				Incorrect Response			
Computer:	01.	____	____%	Computer:	01.	____	____%
Student:		a	75	Student:		b	50
Computer:	Correct			Computer:	Wrong		
	02.	____	____%		02.	____	____%

Group C: Knowledge of Correct Response Feedback

Correct Response				Incorrect Response			
Computer:	01.	____	____%	Computer:	01.	____	____%
Student:		a	75	Student:		b	50
Computer:	a is correct			Computer:	a is correct		
	02.	____	____%		02.	____	____%

Group D: Response Contingent Feedback

Correct Response

Computer: 01. _____ _____%
Student: a 75
Computer: Certain species of insects are sometimes capable of developing, through natural selection, an immunity to insecticides.
02. _____ _____%

Incorrect Response

Computer: 01. _____ _____%
Student: b 50
Computer: The individual insects themselves are not able to build up resistance to insecticides. Certain species of insects are sometimes capable of developing, through natural selection, an immunity to insecticides.
02. _____ _____%

Group E: Combination of Feedback Modes

Correct Response

Computer: 01. _____ _____%
Student: a 75
Computer: Correct. a is correct. Certain species of insects are sometimes capable of developing, through natural selection, an immunity to insecticides.
02. _____ _____%

Incorrect Response

Computer: 01. _____ _____%
Student: b 50
Computer: Wrong. the individual insects are not able to build up resistance to insecticides.
a is correct.
Certain species of insects are sometimes capable of developing, through natural selection, an immunity to insecticides.
02. _____ _____%

Appendix B

Table G1

Analysis of Number of Correct Responses On
First Iteration of Program (Pretest) by Treatment Groups

A. Group Means

	Group A (n = 15)	Group B (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
Means	7.67	9.47	8.60	9.60	11.07	9.28

B. Analysis of Variance

Source of Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	95.92	23.98	2.24	n.s.
Levels	14	205.52	14.68	1.37	n.s.
Treatments x Levels	56	599.68	10.71		
Total	74	901.12			

Appendix B

Table G2

Analysis of Number of Correct Responses Through Second Iteration of the Program

A. Group Means

	Group A (n = 15)	Group B (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
Means	16.53	17.40	28.60	28.67	29.13	24.07

B. Analysis of Variance

Source of Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	2,528.67	632.17	199.42	(p <.01)
Levels	14	71.47	5.10	1.61	n.s.
Treatments x Levels	56	208.53	3.17		
Total	74	2,808.67			

C. Tukey's W-Procedure for Differences Between Pairs of Means

	Group B	Group C	Group D	Group E
Group A	0.87	12.07**	12.14**	12.60**
Group B		11.20**	11.27**	12.73**
Group C			0.07	0.53
Group D				0.46

$.05^{W}5,56 = 1.84$

$.01^{W}5,56 = 2.23$

**Significant (p <.01)

Appendix B

Table G3

Analysis of Total Number of Correct Responses on Posttest by Treatment Groups

A. Group Means

	Group A (n = 15)	Group B (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
Means	25.87	25.73	25.80	27.60	28.67	26.73

B. Analysis of Variance

Source Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	106.67	26.67	3.97	(p <.01)
Levels	14	80.27	5.73	<1.00	n.s.
Treatments x Levels	56	375.73	6.71		
Total	74	562.67			

C. Tukey's W-Procedure for Differences Between Pairs of Means

	Group B	Group C	Group D	Group E
Group A	0.14	0.07	1.73	2.80*
Group B		0.07	1.87	2.94*
Group C			1.80	2.87*
Group D				1.07

$.05W_{5,56} = 2.67$

$.01W_{5,56} = 3.21$

*Significant (p <.05)

Appendix B

Table G4

Analysis of Number of Responses to Criterion by Treatment Groups

A. Group Means

	Group A (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
Means	74.53	54.20	54.00	50.47	61.03

B. Analysis of Variance

Source of Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	7,633.15	1,908.29	65.83	(p <.01)
Levels	14	689.55	49.25	1.00	n.s.
Treatments x Levels	56	1,623.25	28.99		
Total	74	9,945.95			

C. Tukey's W-Procedure for Differences Between Pairs of Means

	Group B	Group C	Group D	Group E
Group A	2.60	20.33**	20.53**	24.06**
Group B		17.73**	17.93**	21.46**
Group C			0.20	3.73
Group D				3.53

$.05^{W}5,56 = 5.50$

$.01^{W}5,56 = 6.67$

**Significant (p <.01)

Appendix B

Table G5

Analysis of Number of Iterations
of Program to Criterion

A. Group Means

	Group A (n = 15)	Group B (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
Means	4.67	4.60	2.73	2.87	2.53	3.48

B. Analysis of Variance

Source of Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	67.38	16.85	37.44	(p <.01)
Levels	14	5.92	0.42	<1.00	n.s.
Treatments x Levels	56	25.41	0.45		
Total	74	98.72			

C. Tukey's W-Procedure for Differences
Between Pairs of Means

	Group B	Group C	Group D	Group E
Group A	0.07	1.94**	1.80**	2.14**
Group B		1.87**	1.73**	2.07**
Group C			0.14	0.20
Group D				0.34

$.05W_{5,56} = 0.69$

$.05W_{5,56} = 0.84$

**Significant
(p <.01)

Appendix B

Table G6

Analysis of Lapsed Time for First Iteration of Program by Treatment Groups

A. Group Means

	Group A (n = 15)	Group B (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
Means	26.36	27.80	29.58	47.10	44.44	35.0

B. Analysis of Variance

Source of Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	5,862.40	1,465.60	32.70	(p <.01)
Levels	14	594.30	42.75	<1.00	n.s.
Treatments x Levels	56	2,509.92	44.82		
Total	74	8,966.62			

C. Tukey's W-Procedure for Differences Between Pairs of Means

	Group B	Group C	Group D	Group E
Group A	1.44	3.22	20.74**	18.08**
Group B		1.78	19.30**	16.64**
Group C			17.52**	14.86**
Group D				2.66

$.05W_{5,56} = 6.91$

$.01W_{5,56} = 8.38$

**Significant (p <.01)

Appendix B

Table G7

Analysis of Amount of Time Needed for
Second Iteration of the Program by Treatment Groups

A. Group Means

	Group A (n = 15)	Group B (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
Mean (in minutes)	15.92	15.94	13.80	17.62	18.59	16.26

B. Analysis of Variance

Source of Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	195.20	48.80	4.62	(p <.01)
Levels	14	333.20	23.80	2.25	(p <.05)
Treatments x Levels	56	591.92	10.57		
Total	74	1,120.32			

C. Tukey's W-Procedure for Differences Between Pairs of Means

	Group B	Group C	Group D	Group E
Group A	0.02	2.12	1.70	2.67
Group B		2.14	1.68	2.65
Group C			3.82*	4.79**
Group D				0.97

$.05^{W}5,56 = 3.34$

$.01^{W}5,56 = 4.05$

*Significant (p <.05)

**Significant (p <.01)

Appendix B

Table G8

Analysis of Lapsed Time on Line From Beginning of Program To Criterion Achievement by Treatment Groups

A. Group Means

	Group A (n = 15)	Group B (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
Mean (in minutes)	5[illegible].1	56.12	44.70	69.00	64.02	58.21

B. Analysis of Variance

Source of Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	5,070.00	1,267.50	19.02	(p <.01)
Levels	14	1,842.40	131.60	1.95	n.s.
Treatments x Levels	56	3,730.72	66.62		
Total	74	10,643.12			

C. Tukey's W-Procedure for Differences Between Pairs of Means

	Group B	Group C	Group D	Group E
Group A	1.09	12.51**	11.79**	6.81
Group B		11.42**	12.88**	7.90
Group C			24.30**	19.32**
Group D				5.81

$.05W_{5,56} = 8.41$

$.01W_{5,56} = 10.20$

**Significant (p <.01)

Appendix B

Table G9

Analysis of Accumulated Response Latencies,
First Iteration of Program, by Treatment Groups

A. Group Means

	Group A (n = 15)	Group B (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
Mean (in minutes)	14.86	14.55	14.72	20.82	15.41	16.08

B. Analysis of Variance

Source of Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	429.12	107.28	4.09	(p <.01)
Levels	14	407.44	29.11	1.11	n.s.
Treatments x Levels	56	1,470.00	26.25		
Total	74	2,306.56			

C. Tukey's W-Procedure for Differences
Between Pairs of Means

	Group B	Group C	Group D	Group E
Group A	0.31	0.14	5.96*	0.55
Group B		0.17	6.27*	0.86
Group C			6.10*	0.69
Group D				5.41*

$.05^{W}5,56 = 5.28$

$.01^{W}5,56 = 6.40$

*Significant
(p <.03)

Appendix B

Table G10

Analysis of Accumulated Response Latencies,
Second Iteration of Program, by Treatment Groups

A. Group Means

	Group A (n = 15)	Group B (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
Mean (in minutes)	7.71	6.28	3.89	4.10	3.00	5.00

B. Analysis of Variance

Source of Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	225.40	56.35	19.98	(p <.01)
Levels	14	37.24	2.66	<1.00	n.s.
Treatments x Levels	56	157.92	2.82		
Total	74	420.56			

C. Tukey's W-Procedure for Differences
Between Pairs of Means

	Group B	Group C	Group D	Group E
Group A	1.43	3.82**	3.61**	4.71**
Group B		2.39**	2.18**	3.28**
Group C			0.21	0.89
Group D				1.10

$.05W_{5,56} = 1.73$

$.01W_{5,56} = 2.10$

**Significant
(p <.01)

Appendix B

Table G11

Analysis of Accumulated Response Latencies to Criterion by Treatment Groups

A. Group Means

Mean (in minutes)	Group A (n = 15)	Group B (n = 15)	Group C (n = 15)	Group D (n = 15)	Group E (n = 15)	Grand Mean (N = 75)
	27.11	24.73	18.84	25.18	18.44	22.85

B. Analysis of Variance

Source of Variation	df	Sums of Squares	Mean Square	F Ratio	Significance
Treatments	4	931.20	232.80	6.56	(p <.01)
Levels	14	661.08	47.22	1.33	n.s.
Treatments x Levels	56	1,986.88	35.48		
Total	74	3,579.16			

C. Tukey's W-Procedure for Differences Between Pairs of Means

	Group B	Group C	Group D	Group E
Group A	2.38	8.27**	1.93	8.67**
Group B		5.89	0.45	6.29*
Group C			6.34*	0.40
Group D				6.74*

$.05^{W}_{5,56} = 6.15$

$.01^{W}_{5,56} = 7.46$

*Significant (p <.05)

**Significant (p <.01)

NUMERICAL AND VERBAL APTITUDE TESTS ADMINISTERED AT THE CAI STUDENT STATION

Joseph L. French and John Tardibuono

This reporting period has been characterized by reorganization of material. Considerable off-line testing has been completed and the items in all tests have been reorganized as suggested by the new data.

The students participating in this phase included those enrolled in an Electrical Engineering associate degree program, a Drafting and Design Technician associate degree program, a group of adult speech clients in residence at the University under a Bureau of Vocational Rehabilitation program, and several secondary school students. Each subject responded to an 80-item numerical test (COMPAT-N), a 25-item verbal test (Pretest), and a 204-item verbal test (COMPAT-V) administered in small group settings. The items were analyzed at the Computation Center with a program yielding the following data:

	Kuder Richardson 20	Estimated Inter-Item Correlation	Average Item-Total Test R
COMPAT-N	.937	.267	.516
Pretest	.624	.163	.403
COMPAT-V			
first 133 items	.977	.139	.373
last 71 items	.994	.191	.437

Now four tests exist. They include: 1) reorganized 80-item numerical test called COMPAT-N, 2) a reorganized 25-item verbal test composed of approximately every tenth item when the 229 verbal items were arranged in order of difficulty known as the "pretest," 3) a new 80-item verbal test composed of approximately every third item of the 229 verbal items known as COMPAT-V80, and 4) a reorganized COMPAT-V which is composed of 204 items divided into four 80-item sections each of which contain 38 items used in another section as described in the last report. Subjects will take all four tests. The new COMPAT-V80 will provide a hedge against the possibility that the 204 items are not adequately scaled in difficulty to secure sufficient reliability when the COMPAT program is employed.

During this reporting period a computer program was developed for the verbal pretest. This program presents 25 slides consecutively and types out a score based on the number of items answered correctly. Based on the pretest score a proctor selects one of the four sections of COMPAT-V for the subject.

It was concluded after reviewing the results of the off-line testing that the numerical items are at the correct level of difficulty. Easier items have not been added to COMPAT-N but the items have been reordered.

Four copies of all materials have been completed and are now ready for use at the terminals. However, processing the 2 x 2-inch slides to meet project specifications consumed more time than anticipated during the past two months and not as many subjects were able to complete on-line testing as anticipated. The arrival of the 1500 system and other project priorities for the 1410 system may slow data collection in the months ahead.

During the next reporting period the four COMPAT tests (COMPAT-N, pretest, COMPAT-V, and COMPAT-V80) and the regular edition of the College Level of the Henmon-Nelson will be used with appropriate groups of subjects. Interpretive scores will be developed for the COMPAT tests based on performance and national norming of the Henmon-Nelson Test. Reliability coefficients will be determined. Based on data collected during the next period, a decision will be reached in regard to the desirability of COMPAT-V versus COMPAT-V80. The materials will be converted to the IBM 1500 computer-assisted instruction system.

SPELLING AND COMPUTER-ASSISTED INSTRUCTION

Helen L. K. Farr and Harriett A. Hogan

Before the course authors began writing programs for computer-assisted instruction (CAI), they asked members of the English departments in two two-year colleges to suggest content topics that might also be of some educational benefit for the college students who were to serve in the investigations. Every teacher mentioned that a great many students needed remedial work in spelling.

From the standpoint of tutorial CAI research, remedial spelling seemed to be eminently suitable for the following reasons:

1. The research literature on the teaching of spelling is extensive. Futhermore, the pattern of diagnostic testing, followed by teaching to remedy the diagnosed deficiencies, and evaluation by equivalent testing after teaching is a pattern generally accepted by teachers (Horn, E., 1954; Horn, T., 1967; Marksheffel, 1964).
2. Spelling is one of the few segments of the English curriculum on which there is almost universal agreement; i.e., accepted variations in the spelling of words are few (Horn, E., 1954).
3. English teachers agree that spelling is a basic skill that should be mastered to some standard criterion level by all students (Horn, E., 1954).
4. Because of its uniformity and its discrete nature, spelling data can be easily quantified and analyzed by statistical methods.
5. The variety in kinds of spelling errors allows for clear specification of errors and the development of remedial teaching objectives and programs.

6. The college English teachers consulted did not feel that they had the time nor the responsibility for teaching basic spelling to post-high school students; therefore, they welcomed an innovation that would undertake to teach spelling.

Following the suggestion of the English teachers, a spelling program was prepared and used in a preliminary field trial to obtain information about its "takeability" and effectiveness (Hogan and Farr, 1966). In that field trial, a positive correlation was found between the achievement gain scores and the number of remedial course segments studied. Confidence in the validity of the pretest as a diagnostic instrument was also established.

Spelling Transfer Study

Although the field trial had indicated that the CAI spelling course was an effective means for two-year technical students to learn spelling, no attempt had been made to examine whether the students transferred their increased knowledge of spelling rules (as indicated by criterion scores) to off-terminal, non-test writing situations. Consequently, the investigation of transfer was selected as the primary focus of the present study.

It was assumed that after the diagnosis and identification of spelling errors, followed by remedial instruction, students would spell more accurately than they had in non-test, writing situations before instructions. Further, it was assumed that students who were merely informed of the number and kinds of spelling errors that they had made on the pretest would show less improvement in non-test, writing situations than the students who had been given the same information and also received prescribed remedial instruction. Accordingly, the main dimension investigated in this study was the difference in the extent to which the instructed students transferred their demonstrated spelling skills, in contrast to those who had not been instructed.

Method

Materials

The materials used in this study were of three types: a) two samples of expository writing done by the students on topics they selected from a list provided (Appendix A); b) selected segments of the CAI spelling program (Hogan and Farr, 1966); and c) a 44-item attitude questionnaire about CAI.

The topics for these writing samples were deliberately planned by the authors to center the attention of the students on the thought content of their writing, rather than on spelling or other mechanics of composition. Furthermore, no mention of spelling was made when the writing assignments were given, and the attitude questionnaire was concerned with CAI and the students' reactions to it, rather than to any aspect of spelling.

From the CAI spelling program, all of the students were given three segments on-line: orientation, word study, and the diagnostic test. The first segment was a short one dealing with orientation to the terminal equipment and its operation. The second segment provided the students with further opportunity to familiarize themselves with appropriate on-line procedures, as well as preparation for maximum benefit from spelling study. The diagnostic segment consisted of a 37-word test which included 50 possible error items representing nine categories of spelling errors (Appendix B). For example, the word "piece" was designated as both a possible homonym error and as a possible "i-e" error.

In addition to these three segments, the students in the experimental group took the on-line instruction prescribed by their diagnostic test performances, and a 37-word 250-item, on-line posttest similar to the diagnostic test.

Students

The 48 students participating in this study were enrolled in post-high school, two-year technical programs. The experimental group consisted of 23 students; the control group contained 25 students. Two of the experimental students were girls; one of the control students was a girl.

2

Procedure

In the field trial, English instructors had, for the most part, acted as recruiters of students. At that time, although the opportunity for experience with a new educational technology had been mentioned, the English teachers had more often emphasized the possible benefit to the students' spelling and English skills.

In this study, students in the same two-year college programs were asked by their instructors in technical subjects to participate in an investigation of CAI. CAI was defined as a new educational technology.

Students were told that: a) the researchers preferred male students (but would accept female students); b) the CAI subject matter for the experiment would be spelling; c) the participants would be paid for the time they spent participating in the experiment only if they completed their assigned tasks; and d) the estimated maximum time required of participants would be seven hours.

From the two institutions contacted, a total of 92 students volunteered to participate. Appointments for all volunteers were scheduled by the CAI terminal proctors in the institutions. At their first appointments, all volunteers were assigned the following three tasks.

1. They were asked to write "a page and a half or for about half an hour" on one of the topics presented to them (Appendix A).
2. After the written assignment had been completed, all volunteers were put on-line and given three segments of the CAI spelling programs: a) the machine-orientation segment; b) a brief instruction on word-study techniques for spelling; and c) the diagnostic spelling test. (Within the test segments of the program, students were informed, by the system, of the number and kinds of spelling errors they made on each of the 50-item tests which covered nine error categories.)

3. Then, the volunteers were asked to complete a 44-item pencil-and-paper questionnaire about their attitudes toward CAI. Finally, when they had finished the questionnaire, the students were told that they would be notified in a few days about whether or not additional appointments would be required.

The course authors then inspected all the terminal printouts of the diagnostic spelling test applying the following selection criterion to each. If a student had made four or more errors in any of the nine possible spelling error categories, he was retained for the study. If he had made three or fewer errors in each of the nine error categories, he was eliminated from the study, since he was, for the purposes of this study, judged not to be in need of spelling instruction.

Those students who had qualified for participation in the study were randomly assigned to either the experimental or control group and were notified, by the terminal proctors, of their next appointments.

At their second appointments (usually a week later) the experimental and control students had different tasks. The control students were again asked to write on a topic from the same list as before. The experimental students began CAI spelling instruction as prescribed by the diagnostic test. If additional appointments were necessary to complete the spelling instruction, they were scheduled until all the indicated remedial sections were finished by the experimental students.

When the experimental students had completed their CAI instruction and taken the proofreading and posttest segments of the CAI program, they were again asked to complete the attitude questionnaire. (The control subjects were not asked to do this.)

Approximately a week after each experimental student had completed his on-terminal experience, he was called back and asked to write again on one of the listed topics. Then he, too, was thanked by the proctors and told that his work in the study was done.

The student-terminal printouts from all of the volunteers were inspected by the authors who discarded some of them, thereby excluding the data of those students from the study. There were three reasons for exclusion: a) a diagnostic spelling test score indicating, according to the criterion built into the CAI spelling program, that remedial instruction was not needed; b) the report or evidence of machine trouble; and c) failure of the volunteer to complete all parts of his assignment.

In summary, for the reasons just stated, usable data for the study was limited to 48, after the data from 47 of the volunteers had been disqualified.

A repeated measures analysis of variance design was employed for the inspection of the data in this study (Lindquist, Type I design).

Results

The effectiveness of transfer from this CAI spelling program was assessed on two dependent measures: a) performance on the spelling pretest and posttest (minimal transfer); and b) performance on two writing assignments (remote transfer).

Spelling test data for the experimental students were analyzed within a 2 x 2 factorial design (two student groups x two test scores and/or pretest and posttest scores) with repeated measures on one factor (pretest and posttest) (Lindquist, 1953). The results of this analysis (see Table 1) indicated that there was a significant improvement in spelling test performance on the CAI posttest ($p < .05$). However, when absolute criterion performance was examined, it was found that the experimental students had entered the program performing at a 66% level of accuracy, and their end-of-program criterion performance was only 76% (see Table 2). These results indicate that some learning had occurred, but that the students' difficulties with spelling had by no means been entirely eliminated.

Table 1

Analysis of Variance Summary Table of Spelling Test Scores for Experimental Students: a 2 x 2 Factorial Design Consisting of Student Groups by Pretest-Posttest, with Pretest-Posttest as a Repeated Measure

Source of Variance	df	SS	MS	F
Subjects	20	1228.62	61.43	-
"B" Colleges A and B	1	289.75	289.75	5.86*
Error Within (B)	19	938.87	49.41	-
Within Subjects	21	551.50	26.26	-
"A" (Pretest-Posttest)	1	242.88	242.88	16.81**
"A" x "B"	1	34.11	34.11	2.36
Error	19	274.51	14.45	-
Total	41	1780.12	43.42	-

*F (df. 1, 22) = 16.81, p <.01

**F (df. 1, 22) = 5.86, p <.05

Table 2

Pretest and Posttest Means

Students	Pretest Experimental		Pretest Control		Posttest** Experimental	
	s.d.	mean	s.d.	mean	s.d.	mean
Number Right*	4.94	33.19	7.07	34.00	4.48	38.75
Number Wrong	-	16.81	-	16.00	-	11.25

*The maximum score possible was 50.

**The posttest was not given to the control students.

The data from Writing Sample 1 and Writing Sample 2 produced no significant differences on the following four analyses: a) the total number of errors; b) the total number of errors in each spelling category; c) the mean number of errors per 100 words written on each sample; and d) the type-token ratios from the first and last 50 words on the two samples (see Table 3). The type-token ratio as it is used here refers to the ratio of different words used (types) to the total number of words (tokens). To avoid the possibility that the absolute numbers of words written and errors made might be functions of the lengths of the students' written samples, the latter two measures (c and d) were analyzed. For the type-token measure on each sample, the first and last 50 words were arbitrarily selected, and the mean type-token ratio on the combined 100 words from each student's Writing Sample 1 were compared with the mean type-token ratio on the combined 100 words from each student's Writing Sample 2.

Table 3

Selected Means from Writing Sample 1 and Writing Sample 2

	Students			
	Experimental		Control	
	Sample 1	Sample 2	Sample 1	Sample 2
Mean Number of Written Words	216.8	206.2	254.3	242.8
Mean Number of Errors per 100 Words	3.55	2.85	2.96	3.38
Most Common Error Categories: Mean Number of errors per 100 Words				
Demons	1.98	1.49	1.92	2.76
Syllabification	1.70	1.35	1.36	1.56
Discrimination	1.04	1.17	0.88	1.40

The experimental students were found to write fewer words ($p < .05$) on the two writing assignments than the control students. However, since all of the students accepted for this study were randomly assigned to either the experimental or control groups, the statistically significant difference in the mean number of words produced does not seem to have been influenced by the experimental treatment. That is, the experimental and control groups wrote, respectively, 10 and 12 more words on the first writing sample than they did on the second. The difference may simply indicate a stylistic difference in fluency of written production for the members of the two groups.

Indeed, none of the measures investigated on the two written samples seems to have revealed any differential effects attributable to the experimental treatment, namely, the CAI spelling program.

Discussion and Implications

The authors recognize that additional information to support the claim that this CAI spelling program does indeed teach spelling rules and information (which the students do not transfer to non-test situations) might have been obtained if the control students had taken the posttest as well as the diagnostic test. Such a procedure is recommended for future studies.

The findings of this study are, on the whole, however, in accord with other reported evaluations of spelling instruction methods (Horn, E., 1954; Horn, T., 1967; Marksheffel, 1964). That is, there is evidence from the experimental students' performances on the diagnostic test and posttest that post-high school technical students can be taught spelling by means of this CAI program.

However, this study also looked at the dimension of transfer of knowledge. The transfer of spelling knowledge which had been demonstrated on clearly identified tests was sought on writing assignments not identified to the students as test situations. Consequently, the reported success in improving spelling which is based on test performance is tempered by the failure to demonstrate transfer in a non-test situation (i.e., in general writing). Even though most studies on methods of spelling instruction do not include the transfer dimension, transfer is, presumably, the real or crucial measure of the success of an instructional program in spelling.

If one accepts the common definition of spelling skill as "the ability to spell correctly the words a student needs to spell in his school experiences and in his future life circumstances," then there is little evidence that any instructional material or method has "succeeded" in teaching spelling skills. This evidence is strengthened by two external facts: a) all of the students in the study were high school graduates; and b) the students had attended a number of different elementary and secondary schools in a variety of school districts. The latter fact is of particular import because Pennsylvania does not have a uniform textbook or spelling course policy.

The finding that there was no significant difference between error categories on the writing tasks performed before and after the CAI spelling instruction indicates that the spelling program did not differentially affect, to a statistically significant degree, the students' performances in designated error categories. In fact, the students made a considerable number of spelling errors on both of the CAI spelling tests as well as on the written assignments when they, presumably, used words of their own choice (see Table 3).

The improvement demonstrated by the experimental students' scores on both their posttests and their second writing samples tend to support the original hypotheses. Those hypotheses were: a) that students who used the CAI program would make fewer errors after instruction; and b) that students who did not receive such instruction would not make fewer errors.

Two possible relations were considered: a) that the experimental students made fewer errors on their second written samples because they avoided using words they did not know how to spell; and/or b) that they used the same words repeatedly (thus decreasing the number of spelling errors recorded, since repetitions of the same error were counted only once). Aside from directly asking the students about whether they consciously avoided words because of uncertainty about their spelling, there seemed no feasible way of checking on the first possibility. Futhermore, the mean rate of nearly three or more words misspelled per 100 words written indicated the students' lack of success, even if they were trying to avoid troublesome words.

The determination of the type-token ratio on the other hand seemed to provide an easy check on whether the same words were being used repeatedly. In this analysis, the first and last 50 words in each written sample were arbitrarily selected as the segments for inspection.

If a student were repeating the words he felt sure he could spell correctly, one might logically expect this practice to be more common on his second writing sample, after his attention had been called to his spelling errors. However, for both the experimental and control groups the type-token ratios were almost identical on Writing Sample 1 and Sample 2. The mean ratio for the experimental students was 76 on both samples. For the control

students, the mean ratio was 77 on Sample 1 and 76 on Sample 2. Consequently, there did not seem to be any deliberate attempt to use fewer different words on the second writing sample. In other words, the students did not differentially exercise the opportunity for reducing spelling errors by limiting either the total number of words written or the number of different words used.

The greatest number of errors per 100 words on Sample 1 and Sample 2 were in the same three spelling error categories: "demons," syllabification, and discrimination (see Table 3). "Demons," as the term was used in this study, indicated a somewhat broader category than is usual in spelling contexts. In checking these writing samples, the traditional lists of "100 Spelling Demons" were augmented by all words containing the same fundamental problems as those in the usual lists. That is, all words were considered to be demons if errors occur because no clear spelling clue is given in the sound of the word (e.g., different vowels with the same sound, silent letters). Syllabification, as an error category here, covered the omission of pronounced syllables, the addition of extra pronounced syllables, and the improper breaking of words continued on the next line. Discrimination errors were those in which similarities of meaning, sound, and/or written forms of different words are confused.

All three of these error categories lie at the extreme ends of the scales of irregularity and/or scholarly disagreement in English orthography (Hall, 1961; Hanna et al, 1966). Thus, it might be assumed that these three error categories are the most troublesome ones for the students as well as for the scholars, and that this is also the case even when they are under no instruction to demonstrate their skills with them. It seems evident that neither many years of spelling instruction nor a few hours of this CAI course has successfully assured that in non-test situations students will spell correctly the words they need or use in their writing.

Conclusions

The main conclusions from this study to investigate the effectiveness of transfer from this CAI spelling course were: a) students in two-year technical courses demonstrated a significant improvement in their spelling abilities as measured by spelling tests, after they had received instruction

from this CAI program; and b) when they were not specifically told, on writing assignments, that "spelling counts," post-high school technical students made far more errors in certain categories than they made in those categories during test situations.

Because, for the non-test writing samples gathered in this study, students were free to use - or not to use - any words they wished, it is reasonable to conclude that the ones they used are the words they "need" to know how to spell: the very words that formal spelling instruction aims to teach (Horn, E., 1954). Likewise, the large number of errors appearing in the writing samples can be taken as an indication of the failure of traditional, classroom spelling instruction to achieve its commonly stated goal of teaching "needed" words.

The novelty of CAI experience and the appeal of its technology for students in technical courses did not produce a markedly greater amount of transfer of spelling ability to non-test situations than had other methods of instruction earlier in the students' educational experiences. This, however, should not be regarded as a failure of this CAI course. Unanimously, on the attitude questionnaire students indicated that they felt that spelling was a subject that could be effectively taught by CAI, and their scores indicate that it was. Since, CAI does not seem to affect their spelling in general writing situations any more than any other method of instruction, it might be of particular value to consider the ways in which students - and teacher - approach spelling instruction. Such consideration is recommended if the present CAI course is revised in the future.

Perhaps, if adults who are in need of remedial spelling are to be "all-around" competent spellers, the spelling course must include an attempt at attitude change, so that spelling is accepted as a skill with "all-around" importance. Otherwise, for many students, and especially for students in technical courses, spelling is likely to remain a subject in which students do as well as possible only on spelling tests and in situations where they know that "spelling counts."

References

Hall, R. A., Jr. Sound and spelling in English. Philadelphia: Chilton Books, 1961.

Hanna, P. R.; Hanna, J. S.; Hodges, R. E., and Rudorf, E. H., Jr. Phoneme-grapheme correspondences as cues to spelling improvement. Washington, D.C.: U. S. Department of Health, Education and Welfare, Bureau of Research of the Office of Education, 1966.

Hogan, H. A. and Farr, H. L. K. Communication skills. 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., December 31, 1966.

Horn, E. Teaching spelling. Washington: National Education Association, 1954.

Horn, T. D. Handwriting and spelling. Rev. Ed. Res., 1967, 37, 168-177.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin Company, 1953.

Marksheffel, N. D. Composition, handwriting, and spelling. Rev. Ed. Res., 1964, 34, 177-186.

APPENDICES

Appendix A

CAI Study: Farr--Hogan

Write at least a page and a half on one of the following topics.
If you want to combine any of these topics you may.
It should only take you half an hour or so to write this article.
Put your CAI student number and the date in the upper right corner of your paper.

Topics:

The Computer in Education or Industry or Business
Considerations in Selecting a Career
Specialized Preparation for Work Today and in the Future
Suggestions for Technical Training

Appendix B

Nine Instructional Areas in CAI Spelling Program:

1. Plurals
2. Prefixes and suffixes
3. Final e
4. The ie-ei combination
5. Syllables
6. Contractions and compounds
7. Discriminating between similar words
8. Homonyms
9. Demons

EFFECTS OF REDUCING VERBAL CONTENT IN COMPUTER-ASSISTED INSTRUCTION PROGRAMS

David Alan Gilman and Nancy Harvilchuck

Several studies summarized by Evans (1966) have compared the presentation of instructional programs by teaching machines with the presentation of the same program by programed texts. A typical finding in these studies is that presentation by teaching machine required from 10 to 40 per cent more instructional time with no significant increase in learning. Several recent studies summarized by Wodtke (1967) indicate that additional time is required for presentation of instructional programs by means of computer-assisted instruction using a typewriter terminal compared to presenting the same material by programed texts.

Several rather obvious possibilities exist for the reduction of time in computer-assisted instruction (CAI). Among these are 1) faster operation of the system hardware, 2) the development of more efficient branching strategies, and 3) the reduction of the amount of verbal material presented by the typewriter terminal.

This study investigated the effects resulting from reducing the verbal content in a CAI program. The rationale for this treatment is that those students of low verbal ability may comprehend material better when taught by programs with short, concise sentences containing few unusual words, rather than when they are taught by programs with a high verbal content.

This study tested three hypotheses regarding the verbal content of computer-assisted instruction programs. The expected findings were as follows:

1) A low verbal content program requires less instructional time than a high verbal content program.

2) Greater comprehension (posttest performance) results from having studied a low verbal content program, as contrasted with having studied a high verbal content program.

3) There is a higher correlation between learning (posttest performance) and verbal intelligence (California Test of Mental Maturity) for students studying a high verbal content program than for students studying a low verbal content program.

Rationale

Evans (1966) summarized several studies comparing the efficiency of presenting programs by teaching machines rather than by programed texts. He concludes that in terms of instructional efficiency "the teaching machines rarely broke even, although they often broke down." In most cases, presentation by teaching machine resulted in an instructional deficit of from 10 to 40 per cent with no significant increase in learning.

Wodtke (1967) cites studies which demonstrate that the typewriter interface in a computer-assisted instruction system is inefficient even for highly verbal college students. When identical instructional programs were administered on terminal and by programed text, there was an increase in instructional time of 25 per cent in the on-terminal group with no commensurate increase in learning. In a second study employing a more highly verbal program, the increase in instructional time was 75 per cent for the on-terminal group with no significant difference in learning when compared to the programed test group. Wodtke attributed the time decrement to the slow type-out rate of the typewriter (approximately 100 words per minute) which is substantially slower than the normal reading speed of the typical student.

The conclusion that students learn better from a program of low verbal content has face validity, but this is also substantiated by studies in reading comprehension and programed instruction theory.

Reading Comprehension. Reading material is said to be "readable" when it can be read by all those who are literate, rather than only those who are highly literate.

Several attempts have been made to identify the principle factors which cause students difficulty in reading and comprehending books and printed material. Most studies (Dale and Tyler, 1934; Lorge, 1939; Flesch, 1948; and Dale and Chall, 1948) find sentence length and the percentage of uncommon words as the major factors contributing to reading comprehension. Chall (1958) summarized the effect of sentence length on readability:

> The two factors common to most reading formulas are vocabulary difficulty and sentence length. Almost every study in reading has found a significant relationship between sentence structure and comprehension difficulty. The most popular way of estimating sentence structure is by sentence length.

Thus, the reduction of verbal content in an instructional program could significantly increase comprehension of programed materials. This may be particularly true for the less literate students.

Programed Instruction. Holland (1966) proposes the "blackout ratio" as a measure for determining the extent to which material is programed. The measure consists of determining which portions of a given set of instructional materials play no role in eliciting correct answers. The first step in performing the measure is to obliterate everything that can be obliterated without changing the program's error rate. All phrases not supporting the answer are blacked-out by black crayon.

The blackout is next validated by testing the blacked-out normal (unblacked-out) programs to demonstrate that error rate has not been influenced to a significant extent.

The percentage of total words blacked-out is an index of the amount of material that bears no contingent relation to the answer. The lower the percentage, the more completely the material has been programed.

Holland has demonstrated that in some programs blacking-out 69 per cent of the words produced no change in criterion test performance. Holland also demonstrated that success on later items in a program was not affected by having studied frames containing blacked-out material in earlier items.

The blackout ratio is proposed as a measure of the extent to which material is programed, but there are also possibilities for utilizing it as a measure of instructional effectiveness. Holland concludes:

> A considerable advantage is, of course, expected to result through the technique of programed instruction. It would, therefore, be reasonable to expect greater learning with programs yielding a low blackout ratio than with programs yielding a high blackout ratio.

Logically, a program of low verbal content could be expected to have a lower blackout ratio (programed to a greater extent) and thus could provide a better learning situation.

In terms of time expended during instruction, reading comprehension, and better learning due to studying more highly programed material, a program of low verbal content should provide a more efficient learning situation than a program having a high verbal content. These results should be manifested as lower time required for instruction and higher posttest performance by students studying the low verbal content programs.

The low verbal content program may also provide better instruction for persons of low verbal ability. These results should be manifested as lower correlations between verbal intelligence and posttest performance by students studying the low verbal content program.

Method

Subjects

The subjects were 36 students from the tile setting and plumbing programs of Williamsport Area Community College, Williamsport, Pennsylvania. The students were selected because they had not yet demonstrated high academic ability and had not received previous instruction in the content materials (significant figures) to be used in the study. All Ss were naive with respect to educational experimentation.

Materials

Two versions of a CAI program used in a previous study (Logan and Wodtke, 1966) were prepared. The subject of the programs was significant figures, or performing calculations to the proper degree of accuracy in a scientific experiment. Subjects who responded incorrectly to a question were provided with typed feedback messages that were contingent on their response to the question.

The two versions of the program differed in only one respect. The high verbal content (HVC) program contained the frames as they were originally written. In most cases, the instruction and feedback messages were long,

complex sentences. The low verbal content (LVC) version of the program was rewritten in short, concise sentences and the unusual words were replaced with more common words. The verbal content of the messages was shortened as much as was possible without losing the meaning or content of the messages.

Both versions of the CAI program were supplemented by three static display diagrams.

Equipment

CAI equipment consisted of two IBM 1050 terminals. Instruction was teleprocessed a distance of 55 miles between the terminals located at Williamsport, Pennsylvania, and an IBM 1410 computer located at University Park, Pennsylvania.

Tests

Prior to the instruction, Ss were administered the California Mental Maturity Test and verbal and nonverbal I. Q. scores were recorded. A pretest consisting of five items was devised to insure that Ss had no prior knowledge of significant figures.

A posttest consisting of 18 multiple-choice items was also constructed. The test contained items designed to measure both mastery and transfer. In an earlier pilot study involving 30 subjects, the test yielded a KR-20 reliability of .86, an average item difficulty index of .68, a mean of 12.52, and an average item-total score correlation of .53.

Design

Ss were randomly assigned to two treatment groups of 18 Ss per group. Ss were pretested with the five-item pretest. No Ss answered more than two questions correctly and most Ss answered all questions incorrectly.

Both treatment groups received instruction through the 1050 terminal. The time required for instruction was recorded by the computer. Immediately following the instruction, the 18-item posttest was administered.

Results

Table 1 shows the means of the two groups. There is a slight difference between the posttest means favoring the LVC group. This small difference was in the direction hypothesized, but was not statistically significant ($p < .10$).

An important factor in programed learning is the time required. Table 1 also shows the mean times required by students to complete the instruction. The difference between the means of the groups was significant ($p < .05$). The instruction by means of the low verbal content program required significantly less time than instruction from the more verbal program.

Table 1

Comparison of Mean and Standard Deviations of Pretest, Posttest, and Time Required for High Verbal and Low Verbal Instructional Programs

		10-Item Pretest	18-Item Posttest	Time Required
(A)	High Verbal Program (HVC) (n = 18)	mean = 0.11	13.26	1 hr. 34.7 min.
		sigma = 0.01	2.38	13.36 min.
(B)	Low Verbal Program (LVC) (n = 18)	mean = 0.18	13.30	1 hr. 20.1 min.
		sigma = 0.01	1.18	27.42 min.
(C)	t ratio	0.45	0.294	2.278
(D)	Significance	n.s.	n.s.	($p < .05$)

The correlation between posttest score and the California Mental Maturity Test score for verbal intelligence was 0.310 for the LVC treatment group. This correlation was not significant at the .05 level. The correlation between posttest score and verbal intelligence was 0.520 for the HVC treatment group. This correlation was significant at the .05 level.

These results are in the hypothesized direction. The hypothesis to be tested was: There is no difference in the correlation between verbal intelligence and posttest score for the highly verbal program and the correlation between verbal intelligence and posttest score for the low verbal program.

The hypothesis was tested by the method prescribed by Wert, Neidt, and Ahman (1954). Results are as follows:

$r_{HV} = 0.520$ $\qquad$ $r_{LV} = 0.310$

$Z_{HV} = 0.5763$ $\qquad$ $Z_{LV} = 0.3205$

$n_{HV} = 18$ $\qquad$ $n_{HV} = 18$

$$\frac{x}{\sigma} = \frac{Z_{HV} - Z_{LV}}{\sqrt{\frac{1}{n_{HV}-3} + \frac{1}{n_{LV}-3}}} = \frac{0.5763 - 0.3205}{\sqrt{\frac{1}{15} + \frac{1}{15}}}$$

$= 0.66$

A table of the normal curve indicates the probability of obtaining a difference in the predicted direction as large or larger than the one obtained, as a result of random sampling from a single population, is 0.255. The null hypothesis cannot be rejected since the probability accompanying the difference is larger than the 5 per cent level.

It should be noted, however, that the foregoing test of significance is highly influenced by the number of Ss. Had a larger n been used, the difference in correlations might have been significant.

Discussion

The major conclusions from this study may be summarized as follows:

(1) It is possible to substantially reduce the verbal content of a computer-assisted instructional program without significantly decreasing the learning which results from a student having studied the program.

(2) The conditions in which instruction is presented by a low verbal content program required significantly less time than instruction by a high verbal content program. This effect results from the slow type-out rate of the typewriter terminal device and the additional time required by students to read and comprehend the longer typed messages.

(3) Although the difference between the correlations of learning and verbal intelligence was not significant, the results indicated a higher correlation between intelligence and learning on the part of the students who studied the low verbal content program.

Reducing the verbal content of a computer-assisted instruction program has definite advantages for efficiently utilizing instructional time. The time saving can be considerable when a typewriter interface is used. There are also advantages for using low verbal content programs with the newer cathode ray tube interfaces, since these devices cannot accommodate lengthy messages.

The use of low verbal content materials may also be advantageous for the slow learner. Further studies should be conducted with a wide variety of programed materials and with larger groups of subjects to ascertain whether or not the lower correlations of learning and verbal intelligence on the part of the students studying low verbal content programs can be replicated.

The widespread use of programed materials is advantageous to students of all ability levels, but particularly to low ability and less literate students. For many low ability students, studying materials that are programed may mean the difference between comprehending the material and being confused. It is recommended that there should be a greater effort to program materials with as low a verbal content as is possible in order that low ability students can more adequately comprehend the programs.

References

Chall, J. S. Readability: An appraisal of research and applications. Columbus, Bureau of Educational Monographs, Ohio State University, 1958.

Dale, Edgar and Chall, J. S. A formula for predicting readability. Educational Research Bulletin, 1948, 17.

Dale, Edgar and Tyler, R. W. A study of factors influencing the difficulty of reading materials for adults of limited reading ability. Library Quarterly, 1934, 384-412.

Evans, James L. Programing in mathematics and logic, In R. Glaser (Ed.), Teaching machines and programed learning II. Washington, D. C., National Education Association, 1965.

Flesch, Rudolph. A new readability yardstick. Journal of Applied Psychology, 1948, 221-233.

Holland, J. G. A quantitative measure for programmed instruction. American Educational Research Journal, 1967 [illegible], 102.

Logan, T. H. and Wodtke, K. W. The effect of rote rule learning on transfer of learning. University Park: Computer Assisted Instruction Laboratory, The Pennsylvania State University, June, 1966.

Lorge, Irving. Predicting reading difficulty of selections for children. Elementary English Review, 1939, 231-232.

West, J. E., and others. Statistical methods in educational and psychological research, New York: Appleton Century Croft.

Wodtke, K. H. Educational requirements for a student-subject matter interface. University Park: Computer Assisted Instruction Laboratory, The Pennsylvania State University, March, 1967.