DOCUMENT RESUME

ED 021 445 | 24 | EM 000 248

By- Starkweather, John A.
COMPUTER SCIENCE INSTRUCTION IN ELEMENTARY GRADES, AN EXPLORATION OF COMPUTER-BASED LEARNING METHODS. FINAL REPORT.
California Univ., San Francisco.
Spons Agency- Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No- BR-5-0652
Pub Date 1 Jan 68
Contract- OEC-6-10-131
Note- 199p.
EDRS Price MF-$1.00 HC-$8.04
Descriptors- AMERICAN HISTORY, *COMPUTER ASSISTED INSTRUCTION, *COMPUTER ORIENTED PROGRAMS, ELEMENTARY SCHOOL STUDENTS, EXPERIMENTAL TEACHING, FEEDBACK, GRADE 7, INDIVIDUAL INSTRUCTION, *INTELLIGENCE QUOTIENT, JUNIOR HIGH SCHOOL STUDENTS, PEER TEACHING, *PROBLEM SOLVING, *PROGRAMED INSTRUCTION, PROGRAMED MATERIALS, PROGRAMED TEXTS, PROGRAMING, SMALL GROUP INSTRUCTION, TEACHING METHODS
Identifiers- COMPUTEST

During the exploratory phase of this two-year project, 234 instructional computer programs were written by 167 junior and senior high school students, instructed as individuals, in small groups, and in whole classes. Then a doctoral study investigated the effectiveness of computer-assisted instruction in the development of problem solving skills. The study compared three conditions of learning from booklets, three conditions of computer training, and one untutored group in each of six eighth grade classes, all crossed on sex and two IQ levels (above and below the class average). Data analysis showed that students below the class IQ average of 113 who used a combination of two types of computer training materials out performed every other group in the three main problem solving functions. A study of paired learners at different IQ levels used the same design as the doctoral study, but no significant results emerged. Negative results were also obtained from another small study which investigated ability changes in seventh grade students following experience in computer programing. Finally, two versions of the language COMPUTEST for the IBM 1620 and a conversational language for use on the IBM 360 remote terminals were developed. The dissertation provides the bulk of this document. (BB)

FINAL REPORT

Project No. H-226
Contract No. OE-6-10-131

COMPUTER SCIENCE INSTRUCTION IN ELEMENTARY GRADES

An Exploration of Computer-Based Learning Methods.

John A. Starkweather, Ph.D.

University of California

San Francisco

California

January 1, 1968

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

Office of Education
Bureau of Research

CONTENTS

SUMMARY

Use of a small computer in an elementary school setting was explored as a means of practice in and testing of problem solving skills. Pupils from the first to eighth grade acted as authors as well as students in the development of instructional computer programs. A variety of methods were used to introduce pupils to conversational use of the computer. In the course of this activity an improved programming language, COMPUTEST, was developed, usable after only brief instruction by elementary pupils and their teachers, but providing mechanisms for complex language recognition and logical branching when required.

Orientation of pupils and teachers during the first year of the project resulted in production of over 200 brief programs which dealt with many different content areas. Approximately 600 pupils and 30 teachers were involved. A formal experiment demonstrated greater effectiveness of computer instruction when compared to programmed instruction through booklets. Less extensive studies investigated students' ability to work in pairs in response to computer instruction, and attempted to evaluate the effect of programming experience on standard achievement tests.

A problem solving approach to programming activity gave best results in the development of superior programs. Constant feedback of results to the student author allowed rapid program improvement to overcome unanticipated difficulties. Best results in author instruction were obtained by providing explicit instruction and examples of COMPUTEST techniques.

A further extension of the COMPUTEST approach to conversational programming resulted in PILOT (Programmed Inquiry, Learning Or Teaching), a programming language designed to operate from remote terminals connected by telephone to a large computer.

INTRODUCTION

This project explored the use of a small computer with elementary school pupils as a means of practice and testing of problem solving skills. Pupils were prompted to act as authors of instructional computer programs and the computer was not used for drill in the learning of specific information. Emphasis was placed on the recognition and solution by both teachers and pupils of problems which could be described by a sequence of instructions. Methods of feedback from pupil to program author were developed in order to capitalize on student initiative and improve programs during their early use.

A major hypothesis of the project was that skill in problem oriented questioning can be developed by practice in conversational computer programming and by the testing of such programs. A second hypothesis was that youngsters can produce original solutions through the use of the computer to problems which require the development of sequences of instructions. We intended to prompt students to write questions which anticipate multiple response possibilities and to program the appropriate handling of responses to their questions. This implied the development of a language for such programming, usable by elementary pupils and their teachers. A rudimentary form of such a language was available prior to this study but its characteristics for use by children had not been explored and it was recognized that many improvements would have to be made to meet operational needs.

METHODS

1. <u>Exploration of conversational programming by elementary pupils.</u>

The first year of the project was exploratory on the part of computer staff, school district teachers, and their pupils. Approximently 600 elementary students had some contact with the computer operation. This ranged from brief orientation to programming instruction, personal operation of their own program on the computer, or acting as subjects for the operation of their classmates' programs. After a classroom became available in an intermediate school (grades 7-8) students from this school had a considerable advantage in access to the computer so that about half of the total number of students were from these grades. Lower grades were involved, however, and programs capable of successful operation were written by pupils in the first and second grades. About 30 teachers of the Dixie School District were involved at one time or another in bringing pupils for program writing and testing. Programs written by pupils during this first year were collected, organized and reviewed during the second year in an attempt to improve upon the exposure of youngsters to the programming process. Dr. George Stone joined the project during this later stage and has prepared a portion of this report as an independent reviewer of this exploratory programming by pupils. It will be found in the section labeled Findings and Analysis under the title, "Description and Evaluation of Childrens' Computest Programs."

2. <u>The use of computer-assisted learning methods for the training of problem solving skills.</u>

Work during the second year centered on the development and use of especially developed supplementary curriculum materials in 8th grade American History, prepared both with the use of computer methods and with other methods designed to offer certain comparisons of method. An experiment was designed to compare the effectiveness of computer presentation (with logical branching and a display of specific material depending upon a subject's responses) with passive and linear machine display and also with similar material presented by means of booklets. The study also included examination of training methods for general cognitive thinking skills, using materials developed in a related but separate project. This study formed the basis of a doctoral dissertation for Louis W. Stokes and this dissertation is included as a major part of this report. It is entitled "Training For Problem Solving Skills Utilizing a Computer-Assisted Instructional Method." Most of the formal and experimental results of the project are included within this dissertation.

3. <u>Evaluation of student ability to work in pairs.</u>

A brief study with two classes of eighth grade pupils was carried out in the same context as the experimental work described above in order to explore possible advantages of students with differing IQ levels working together in response to computer presented material.

4. <u>The effect of programming experience on standard achievement tests.</u>

A brief summer school session was used to work intensively with a few seventh grade pupils, instructing them in the rudiments of computer programming with the Computest language. Their abilities both before and after this experience were compared with comparable students in a more usual summer school program.

5. <u>The development of an improved programming language for conversational use.</u>

During the course of the project two versions of COMPUTEST for the IBM 1620 computer were produced and work was begun on a related but improved language for operation into a larger computer (IBM 360) from remote terminals.

FINDINGS AND ANALYSIS

1. Description and Evaluation of Children's Computest Programs.
G. Stone

During the term of the contract, 234 COMPUTEST programs were written by 167 elementary school children. A few children wrote as many as six or seven programs; most children wrote only one or two. During the first year of the project, children came into the Computer Center on the basis of their own interest. Instruction was casual beyond an introduction to basic COMPUTEST mechanisms. It consisted mostly of allowing children to take and to look at other simple programs written by children or by staff members. Promising students were given special attention. A summer session was conducted in 1966 in which it was attempted to integrate experience with the COMPUTEST system into a full-size class with concurrent presentation of other subject matter.

During the spring of 1967, the programs written up to that time were evaluated and a variety of shortcomings and characteristic errors were identified. It was concluded that the average level of the program products could be substantially improved by a more focussed approach to the programming task, allotting more time, more contact with the computer, and providing formal instruction in programming both in class and by COMPUTEST. In this section, we describe and illustrate characteristic features of children's COMPUTEST programs under the several circumstances of learning.

Evaluation of children's COMPUTEST programs--or of any such programs -- may be broadly subdivided first on the basis of whether the techniques being appraised are those of programmed instruction or test construction in general, on the one hand, or those specific to COMPUTEST on the other. We shall place less emphasis on the more general techniques. To attempt to cover them in detail would take us far afield into the theories of instruction and testing, where this project had limited objectives.

Evaluation of programs from an instructional viewpoint.

Three common errors were found. A lack of structural interrelationships among question items was the rule. The modal program was a series of unrelated questions about a common topic with a testing or quizzing rather than an instructional emphasis. (programs 174, 186, 220). Variation about this mode ranged from collections of unrelated questions (program 258); to a well organized, logically structured game (program 109); or exercise (program 201).

In many cases, the questions asked of students were impossibly specific and detailed. In some cases, students called upon their own expertise, failing to make proper allowance for their deviation from the knowledge level of their classmates (program 186). In other cases, tests of specific details were built without an attempt at related instruction (program 220). Sometimes the children's programs failed to inform the student of the correct answer after an error (program 144).

The third general deviation from usual instructional norms was in the excessive use of "cute" and downright insulting comments as feedback. The opportunity to make these remarks seemed to be the focus of involvement in the programming effort for many children. (programs 119, 144, 186).

We cannot say to what extent these "errors," from the point of view of computer assisted instruction or of other systematic interactive use of the computer, could be overcome by some instruction or examples. Students in the second summer session, whose first exposure to COMPUTEST included complex didactic programs, submitted a higher proportion of interesting programs but did not eliminate these tendencies. It is perhaps not surprising that five weeks of instruction in which general instructional principles were not taken up failed to eliminate them.

The ubiquity of the "quizzing" mode may be influenced by teaching practices in the schools and by the undoubted ease of constructing and responding to simple quizzes. In fact, this ease may influence teacher practices in classroom situations and their influence on children can be clearly seen in this regard.

Evaluation of programs from programming viewpoint.

Our appraisal of the children's use of specific COMPUTEST techniques is based on the degree to which they utilized the many potentialities of the system. These potentialities may be subdivided into four major areas: formulation of questions, answer recognition techniques, replies and program logic. While these areas are not entirely independent (for example, really sophisticated answer recognition techniques require sophisticated logic), at the level of competence where most of the children were operating there was no necessary relationship. We may conveniently consider these topics separately. Although these areas have been mentioned in the order in which they occur in a student's view of a question, it is necessary to begin with the discussion of answer recognition techniques, since these underlie the principles of question formulation.

The highest use of an interactive system like COMPUTEST lies in the categorization of minimally constrained responses. Thus we can evaluate question formulation and answer recognition techniques in terms of the degree of constraint imposed. Constraint may be imposed explicitly in the statement of the questions of implicitly in the kinds of answer recognition employed.

Answer recognition techniques.

Maximum constraint is found in true-false or multiple-choice question formats. Relatively few of the children relied entirely on this device. (The percentage of the total set of programs that used each of the answer recognition devices discussed is given in Table 1.) Program 119 exemplifies the best use of the multiple choice format. From the point of view of subject matter, this was a good program. There seemed almost to be a negative correlation between quality of content and complexity of the program, as though the children chose some aspect of the overall task to emphasize at the expense of others. Some of the shortest programs utilized a maximum of programming devices as children were prompted to try the operation of these techniques. (program 144)

Table 1

Percentages of 234 programs using various answer recognition techniques. Since most programs used several techniques, the percentages fo not sum to 100%.

Explicit Answer Constraints	Percentage of programs using the technique
True-false	3
Multiple choice	16
Cued response	11
Answer Recognition techniques	
Match 1	80
Group n	50
Multiple match	32
Alternative forms of response (includes COMPARE option)	17
Alternatives involving RESCAN	6
Alternative words or concepts	15
Single concept extracted from text	17
Related concepts extracted from text	14
Alternative accepted in sentential structure	1

Only slightly less constrained than the multiple-choice questions is the one that requires a single word reply. This type of question was by far the most frequently found in the children's programs. The COMPUTEST option normally used for this kind of answer recognition is "Match 1," and it is suggestive of the power of the device that is the normal or preset answer evaluation mode of the system. In using Match 1 in the simplest way, a programmer asks a question that strongly invites a one word answer (programs 144,Q6; 186, Q4). An extension of this approach uses "Group n" to demand a string of words. In fact, this capability was involved in a frequent error of over-specificity in the response designation. The most extreme example of this may be seen in Program 174 Q10 where the child specified a 12 word sentence as the correct answer. It is more appropiately used in program 220, Q4, where a match of 9 out of 9 elements was required to get a "correct" reply.

In its normal operation, "Match 1" is able to extract the correct one-word answer from text. When the question is worded in such a way as to elicit more than a one-word response, the possibility arises that there will be variability in the way the key word is written. For example, a number might be written as a numeral or spelled out. Or, the key word might vary in number, case, or tense. These alternatives are most simply handled by two devices, the multiple R-list (program 220, Q3), and use of the compare option (program 227, Q2).

The next level of relaxation of constraint in answer recognition techniques accepts the possibility that alternative words may be employed to express a particular idea. Multiple R-lists are required to cope with this possibility (program 144,Q3, Q4) and in many cases it may be impossible to arrange to detect all of the possibilities within a single format. Then two sets of R-lists and A cards must be used, along with the "Rescan" option. This useful technique was rarely adopted by the children.

The final relaxation of constraint comes when the answer is evaluated by the program for the coordinated presence of several concepts. This approach is not logically related to the technique mentioned earlier of seeking multiple, parallel concepts in a single answer. Instead, related concepts and, often, relation terms are included in the R-list with the demand that two or more matches be made. This level of sophistication was reached by very few of our young programmers. Program 135 used this approach throughout, although not too well. A single question that demonstrated considerable analytic power is in program 144, Q2. It is, of course, possible to apply the techniques for recognition of alternative answers, described above, to one or more of the elements in a relational sentence and to apply the ligic of alternative or multiple answers to sentential answer lists. Such complexity taxes the capabilities of mature and experienced programmers. Most often, the children's effort to evaluate sentences or extended phrases was limited to the unorganized inclusion of multiple concept words with the use of "Match n." (program 135).

Question Formulation.

An alternative to the use of rich and complex answer recognition techniques is to formulate questions in such a way as to markedly reduced the range of answers likely to be given by the subjects. We have already mentioned the use of multiple-choice or true-false format, in which the student is explicitly instructed as to the set of acceptable answer alternatives. Another way of imposing constraint is through an explicit statement of format requirements (program 220). Such "format cues" were used by a very few of the children.

A more subtle calss of cues provides some information about the domain in which the correct response is to be found by inclusion of related words in the statement of the questions. Consider these three questions:

Whow did the American slaves gain their freedom?
Who freed the American slaves?
What president freed the American slaves?

It is very clear that the range of possible answers decreases greatly from the first to the third of these questions.

Evaluation of "subject matter" cues is difficult to make explicit and quantitative. It was said that most of the childrens' questions strongly invited one word answers. For the most part, this invitation is given by means of subject matter cues. Therefore, we may assume that the children were using subject matter cues effectively. A few programs were notable exceptions to this rule. Program 174, Qs 3, 4, 10 are examples. In these cases, reading the question alone does not give much indication of the kind of response desired.

Programmed replies to sutdent responses.

COMPUTEST provides two commands that are primarily used to produce a textual output from the computer's typewriter if the student's response matches (G) or fails to match (B) the right answer list. Logically, there are five classes of replies that can be made:

1. No reply. The program continues without further reference to the last question.
2. Undifferentiated reply. The same comment relative to the preceding question is made whether the student's response was "correct" or "incorrect."
3. Feedback as to the correctness of the student's response.
4. An evaluative comment upon the student's response.
5. Further information about the topic of the question. In the case of "B" response, this information usually takes the

form of the correct answer.

Table 2 presents the percentages with which each of these types of reply was used.

Table 2

Percentages of 234 programs using the several classes of program replies. Since most programs used several types of reply, the percentages do not sum to 100%.

<table>
<tr><th>Reply</th><th>"G" replies</th><th>"B" replies</th></tr>
<tr><td>1. None</td><td colspan="2">1%</td></tr>
<tr><td>2. Undifferentiated</td><td colspan="2">1%</td></tr>
<tr><td>3. Feedback only</td><td>33%</td><td>14%</td></tr>
<tr><td>4. Evaluative
(includes feedback)</td><td>82%</td><td>44%</td></tr>
<tr><td>5. Informative</td><td>4%</td><td>78%</td></tr>
</table>

The "No reply" class is mainly called upon when the programmer uses a student response either to learn about student characteristics prior to some branching, or when a response is used to permit delays in presentation of material--for example, when a question is presented after a slide had been viewed. It was this second purpose that led two of our programmers to use the "No reply" category. (program 303).

The two children who made undifferentiated replies gave them after their students answered questions about themselves. The replies were simply acknowledgements. (It is interesting that neither of these children used, in their programs, the information thus acquired.)

The significance of reply classes 3, 4, 5 differs between the "G" replies, "Correct" was almost always used as a variant of class 4. In other words, after saying "Good" or "Great" a few times, a "Right" was apparently considered by the programmer to convey the same message. Only six children used class 3 replies alone. Of these,three used them for both "G" and "B" replies, while three gave corrective information with their "B" replies.

"Wrong" was mixed with other evaluative comments to some extent, but much less frequently than "Right." "Sorry" and "Too bad" were much more common. In Table 2, the last two replies were counted with the evaluative replies.

Additional information was a rate exception in "G" replies, and lent programs where it was used a quality of graciousness (program 71). In contrast, only 22% of the programs failed to privide any corrective information; and such programs seemed surly and unconstructive to us (program 258).

The most striking thing about the childrens' replies was their variety, informality, and sometimes, wit. Some also seemed vulgar to adult readers. Programs 119, 135, 144, 174, 186, 220, 258, 302, 303 all illustrate these characteristics. It seemed to us that for many of the children the primary activity was the anticipation rarely the realization because of time pressures) of astounding their friends with the daring and wit of the replies. A child choosing this emphasis need not be concerned with the quality of the programming itself, since every kind of item gives opportunity for "G" and "B" replies.

It is only when the programmer becomes involved in the effort to astonish has friends with the cleverness of his answer recognition techniques and his anticipation of their unconstrained responses that sophistication in programming begins to emerge. Such involvement is very unlikely until children (or adults for that matter) have had the opportunity to observe students at work on programs they have written. Thus, the emphasis on the replies provides motivation for the first program, with later programs or later versions of the first program providing the occasion for involvement in programming techniques. More will be said of this when the three approaches to teaching the use of COMPUTEST are discussed.

Program logic.

It was in the area of program logic that our expectations of the outcomes of exposing children to COMPUTEST suffered their most grievous disappointments. With a few notable exceptions, even the simplest deviations from the pure "quiz" mode were not used. Only 15 programs used the RESCAN option to evaluate answers against more than a single R list. Only 4 made program decisions on the basis of scores tallied during earlier portions of the program. Forty-eight examples of branching (exclusive of RESCAN) were found, occurring in 37 different programs. Table 3 shows the distribution of these uses of branching into several classes.

Table 3

Number of programs using various program logic devices.

Use of branching	Number of Programs	
Early termination option	12	
Other student option	3	
Single question branches	20	
Special comment for specific errors		9
Present simpler question on errors		8
Present harder question on correct response		3
Complex logic in games, problem solving, etc.	8	
Complex program loops in quizzing	6	

As was noted earlier, the programs using the more complex logic were not necessarily the most interesting from other points of view. In the games and exercises it was common to find a relatively simple logic module repeated until the overall size of the program was quite large.

Complex progrm logic was used in only four of the six programs tabulated as having complex loops in quizzing. The other two involved only a rather trivial repetitive looping in very short programs. Program #302 represents the best and most original use of branching in an instructional or testing type of program. Intelligence, wit, and creativity are apparent throughout the program. The girl who wrote it had just completed seventh grade.

Effectiveness of three teaching methods.

We have explored three methods of making COMPUTEST available to the children:

1. Permitting children to use the computer on the basis of their own interest, self-scheduled and with casual instruction. (This group is referred to as "individual porgrammers.")

2. Integration in a full class setting (Summer Session I).
3. Focussed experience and instruction in small groups (Summer Session II).

No controlled comparison of the three methods is possible, since there was no matching of the kinds of students that took part in the different approaches and no equilization of the amount of computer time available in the three circumstances. Nevertheless, the results seem to permit clear conclusions to be drawn. The greatest productivity came from a very small number of children who developed an interest in the computer that could only be described as passionate. One of these boys, in particular, was on hand during almost every free hour when the Computer Center was open. His accomplishments included:

A test on the clarinet.
A test on famous persons.
A test on mapping skills, using a map in conjunction with his COMPUTEST program.
A test in Spanish.
A complex program to cast horoscopes.
An intricate effort to simulate a slot machine (prompted by a non-COMPUTEST, 1620 demonstration program that does so.)

Eventually he graduated to Fortran programming. The individual programmers because of a requirement to write ten questions, wrote longer programs on the average than the children in formal classes. Lengths of programs can be best described in terms of the number of A cards used. These numbers ranged from a low of one (a clever program with a program loop) to a high of over 100 (a less interesting program that identified the number the subject is thinking of) The overall median length was 11 A cards. The median in the individual group was 12; for the first summer session, 6; and for the second summer session, 9.

In spite of the level of involvement shown by a few of the individual groups, our overall success at teaching the use of COMPUTEST's capabilities was highes in Summer Session II, in which focussed teaching of programming devices was used. Some objective basis for comparing the three groups of children is provided by displaying the percentages of each group that used each of a number of identifiable program devices. These are shown in Table 4. It can be seen that a farther advance in the use of program logic was the main thing that differentiated our second summer session from the earlier groups. The simpler answer recognition techniques were adopted by sizable number of children in all groups. Methods of program logic were fairly effectively conveyed in the second summer session, while the first summer session managed to instruct a fair number of students in the rudimentary use of answer recognition in sentences.

Table 4

Percentage of each programming group that used various program devices.

	Individual	Summer 1	Summer 2
Explicit Answer Constraints	%	%	%
True-false	3	0	0
Multiple choice	14	7	0
Cued response	12	2	21
Answer Recognition Techniques			
Match 1	85	65	63
Group n	49	51	53
Multiple Match	33	30	32
Alternative forms of response (Including Comp)	3	56	53
Alternatives involving rescan	7	2	11
Alternative words or concepts	14	19	21
Single concept extracted from text	12	33	26
Related concepts extracted from text	13	21	5
Alternatives accepted in sentenial structure	1	2	0
Program logic			
Early termination and student options	5	7	21
Single question branching	6	2	47
Complex logic in games, etc.	4	0	0
Complex program loops in quizzing	0	0	37

Conclusions

The children who took part in these studies were able to master the elements of COMPUTEST programming to varying degrees. In large classes with minimum access to the computer, the program products were mostly uninteresting. A few children who, by their own initiative, had extended experience with the COMPUTEST system, produced programs that were original and interesting from the point of view of program logic, but they made relatively little use of the instructional and interactive capabilities of the system. Best results were achieved by providing explicit instruction and examples in COMPUTEST techniques.

2. Training for Problem Solving Skills

This section briefly describes the experimental investigation of the effectiveness and efficiency of specially developed computer-assisted learning materials for the training of productive problem-solving skills and attitudes. This was the major formal study completed by this research effort, and it is more extensively reported in a separate section entitled "Report of the Major Study: Training for Problem Solving Skills Utilizing a Computer-Assisted Instructional Method." That section contains the doctoral dissertation of Louis Stokes with the exception of extensive appendices to the dissertation. Readers interested in the materials of the dessertation appendices may obtain copies from University Microfilms, Ann Arbor, Michigan.

A basic assumption of the experimental study was that computer-based materials make possible a more interactive training situation, involving open-ended questioning and feedback, and hence have the potentiality of eventually leading to a truly individualized instructional setting for the training of these thinking skills. The basic pedagogical rationale concerning the structure and content of the training materials, and the underlying theoretical assumptions concerning the problem-solving skills to be trained for, were derived from the work of Crutchfield and Covington (1965). This formed the basis for the method of investigating the question of the effectiveness of the CAI training materials developed. The question of the efficiency of the CAI materials developed was mainly one of determining the feasibility of utilizing a special computer language, COMPUTEST. This question of efficiency centered around two main points of user-simplicity of the COMPUTEST language and its potentiality for permitting full utilization of the computer system.

The experimental design consisted of seven different training conditions, completely crossed over the two categories of sex and two levels of IQ (above and below the class mean of 113). This design was completely replicated within six randomly chosen eighth grade classes with a final total of 168 subjects. Of the seven groups, three groups constituted the major computer-training conditions. One computer group worked on selected lessons from the <u>General Problem Solving</u> series (Covington, Crutchfield, and Davies, 1966) in a computer-simulated teletype situation. A second computer group worked on specially developed computer programs consisting of "fictitious" American history materials which were presented on an IBM 1620 computer. A third computer group received the computer-training materials of both the preceding two groups. The other four groups consisted of three "active control" comparison booklet groups and an untutored group of subjects. The training period lasted for approximately three weeks and consisted of eight to ten 40-minute periods of instruction for most subjects. Two periods specially developed posttests yielding measures of problem-solving performance and measures of various relevant attitudes were given at the completion of the training period.

Statistical analyses were carried out by means of the planned comparisons method, with separate analyses being performed on so-called Middle IQ subjects (mean=104) and so-called High IQ subjects (mean=124). The most consistent and statistically significant performance finding was that the Middle IQ subjects who worked on both types of computer -training materials outperformed every other experimental group of both the Middle and the High IQ subjects in the three main problem-solving functions of Problem-Formulation, Idea-Generation, and Idea-Evaluation. These subjects also rated the posttest problems as positive to a significantly greater degree. This superior performance was primarily attributed to the beneficial interaction effects of the two different types of computer-training materials; a possible alternative explanation, namely, the lengthened time of involvement in the training period, was not sufficient to account for the results obtained.

The three computer groups, especially the computer history subjects, all positively rated their training materials significantly higher than did the other comparison booklet groups. This was true for both the Middle and High IQ subjects. The basis of this favorable rating was the problem-solving nature of the training materials. This was interpreted as resulting from the intrinsically attractive interactive nature of the computer-training materials; an alternative explanation in terms of a "Hawthorne" effect of the novelty of the computer was not substantiated by the facts of the study. Positive changes in problem-solving attitudes having to do with approaching a problem through more than one idea, persistence in working on difficult problems, and higher self-evaluation of oneself as a problem solver were almost all statistically significant in favor of the computer-instructed subjects for both the Middle and High IQ subjects.

The second major question of the study on the efficiency and effectiveness of using the COMPUTEST (and the analogous PILOT) computer language for the development of these CAI productive thinking materials was considered to be positively answered on two main grounds: First, the ease of development of the history computer programs by the author who had no previous programming experience; second, the fact that these computer programs were sophisticated enough to establish a valid instructional situation, as evidenced by the performance and attitude gains on the posttest for certain of the computer-instructed groups.

3. Evaluation of Students' Ability to Work in Pairs.

A small exploratory study concerning the effects of a paired group training situation was performed which involved two eighth grade classes with a total of 56 subjects. The basic design of this study was to pair a subject (High IQ) above the class mean IQ of 113 with another subject of the same sex who was below the class mean IQ (Middle IQ). It had been suggested the Middle IQ subject would benefit from the example and help of the brighter student, while the more intelligent student would also benefit from his active involvement in explaining the problem to his partner.

The same training materials and posttests, the same experimental design of seven experimental groups, and the same experimental format and procedures were utilized in this study as in the major experiment on problem solving skills with the exception that subjects worked in pairs, as described above, rather than working separtely on the training materials. All subjects did, however, take the posttests separately without any assistance from their training partners.

The overall results, which were obtained from the same statistical design of the planned comparisons method, did not indicate any increased benefit from this type of training situation over the previous format of the subjects working separately on the training materials, either for the Middle IQ or the High IQ subjects. There were no consistent and statistically significant differences in the three performance measures of Problem-Formulation, Idea-Generation, or Idea-Evaluation among any of the combinations of training groups. The most important consistent finding in the attitude measures was found in the Evaluation of the Training Materials for the Middle IQ subjects. All five of the measures showed differences in favor of the computer trained subjects versus the comparison booklet subjects in the positive evaluation of the special instructional materials, with four of these five measures being significant at the .05 level. The computer subjects in the Middle IQ group also tended to score higher on the attitude measures of Self-Evaluation as a problem solver with all of the measures being in favor of the computer trained subjects, two of them significant at the .05 level.

When these results are compared with the results of the main experiment it is found that there is a general lack of performance and attitude gains by the computer subjects in this exploratory paired group training situation. Most obvious of all is the lack of performance superiority by the Middle IQ subjects who received both types of the computer training materials, for in the main study these subjects outperformed every other group in the experiment. Also lacking is the significantly higher positive rating and evaluation of the special conputer training materials by the High IQ computer subjects and their significantly higher positive evaluation of themselves as problem solvers in the Self-Evaluation inventory. Both of these were consistently and significantly higher in favor of the High IQ computer subjects in the main thesis experiment. This overall general poorer showing by the computer subjects calls for an explanation of why the pairing of two students to work together on the instructional materials has led to a poorer performance and attitude showing rather than a superior one as had been initally expected.

The key to this problem apparently lies in the fact of pairing together a High IQ student with a Middle IQ student. It seems that the difference in reading speeds between these two IQ groups is a major factor in explaining the results obtained. An examination of the subject protocols indicates that a number of the Middle IQ subjects complained about the manner in which their High IQ partners worked on the training materials. It appears that the High IQ subjects, who would usually finish reading the materials first, would tend to fool around while they were waiting for the slower student to complete the reading rather than help him to get done quicker. Thus rather than aiding the Middle IQ subject, the High IQ subject actually hindered him from learning the instructional materials.

It is interesting to note that both the Middle and High IQ computer subjects reported that their partners were distracting to a greater degree than did the comparison booklet subjects. This may have resulted from the sequential nature of the computer training materials in which each question had to be answered first before both subjects could continue. Thus the brighter student would be more anxious to continue and would even feel that the slower student partner was frivolous by not finishing as quickly or not putting down as intelligent an answer as he should. However, in the booklet reading materials the faster reader could read ahead in the story without having to wait for his slower partner to finish each page; this practice would avoid the problem of the differences in reading speeds which could not be done by the computer instructed subjects. In fact, this practice of the brighter students reading ahead in the booklet materials was observed by the author in training situation.

It appears then that the significant factor affecting the results of this paired group training situation is not the level of intelligence, but rather the reading speed and ability of the students. Such reading ability tends to be closely correlated with IQ. It appears in future research regarding subjects working in small groups on the computer-assisted instructional materials that the reading speed of the various members of the groups should be considered an important experimental design factor.

4. <u>The Effect of Programming Experience on Standard Achievement Tests.</u>

During a five-week summer session in July, 1967, two groups of six pupils each were allowed to volunteer for daily instruction of two hours each day in COMPUTEST programming. These were pupils about to enter the seventh grade and were from a geographically remote part of the school district which had not had prior access to the computer. The Reading and Arithmetic sections of the California Achievement Test were used with these pupils at the beginning and at the end of the five week period, using an alternate test form for the second testing

Another group of 12 pupils of the same grade were chosen to be as comparable as possible in terms of general intelligence scores, pupils who attended an experimental "demonstration school" during the same five week period. These pupils, who acted as a control group, took the same tests at the beginning and end of the five weeks.

Table A shows the scores of experimental and control groups in terms of percent correct for the first and second testing with the two tests. Table B presents the same information in terms of change scores for individuals.

Table A

Percent Correct Scores

	California Reading Test	
	First Testing	Second Testing
Exper.	93.0	95.6
Control	91.4	92.1
	California Arithmetic Test	
	First	Second
Exper.	90.0	91.7
Control	86.3	84.7

Table B

Distribution of Change Scores:
First to Second Testing

California Reading Test

						X	0		X			0	X			
			0	0	X	X	0	X	0X		0X	0X	X		0	X
-8	-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7	+8

X: Experimental Group (Mean Change = 1.82)
0: Control Group (Mean Change = 0.89)

California Arithmetic Test

							0									
							0X	X	X	0X	X					
0					X	0	0X	0X	X	0X	X				0	
-8	-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7	+8

X: Experimental Group (Mean Change = 0.64)
0: Control Group (Mean Change = -0.22)

Although both tests show a mean advantage for the experimental group who received computer programming instruction, it is clear that individual overlap is considerable, and statistical significance cannot be demonstrated with these small numbers of subjects.

The tests were chosen with the hope that they would partially represent elements of ability in problem solving, data interpretation, or questioning. Generally standardized tests are indirect in their assesment of such abilities. Our experiment was further weakened by the use of tests at too low a grade level (6) with students of very high test ability. Tests at a high school level would have had an increased chance to measure differences with these selected pupils

5. <u>The Development of an Improved Programming Language for Conversational Use.</u>

A. COMPUTEST

There are currently three forms of the COMPUTEST system. These are (in order of development):

1. COMPUTEST I. This is a primitive version which operates on an IBM 1620 computer having 20,000 digit memory and card/input/output, but no auxiliary storage. The computer may be either a model 1 or a model 2 but must have the automatic divide and indirect addression special features.

2. COMPUTEST II (20K). This is a greatly improved version which makes use of one or more 1311 disk drives attached to a 1620 computer with a 20,000 digit memory. This version has improved features such as:

   1. An author mode..aids the author in debugging his program.
   2. A desk calculator mode...(Expensive Desk Calculator 1620.11.0.043)
   3. Subroutines (coded in COMPUTEST)
   4. Forward or backward program branching by use of statement labels.
   5. Built in linkage to user coded (1620-SPS) routines.
   6. Restart (subject may stop at any point and continue another day).
   7. Additional "options".
   8. Increased COMPUTEST error checking.
   9. Additional "L" storage areas.
   10. Increased typewriter input area (allows 500 characters instead of 200).

3. COMPUTEST II (40K). This version is functionally the same as the 20k version of COMPUTEST II but modified more rapidly when a 40,000 digit memory is available.

The program specifications and test data for the COMPUTEST I version may be obtained from the IBM Program Information Department, 40 Saw Mill River Road, Hawthorne, New York, Bile No. 1620. 2.0.052.

We have chosen not to submit COMPUTEST II to the Program Distribution Library, since initial generation of the system is complex and might be troublesome unless demonstration is provided. If you are interested in COMPUTEST II, please contact:

Dr. John A Starkweather
Computer Center

University of California Medical Center
San Francisco, California 94122
Telephone (415) 666-2012

Please include information about the configuration of the 1620 that you wish to use. Minor adaptations to your specific machine may make COMPUTEST more efficient.

B. PILOT

PILOT is a system designed to permit natural communication with computers by providing the facilities for a conversational dialogue to take place between a person and a computer. It is designed for use with dial-in terminals making telephone connections with a computer having time-sharing capability. Machine assisted learning (computer aided instruction), specialized inquiry systems, simulated diagnostic interviews, and similar endeavors, are only a few of the ways in which PILOT (Programmed Inquiry, Learning Or Teaching) might be used.

This system allows the specialist in a non-computer field to write interactive conversational programs, without having a knowledge of the technicalities of the machine. Likewise, the subject needs no special knowledge to converse with a PILOT program. He needs only the ability to press appropriate keys on the typewriter.

A typical conversation with a computer using previously written PILOT program might have the following form:

1. Information or a question is presented by the computer.
2. The subject types a reply.
3. Recognition techniques permit the computer to make decisions based on the reply.
4. The computer may make one of several comments, and/or ask another question based on its previous decisions.
5. The conversation continues from (2).

The PILOT language permits the author to describe to the computer, in a natural way, how to make these decisions and what to do about them.

Responses can also be saved and reviewed later. This and many other features of the language are optional so that the author may use only what he needs.

The system is designed so that its use is not restricted to a particular manufacturer's equipment. PILOT is general enough so that it may be used at any level of program complexity on a range of machines from a small computer with a typewriter to a large system with many typewriters and visual displays, using one or many programs simultaneously.

Up to date specifications and other information about PILOT may be obtained from:

Dr. John A. Starkweather
Computer Center
University of California Medical Center
San Francisco, California 94122
Telephone (415) 666-2012

CONCLUSIONS

Best results in prompting children to act as program authors were achieved by explicit instruction and examples of COMPUTEST techniques. Children who had extended experience produced programs which were original and interesting from the point of view of program logic. A major logistic problem resulted from the availability of only one typewriter on the available computer, for attempts to teach programming in large groups with minimum access to the computer resulted in uninteresting productions.

Computer-based materials made possible and interactive training situation, and allowed a truly individualized instructional setting for the training of skills for problem solving. Groups taught by computer interaction learned these skills more effectively than those who worked with similar materials in booklet form. Suggestive PILOT data indicates that problem oriented programming activity can result in an improvement in children's general problem solving ability.

Student initiative and involvement in programming activity proved to be a powerful aid in the production of superior program sequences. With a means for constant feedback of results of program operation to a student author, the author is motivated to overcome unanticipated difficulties by modification of program strategy or of question presentation. Such a problem solving approach to programming can effect great savings in the development of new programmed curricula. It requires a programming system with means for feedback to the author and a system which makes it easy to change the program. Future systems should collect comments easily from users and present them to authors for this purpose.

Considerable further work is required to produce and author language for computer-assisted instruction which meets these needs, as well as allowing access to all the other powerful aids the computer may provide. The PILOT language is being developed in this direction.

PUBLICATIONS

Preparation of the following publications was supported in part by the contract being reported here. Reference to this support has been included in each instance when appropriate.

1. Starkweather, J.A., Computer-assisted learning in medical education. Canad. Med. Assoc. J., 1967, 97, 733-738.

2. Starkweather, J.A. Computer methods for the study of psychiatric interviews. Comprehensive Psychiatry, December, 1967.

3. Starkweather, J.A. Computer simulation of psychiatric interviewing. In N.S. Klein and E. Laska (Eds.) Use of Electronic Devices in Psychiatry. New York: Grune and Stratton, in press.

4. Starkweather, J.A. Educational use of computer assistance. Proceedings of 8th IBM Medical Symposium, 1967

5. Starkweather, J.A. Computer aids to content recognition. In G. Gerbner (Ed.) National Conference on Content Analysis, New York: Wiley, in press.

6. Starkweather, J.A. PILOT: A system for Programmed Inquiry, Learning, Or Teaching. In G. Gerbner, (Ed.) National Conference on Content Analysis, New York: Wiley, in press.

7. A Chatty Computer in Room B-3. Educom, March 1966, 3-7.

Report of the Major Study:

Training for Problem Solving Skills Utilizing a Computer-Assisted Instructional Method

Training for Problem-Solving Skills
Utilizing a Computer-Assisted Instructional Method

By

Louis W. Stokes
(Louis Walter Stojkiewicz)

A.B. (Duquesne University) 1963
M.A. (Hollins College) 1965

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

Dr. Richard S. Crutchfield
....

Dr. Martin V. Covington
....

Dr. John A. Starkweather
....

Committee in Charge

TABLE OF CONTENTS

IX. APPENDIXES

ACKNOWLEDGEMENTS

I am indebted to Dr. Richard S. Crutchfield for his advice during the initial stages of this research and his editing of the final manuscript. Dr. Martin V. Covington's helpful criticisms have led to a more readable final report. Special thanks are due to Dr. John A. Starkweather for his most generous offer of both the computer and clerical facilities of the Dixie Computer Learning Project and the financial support of the author for the research of this thesis. This study would not have been possible without this assistance.

Dr. Gerald A. Mendelsohn offered valuable suggestions concerning the statistical analyses of the data and Mrs. Eleanor Krasnow gave important assistance in writing the computer program to run the planned comparisons analyses.

Thanks are due to the Dixie School District and its superintendent Mr. Dennie B. Willis for their cooperation in permitting this study to be conducted in the Vallecito Junior High School. Mrs. Blanche Wells, Mr. William Turner, and Miss Sandra O'Leary are gratefully thanked for their extensive help during the execution of this study.

Initial research on the development of the computer-training materials was partially supported by a grant from the Carnegie Corporation of New York.

The principal research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy. This work was supported in part by contract USDHEW-OE-6-10-131, U.S. Office of Education, John A. Starkweather, Ph.D., project director.

I. INTRODUCTION

1. THE CONCEPT OF INDIVIDUALIZED INSTRUCTION

The increasing interest in the concept of individualized instruction in recent years has developed out of the realization by psychologists and educators that within the area of mass education the most crucial problem facing them is the development of innovative instructional methods and materials. The need is for instructional methods and materials that will equip students with those cognitive skills, attitudes and motivational dispositions which are necessary, and which will become more essential, to deal with the present complex problems and the yet unknown future problems which will face mankind in his relationships to himself and his environment.

Crutchfield (1965) has succinctly pointed out three critical reasons which necessitate the rapid implementation of individualized instructional methods within the ongoing educational process of today. The first is pedagogical in that behavioral research has come to show that to make the instructional process optimal, account must be taken of the individual's specific background, capabilities, and distinctive cognitive style. The second reason is motivational in that the individual student, within present-day educational systems, seeks and needs a form of instruction which is suited to his needs and wants and is therefore meaningful to him. The third is social in that the individual must be optimally trained in his own unique cognitive skills and attitudes so that he will best be able to bring his unique talents to bear against the complex and unknown problems of tomorrow's world, and this will be best achieved through educational instruction which is suited to his particular talents and needs.

In recent years the increasing interest in the concept of individualized instruction has led to the revitalization of an important pedagogical area of research, that revolving around the teaching of problem-solving skills and attitudes through systematic direct training materials which make use of individualized instructional methods. The recent advances in programmed instruction, and especially the innovative technological method of computer-assisted instruction, have developed directly out of the necessity of finding some type of instructional method which would implement the concept of individualized instruction in the actual school setting. With the advent of computer-assisted instruction, it is now possible to feasibly consider the training of vast numbers of students through large-scale individualized instructional methods.

This thesis is directly concerned with evaluating the unique features of the computer-assisted instructional method through means of developing systematic computer training materials for productive problem-solving skills and attitudes. The rationale and thesis objectives will be explained in more detail after we first make a closer examination of the background of the pedagogical research on the systematic training for problem-solving skills, and the development of computer-assisted instructional methods.

2. THE TRAINING FOR PRODUCTIVE THINKING SKILLS

Until the past decade there have been very few studies concerned with research on problem-solving skills involving the use of training materials, and scarcely any attempts at the systematic direct training of such skills and attitudes. The early paper (1926) and the later monograph, originally written in 1935, <u>On Problem Solving</u> (1945) by Duncker, the book by Wertheimer, <u>Productive Thinking</u> (1945), and the several papers written by Maier (1930, 1933) constitute perhaps the most pertinent early work in this area; but even here, as with other early investigations of the thought processes, there was no systematic attempt to develop and utilize training materials for extended research and actual educational usage.

One possible explanation of the dearth of studies involving the direct training of thinking skills is that the researchers felt they could not proceed to application until they had a far better grasp of the underlying theoretical processes involved. Another explanation might be found in the applied educational milieu of the lockstep classroom system. Since, as Crutchfield (1965) points out, the training for productive thinking skills requires an individualized instructional method so that the person can develop along the paths of his particular cognitive and attitudinal strengths, it may be that the concept of individualized instruction had to be accepted as important and feasible first before any progress could be made in attempting to train for problem-solving skills.

In any case, concurrent with the recent rise of interest in individualized instruction, there has developed a strong interest in the general factors involved in training for problem-solving skills and related attitudes. This can be exemplified in examining the series of Utah Conferences on scientific creativity and productive thinking. In the selected papers of the first three conferences held in 1955, 1957, and 1959 (Taylor and Barron, 1963) there is only one paper (by Parnes) concerned with the direct training of thinking skills; the few other related papers have to do with general environmental conditions and specific situational determinants of creative and problem-solving behavior.

In the later two Utah Conferences which were both published in 1964 (Taylor, 1964a, 1964b) there is a substantial increase in the number of studies reported which are concerned with the fostering of thinking skills. However, many of these are still mainly concerned with the indirect effects of environmental changes on the development of productive thought processes rather than with direct teaching for these skills.

Torrance has emerged as one of the best known of those researchers concerned with the problems of creativity and problem-solving skills in the educational system, as witnessed by his recent books on the subject (1962, 1963, 1965). The major emphasis of his work has been basically upon the development of criterion items to identify creative problem-solving behavior, and upon those situational factors which aid or inhibit creative thinking and problem-solving behavior, such as evaluative teacher behavior, differential sex reinforcement, peer orientation, and cultural factors.

Two of the prominent investigations which have been concerned with the direct training for problem-solving skills are the studies of Parnes (1962, 1964) and of Suchman (1960, 1961). Parnes has utilized the problem-solving methods developed by Osborn (1957) and has found the instructional materials to be significantly beneficial for the trained subjects in different problem-solving tests; Suchman's work on inquiry training has also produced positive results. But the critical shortcoming of these investigations is that they are attempting to train for problem-solving skills by means of general group classroom practices which do not permit either efficient or effective individualized instruction when, in fact, it appears that the optimal training for these skills requires an individualized instructional method which is particularly suited to each individual's cognitive and attitudinal make-up.

But if the optimal training for problem-solving skills and attitudes requires individualized instruction, what other method is there save a one-to-one tutorial situation which, while it may be effective, is neither efficient nor practical for large-scale education? To answer this question we must now examine the recent work done by Crutchfield and Covington and their approach to the problem through the method of programmed instruction.

3. THE CRUTCHFIELD AND COVINGTON PRODUCTIVE THINKING PROJECT

Crutchfield has stated the challenging problem facing anyone concerned with the teaching of problem-solving skills, namely "the dilemma that creativity training must be individualized as far as possible and yet that the materials and methods should be suitable for easy administration to entire classes..." (1965, p. 16). As already seen, none of the other attempts to train for productive thinking and problem-solving skills have been able fully to meet this challenge. Crutchfield and Covington believe that the best way to resolve this dilemma would be to cast the training materials into a loosely programmed self-instructional form. As they state, "The self-pacing, self-directing, and self-administering features of programmed instruction lend themselves directly to the requirements of creativity training, for these characteristics do place the focus of cognitive initiative in the individual, and they open the way for an optimal accommodation of the program to the distinctive cognitive style of the individual" (Crutchfield and Covington, 1965, p. 8). Some of the very procedures of orthodox programmed instruction, such as homogeneity of content and thought processes, effortless learning, authoritative lockstep sequences, and clarity and precision of each step, may be directly inimical to productive thinking. However, Crutchfield and Covington write, "But in fact, all of these features of programmed instruction potentially detrimental to creativity can be mitigated in their effects by avoiding overly strong commitment to rigid forms of programming and by inventing new programming techniques that are positively adapted to the requirements of creativity training" (1965, p.8). The features of programmed instruction which they feel are crucial for the training of problem-solving skills are: (1) the diagnostic testing of the individual as he works through the problem, (2) the appropriate branching techniques for greater freedom of choice of materials and alternative paths, and (3) the ability to create far more flexible forms of feedback which can be optimally suited to the distinctive responses of the particular individual. These are the special features of programmed instruction which they hoped would enable them to solve the dilemma of seeking to train children in thinking skills by using an individualized instructional approach which could be administered in the general classroom setting.

The nature of the materials programmed for the training of problem-solving skills depended on certain theoretical assumptions and pedagogical goals which Crutchfield and Covington (1965) have had in mind. They believe that training for productive thinking requires both the strengthening of certain cognitive skills which are central to the problem-solving process and also the encouragement of certain attitudes and motivational dispositions which favor the use of these skills. Three cognitive skills central to the productive thinking process are: (1) the ability to realize and formulate the problem from the given data, (2) the ability to generate many possible ideas for the problem solution which are not only uncommon, but also relevant to the situation, (3) the ability to evaluate these ideas, testing them against the demands of the problem facts.

But more than just the teaching of cognitive skills is necessary; the student must also be taught those attitudes and motivations which will lead to the optimal use of those cognitive skills. Four such attitudes and motivations which must be developed are: (1) a high value placed by the person on actually working on thought problems; (2) a self-confidence in his own problem-solving ability; (3) a maintenance of an openmindedness about the problem and the avoidance of premature commitment to one particular solution attempt or idea; (4) a readiness to continue working on a problem even if it proves difficult.

Above and beyond the specific thinking skills mentioned above, Crutchfield and Covington believe that there is a master thinking skill which enables the individual optimally to organize and utilize his specific skills in the actual problem situation. Thus they feel that the training materials must be of a nature which permits the subject to practice his specific thinking skills within "the global context of whole and relatively complex problems" (Crutchfield and Covington, 1965, p. 9). They would therefore take the opposite approach from Guilford (1956) who would train for each of these specific skills through separate factor-pure tests and training devices. Beyond the importance of a "creative acts-in-miniature" approach for the training of this master thinking skill, there is the important effect of this approach on strengthening the student's problem-solving attitudes and motivational dispositions listed above. This is especially true in developing persistence and enjoyment in working on problems, for these attitudes and motivations cannot be fully achieved through simple meaningless cognitive drill tasks but can only be maximally developed through working on training materials in which there are whole, complex and meaningful problems.

As they point out, "The challenging task of programming which faces us involves both the working out of appropriate methods and materials for creativity training and the casting of them into an effective programmed instruction form".

(Crutchfield and Covington, 1965, p.9). We have examined the basic rationale for their materials and methods; now let us examine the actual materials they developed and the results they obtained from several studies of these materials in actual school situations.

4. THE CRUTCHFIELD AND COVINGTON PROJECT MATERIALS

The Crutchfield and Covington project (Crutchfield, 1965, 1966; Covington 1965, 1966a) has developed a series of 16 lessons (each approximately 40 pages in length) in a semi-cartoon visual format within an auto-instructional form which undertakes to directly train for those cognitive skills and related attitudes of productive problem-solving discussed above. The lessons in this series (Covington, Crutchfield, & Davies, 1966: The Productive Thinking Program, Series One: General Problem Solving)[1] center around complex and compelling problems presented in story form which the students are called upon to solve. Each lesson is a complete problem-solving episode, containing all of the principal steps and processes inherent in creative problem-solving. As the problem develops the student is systematically led through the successive steps of the lesson which require him to learn about and practice a variety of problem-solving skills. On certain pages of the problem, the student is required to write out his responses and ideas; feedback to these responses is then provided on the following pages where he finds a range of illustrative ideas appropriate to the problem at that point.

The booklets have a continuous story line which follows the adventures of two school children, Jim and Lila (brother and sister), as they try to solve the series of detective problems and other mysterious and puzzling occurences which they become involved in. Jim and Lila are intended as examples for the students to imitate as they work through the problems. They are assisted by the help of their Uncle John, a high school science teacher and spare-time detective, who assists them through advice and encouragement as they gradually learn to work on problems by their own abilities. Each of the lessons is self-administering so that the child can work on the problem at his own rate of speed, in accord with his particular reading level and intellectual capacity.

Three major series of studies under controlled experimental design conditions have been performed using these materials. The first series (Covington and Crutchfield, 1965; Crutchfield, 1966) involved the initial school use of these materials and was carried out in two steps using over 480 students in the fifth and sixth grades in the Berkeley, California, School District. Overall, the results on the cognitive skills showed that the trained subjects scored at least twice as high as carefully matched groups of control students in the following problem-solving functions: number of problems solved, quality of ideas generated, relevance of questions asked, and sensitivity to cues and factual clues in the problem. Moreover, these gains occurred more or less equally over a wide spectrum of individual differences: among low achievers as well as high, among boys and girls alike, among the culturally disadvantaged as well as the advantaged. The results concerning the attitudes and motivations did not show more than a modest amount of change in favor of the trained students in these first studies.

The findings of these initial studies led to the revision of the materials which were then used in a full-scale study involving all 47 fifth grade classes in Racine, Wisconsin with the cooperation of the Research and Development Center for Cognitive Learning at the University of Wisconsin (Olton, et al, in press). Results show that the trained students again performed better than the control students in thinking and problem-solving performance on a wide variety of productive thinking measures. Again these significant instructional benefits occurred for all types of students regardless of sex, IQ, or classroom environment. Although these results were significant statistically, the magnitude of the differences was not as large as was found in the initial studies. This was believed to be due to the stringent conditions placed on the materials in that teacher participation was deliberately kept to a minimum in order to assess the materials as an entirely self-contained program of instruction (Olton, et al, in press).

An additional series of experiments has just been completed by the Crutchfield and Covington project in an effort to examine the effects of the materials when used in an enriched environment of supplementary student workbook materials and teacher interactive discussions. The results (Crutchfield and Covington, 1967) indicate that the effect of the materials on the trained students are substantial for cognitive skills, tests of which show large differences between the fifth grade experimental and control groups (and somewhat smaller differences for sixth grade students). Also, there are found to be significant increments in the positive attitudes of the trained subjects toward problem-solving activity and their self-concept as a thinker.

5. CRITIQUE OF THE CRUTCHFIELD AND COVINGTON PROJECT

The results of the series of experiments described above have been impressive and encouraging. It is clearly evident that it is possible to train directly for the cognitive thinking skills necessary for solving complex problems through the use of training materials developed according to the pedagogical and theoretical rationale of the Crutchfield and Covington project. Although the effects are not as strong, the indications are the same for fostering of the related attitudes and motivational dispositions. The project has successfully met the challenge set for itself in developing training materials which "must be individualized as far as possible and yet...be suitable for easy administration to entire classes" (Crutchfield, 1965, p. 16).

It is obvious however that a booklet used as the vehicle of programmed instruction has inherent limitations which cannot be overcome to make it a completely individualized instructional method. To be sure, the programmed instructional features of immediate feedback and active responding are present and the students, by themselves, can proceed at their own rate of work on the materials. But one of the essential features of individualized instruction is the evaluation of the student's response according not only to his present answer, but also to his relevant past pattern of responses, on the basis of which he can then be branched into the appropriate segment of the program. It is not possible to do this sophisticated evaluation and branching within a booklet format. Indeed, the booklet materials might better be considered as "personalized" instruction to use Grubb's term (1967b), rather than individualized instruction since they lack this crucial self-adjusting feature.

The advent of computer-assisted instruction has now given us the possible technological means to achieve a true individualized instructional method which can also fully implement those salient features of programmed instruction. It seems not only appropriate, but indeed imperative, to examine the feasibility of developing computer-assisted instructional materials similar in rationale to those developed by the Crutchfield and Covington project and to evaluate the effects of such materials when used in an interactive computer environment where actual individualized instruction is possible. If such encouraging training effects can be found by using a restrictive linear programmed booklet model, what might be the effects achieved with a real interactive training situation?

The answer to this question now becomes one of considering what technological devices are possible for such an interactive model. Let us turn to the other area of research, the technological one, which has directly developed out of the milieu of interest in actualizing individualized instruction and examine what possibilities it holds for effective and efficient implementation of this individualized instructional approach.

6. COMPUTER-ASSISTED INSTRUCTION: BACKGROUND

In order to arrive at a basic understanding of the term computer-assisted instruction, it is important to clarify the distinctions between the terms auto-instruction, programmed instruction, individualized instruction, and computer-assisted instruction. Auto-instruction, quite simply and basically, means the materials are in such a form that they can be self-administered and worked on at the student's own rate of speed. The other forms of instruction basically incorporate this aspect in their actual practice. The term programmed instruction means more than mere self-pacing and self-administration of materials as might be used, for example, with a simple reading lesson. Programmed instruction implies that there is a deliberate attempt to incorporate within the materials certain learning principles of active responding on the part of the student, immediate reinforcement to the response, and an appropriate remedial feedback to the response.

The term individualized instruction, in turn, means something more than simple programmed instruction as used in mechanical teaching machines, or booklets. On the one hand it permits a freer response situation for the student which is appropriately evaluated and responded to. On the other hand, it utilizes the past "sufficient history" of the individual, as Atkinson (1967b) terms it, for the decision of how to branch the subject in the program as well as the present response the person puts down for that item; whereas in programmed instruction usage the only information used for branching is the actual present response itself. There are other related areas in which there are practical differences between these two terms, but these are the basic intrinsic differences.

Any instrument designed to be used for individualized instruction must be able to handle these three basic functions which form the core of the interactive processes: (1) there has to be a way of response recognition which permits the student to respond freely and which correctly identifies those responses; (2) there has to be an evaluation and scoring procedure which can keep tally of the subjects' patterns of responses as well as any other personal historical information which may be deemed important in the evaluation of the student's responses; and (3) there has to be a means of complex branching so that the most appropriate feedback and remedial program segments can be given to the subject in relation to his ongoing particular state of ability and performance. These are the three attributes against which any method which claims to implement the individualized instruction concept must be assessed.

Although the traditional one-to-one human tutorial interaction is generally posited as the model individualized instructional situation, it is not feasible as a mass education technique. But some authors would even go further than that and argue that it is really not a complete individualized instructional situation at all since it is not possible for a human tutor always, or even generally, to make use of the pertinent sufficient history of the individual in evaluating his response (Atkinson, 1967b). A quick look at the inherent limitations of using programmed instruction booklets, even the attempt at using scrambled booklets, shows that this method is not adequate to fulfill the three functions mentioned above. The mechanical teaching machines, at first glance, seem to offer some new possibilities, such as branching techniques, but even here their inherent limitations on all three counts prevent them from seriously being considered as adequate individualized instructional devices.

It is thus becoming evident that the only instrument which offers any feasible promise of implementing large scale individualized instruction is the computer-assisted instructional system (Atkinson, 1967b; Suppes, 1965). The computer-assisted instructional system is the only method which potentially offers the technological possibility of fulfilling all three necessary functions of individualized instruction of: (1) response recognition, (2) evaluation and scoring, and (3) complex branching.

We shall now examine the brief history of computer-assisted instruction to see if, indeed, the actual future of this instructional method is as promising as would seem theoretically indicated.

7. COMPUTER-ASSISTED INSTRUCTION: HISTORY

Initially, computers in education were conceived of mainly for purposes of administration and accounting. However, in recent years more and more attention is being given to the use of the computer in the actual instructional aspect of the educational situation. Though a flood of report releases has poured forth about the possibilities and potentialities of computer-assisted instruction, as recently as September, 1967, Atkinson writes, "...the majority are vague speculations and conjectures with little if any data or real experience to back them up" (1967a, p. 1). The need clearly is for more experimental research and actual data.

There are, of course, numerous research projects in the area of computer-assisted instruction, but the number is not as large as one might expect from all the speculative publicity. Zinn (1967) gives an excellent short summary of twenty-six current computer-assisted learning projects. In general, these projects have shown that computer-assisted instruction systems can be feasibly installed and operated in experimental classroom settings. Several projects are operating with a sufficient number of terminals so that actual programs of school training can be performed (Atkinson, 1967a; Bitzer, et al, 1967; Carter and Silberman, 1965).

The work at Stanford done by Suppes and Atkinson has been fundamentally concerned with "drill and practice" materials for lower grade children in the basic skills involved in mathematics (Suppes, 1966a, 1966b, 1967; Suppes, et al, 1966) and in reading (Atkinson, 1967a, 1967b). The project is being used as an actual part of the regular school instruction and the overall results have been very encouraging.

At the University of Illinois, the PLATO (Programmed Logic for Automatic Teaching Operations) project, as well as the SOCRATES project of Stolurow (1965), have been investigating the possibilities of individualized instruction for large numbers of college students through computer-assisted instruction (Braunfeld and Fosdick, 1962; Bitzer, et al, 1966; Lyman, 1967). There are even regular credit courses being given by means of computer-assisted instruction in Fortran programming, library usage, and electrical engineering (Bitzer, et al, 1967).

The System Development Corporation (SDC) has developed the CLASS System (Computer-based Laboratory for Automation of School Systems) which has twenty teaching stations (Coulson, 1962, 1964; Carter and Silberman, 1965). Earlier studies which involved the testing of specific hypotheses about the learning situation (Coulson, et al, 1961, 1965) have given way to a more informal tutorial situation in which there is a continual process of computer and instructor interaction with the students for the purpose of developing hypotheses about programming strategies (Carter and Silberman, 1965). An interesting study involving academic counseling has also been performed (Cogswell and Estavan, 1965).

The work carried on at the Pennsylvania State University is basically concerned with the same problems (Mitzel, et al, 1966). The reader is referred to Zinn (1967) for a detailed account of these and other computer projects.

Generally speaking, the problems of hardware-- i.e., the actual electronic computer and its related physical accessories-- have basically been solved, and the crucial problem now is the development of good programming materials to use on these systems. As Suppes puts it, "In fact, the principal obstacles to computer-assisted instruction are not technological but pedagogical" (1966b, p. 208).

In examining the computer-assisted instruction projects described above, and the other projects involved in this area, there is only one project (Moncreiff, 1965; Leonard and Wing, 1967) which has reported putting main pedagogical emphasis on generalized problem-solving skills. The other projects have been basically concerned with the mastery of standard subject-matter content. In some cases there has been some indirect emphasis on specific thinking skills, but these skills have been ones closely tied in with the content of the subject materials in such areas as statistics (Uttal, 1962), mathematics (Suppes, 1965), and engineering (Mitzel, et al, 1966).

Leonard and Wing (1967) describe three different types of interaction games which were developed to teach basic economic principles to sixth graders.

in the basic design of these interactive programs the students make certain economic decisions on the basis of the facts they receive in the program, then the computer responds by telling them the consequences of their decisions. The student is continually presented with new information and "chance happenings," setting up a series of problematic situations which he works through to the solution of the game. The Sumerian Game trains for basic economic principles in the neolithic revolution of Mesopotamia during the fourth millennium B.C.; the Sierra Leone Development Game simulates the economic problems of newly emerging nations; and the Free Enterprise Game simulates the economic decisions which the owner of a small toy store faces in the competitive business world. In a small study using only 26 sixth graders trained on the computer game programs compared with other students taught in a conventional manner, the authors briefly report that the experimental groups performed better than the controls in one game posttest while the controls did better in the other posttest.

Leonard and Wing (1967) end their article by pointing out that the nature and development of these games preclude the use of any instructional method other than electronic computers for the feasible simulation of the environmental situations and problems which the students work on. This is their conclusion to their initial question, "What can the computer do better than conventional texts, programmed books, manual teaching machines, or board games?" Outside this one study, there appear to be no current projects which use the computer in training for generalized problem-solving skills by utilizing an interactive game structure. This is one of the basic concerns of this thesis.

Another basic problem affecting the actual development of computer-assisted instructional techniques which has directly concerned this thesis and which will now be discussed is the author language used for computer programming.

8. COMPUTER AUTHOR LANGUAGES: COMPUTEST

One initial problem that confronts the person interested in developing computer-assisted instructional programs is the obvious need for translating the written materials into an appropriate computer code. The choice of which of the computer language systems will be used is directly related to and affects one's entire scheme of how computer-assisted instruction will fit into the existing educational system. If the computer language is complex and difficult to learn, it is quite certain that the programs will be written and coded by experts in the field and not modified or revised for local use by the teacher; nor will the average teacher ever likely become involved in writing programs for the pupils to use. This approach would be a centralized one in which local schools simply used unmodified materials given to them by the outside agency, without teacher or pupil interaction with the computer in the writing of their own programs. Indeed, in this case not even the content specialist would be likely to code the materials, but rather they would be given to special coders to do the job.

Zinn (1967) points out that most users of the computer for instructional purposes will not have a command of complex programming logic, or a programming assistant to code the materials for them. Thus he states two characteristics of a computer author language which he feels are essential: (1) it must be simple enough so that the user can write the self-instructional materials he intends to use on the computer in his own language with a minimum of restrictions; (2) it must also contain the programming sophistication which will permit the experienced user to write materials which utilize the capabilities of the computer to the fullest.

Zinn then goes on to give five examples of computer languages which are in current usage. The main language in use for this purpose is IBM's Coursewriter;

other languages, such as SNOBOL and MENTOR are mainly specific to the particular individual computer projects. Another of the languages mentioned is COMPUTEST, which has been developed by Starkweather (1965, 1967a).

In actual usage, IBM's Coursewriter language requires a fair knowledge of programming logic and familiarity with the computer system before it is able to be effectively used. Starkweather has deliberately attempted to develop a computer language which would be easier to use by an inexperienced person (Zinn's first point) while at the same time offering for the more experienced user all the computer controls necessary to use the system to its capacity (Zinn's second point). Starkweather (1967a, abstract) describes his author language as follows:

"COMPUTEST[2] is a problem oriented programming language for computer assisted instruction, testing and interviewing. Sequences of instructional material and test questions may be written in natural language and a variety of cues may be used for the recognition of an answer from the typewriter input. Variable comments and choice of the next question to be asked may be determined by the evaluation of an answer. Scoring and data collection is optional for each question."

Starkweather (Hodge, 1966) had two basic reasons for his desire to build a programming language which would be easy to use by inexperienced persons. The first was his belief that computer-assisted instructional materials should have the flexibility to be modified to suit the needs of the particular area or group which is using the materials. This would necessitate a simple computer language which the interested teacher could easily modify as regards to accepted responses, wording of the feedback output, etc. In this way it is hoped that teachers would be able to write their own special programs dealing with those subject areas which were most pertinent to their pupils. Starkweather's second basic reason was his wish to make the language simple enough so that even elementary school children would be able to write their own programs on the computer. He expected that by having children write such programs they would soon learn to develop certain thinking skills, such as the ability to formulate unambiguous questions and to think of a wide range of possible answers to their questions rather than just considering one answer to be correct.

The COMPUTEST language, and the Dixie Computer Learning Project of Starkweather, offered the best technological means of examining the feasibility of utilizing computer-assisted instruction techniques for the training of generalized problem-solving skills as set forth in the Crutchfield and Covington reports. This thesis developed out of the synthesis of methods from these two projects.

9. ASSUMPTIONS, RATIONALE, AND OVERALL OBJECTIVES OF THIS THESIS

The working assumptions underlying this thesis are the same as those which have laid the foundation for the work by Crutchfield and Covington (Crutchfield, 1964, 1966). They are the following:

1. Virtually all individuals fall far short of their creative thinking potential and they can be helped in this regard through appropriate instructional methods.

2. The basic skills involved in productive thinking are general skills which can be directly trained for.

3. The facilitation of productive thinking performance is not so much a complete training of new skills, as it is a sensitization and activation of cognitive skills already possessed to some degree by the individual and an encouragement of related beneficial attitudes and motives in the individual.

4. In view of the above assumptions, even relatively modest efforts in

appropriate instructional programs may bring about significant increases in productive thinking performance and attitudes.

The major assumption upon which this thesis is based is that the best instructional method for training for productive thinking skills is through an individualized instruction approach in which the subject can freely respond and, on the basis of that answer as well as other pertinent information of his past responses, be branched into the most optimal feedback for him within the program. The Crutchfield and Covington project has made a start in this direction with their use of a programmed booklet format, but it is probable that they have taken programmed instruction as far as it can go without really having achieved a basically individualized instructional method.

The overall purpose of this thesis is experimentally to examine and evaluate the feasibility of utilizing a computer-assisted instructional method for the direct training of generalized productive problem-solving skills based on the theoretical and pedagogical rationale of the Crutchfield and Covington project. To achieve this goal, specially developed computer instructional materials are used to train certain groups of experimental subjects in productive problem-solving skills, while comparison groups of subjects, who do not receive the computer materials, work on other materials during the training period. The analysis and evaluation of such a computer-assisted training approach is actually a twofold problem; it involves the evaluation both of the effectiveness of the particular computer training materials and of the efficiency of the computer language in which the training programs are written. Although these two aspects are intimately interrelated, they can and must be examined separately for a complete evaluation of a computer-assisted approach expressly designed to train for productive problem-solving skills.

The effectiveness of the computer training materials can be ascertained through posttest measures which evaluate the computer subjects' performance and attitudes against the performance and attitudes of comparison subjects. While, in one sense, an "empty" control group of subjects who take only the posttest --without any kind of material during the training period-- would be sufficient as a comparison group against which to test the effectiveness of the computer training materials, it seems desirable to utilize additional comparison groups in order to achieve an even more stringent test of the effectiveness of the computer instructional programs. Accordingly, these additional comparison subjects receive experimental materials which involve them during the training period for the same length of time as the computer subjects. This is intended to help better equate the motivational and attitudinal effects resulting from participation in the experiment. The posttests are designed to yield seven measures which have been selected to determine the effectiveness of the computer training materials in training for productive problem-solving performance skills and attitudes. Included are three problem-solving performance measures concerning the abilities: (1) to formulate the problem, (2) to generate many uncommon yet relevant ideas, and (3) to evaluate these ideas against the given facts. Also there are four measures concerning the following relevant attitudes: (4) a high positive value placed on working to solve thought problems, (5) a high self-confidence in working on such problems, (6) maintenance of an openmindedness about the problem and the avoidance of permature commitment to one particular idea or solution attempt, and (7) a persistence in continuing working on difficult problems despite setbacks.

The question of the efficiency of the computer program has to do with the basic evaluation of the COMPUTEST language, that is, the question of whether it is practical and feasible to use this language in writing computer programs which will fulfill the three basic and essential requirements of an individualized instructional system: (1) response recognition, (2) evaluation and scoring procedure, and (3) complex branching as mentioned above.

The answer to this question of efficiency of the programming method is in one sense even more important than the question of the effectiveness of the particular training programs developed. For regardless of how effective the particular training materials might be, the outcome would be purely of academic interest without any relevancy for actual school use if the materials were drastically inefficient to develop. On the other hand, even if the particular materials used were not strikingly effective but did show some positive training merit, and the programming method was basically efficient to use, then it would be possible and feasible to develop new training materials based on the strong points and eliminating the weak points of the initial materials.

The choice of the particular subject content area of the computer-assisted training materials used in the experiment must be taken into consideration in the evaluation of their effectiveness and efficiency. In interpreting the final experimental results, one must, of course, take care not to overgeneralize the findings to all such computer materials and to the CAI approach in general. Other curriculum content areas have their own special requirements and problems for effective instructional procedures, and the value of computer-assisted instruction may vary among the content fields.

For this thesis the decision was made to choose the field of American History as the instructional content. This decision was based on the fact that the subject matter of American History lends itself very readily to the development of instructional materials directly concerned with training for general problem-solving skills and attitudes similar to those which have been discussed above in conjunction with the Crutchfield and Covington project. This choice was further encouraged by recent concentrated research efforts of various social science educators (Fenton, 1966) to develop and utilize various problem-solving approaches for the teaching of history in the secondary school situation. This made the choice of American History as the subject area for the computer programs of this thesis particularly appropriate and meaningful in light of these current curriculum reform efforts.

With this overview of the assumptions, rationale, and overall objective of this thesis in mind, we shall now examine the actual training materials developed for this thesis and the experimental design by which the study was conducted.

II. EXPERIMENTAL DESIGN AND METHOD

I. TRAINING MATERIALS

One of the two main questions with which this thesis is concerned at this exploratory stage is whether it is possible to develop a set of training materials for productive thinking skills and attitudes using a computer-assisted instructional method so that the experimental subjects taking these materials can later perform better on solving problems than the comparison subjects who do not work on the computer training materials. The comparison subjects work either on booklet materials, or have no work at all during the training period.

Accordingly, the subjects were divided into the following seven main groups depending on the type of training materials they received:

A. Computer Subjects

1) Computer General Problem-Solving Group

2) Computer Programmed-History Group

3) Combined Computer General Problem-Solving and Computer Programmed-History Group

B. Comparison Subjects

4) Booklet General Problem-Solving Group

5) Booklet History-Reading Group

6) Booklet History-Reading Group with Thinking Guides

7) Untutored-Posttest Group

The descriptions of the training materials used by each of the different groups are given below.

A. Computer Subjects

1) Computer General Problem-Solving Group (Computer-GPS)

This group received slightly modified versions of eight lessons selected from The Productive Thinking Program, Series One: General Problem Solving (Covington, Crutchfield, and Davies, 1966)[1], namely, lessons 1, 2, 3, 4, 5, 10, 14, and 15. These lessons, which directly train for problem-solving skills and attitudes through direct explication of thinking guides and examples, are included in this thesis to determine the effects of such instructional materials when used in a simulated computer interactive setting. Although one approach would have been entirely to recast the lessons into complex branching computer programs, this task was beyond the scope of this thesis. However, by arranging these materials in the manner described below, it is believed that the main beneficial effects of a "computer" interactive training setting could still be achieved with only minor textual changes in the original lessons. Even though these materials were originally developed for fifth and sixth grade use, it is believed that they are written in such a way that eighth grade students could still substantially benefit from them in varying degrees.

The eight lessons were selected on the basis of the particular problem-solving skills trained for within each lesson as well as the appropriateness and general continuity of the story content within the overall context of the other training materials of the thesis. These materials were presented in a simulated computer setting which utilized a Teletype typewriter connected to a papertape drive unit as the simulated remote "computer" terminal and a rear screen random-access slide projector. Each page of the booklet lessons had been previously photographed onto a separate slide after the necessary minor textual modifications had been made, such as changing the word "page" to "slide", eliminating pagination, and changing the instruction "write" to "type". Besides these necessary minor procedural changes, no other modifications were made in the program content. Preliminary study showed that the average eighth grade student could do approximately one and a half lessons per 40-minute class period; therefore, the design plan of having five units of materials for each training group led to the eight lessons being divided into five units at natural breaking points which did not interrupt the sequential nature of the programs.

The actual presentation of these materials to the subjects went as follows. First the subject would type his name on the Teletype typewriter which would respond by typing back an instruction to read a particular slide. Each page of the booklet lessons had been photographed on a separate slide and the subject would continue reading the slides sequentially until he reached a slide which required him to respond. In the original booklet lessons he would have written his answer on the page; here he typed his answer on the computer-simulated remote-terminal typewriter. When he finished typing his answer, he would press a button indicating he was through and the "computer", after an appropriate time lapse, would type back a noncommittal feedback response such as, "That is a good idea, now go to slide 46". These feedback responses were prearranged on a paper tape to be typed back at the particular response points. Appendix A shows the feedback responses which were actually typed back to all the subjects regardless of their answers as they worked through the materials. Thus, irrespective of the subject's response, he was instructed by the feedback to jump to a designated slide which would then continue the next part of the lesson. The general scheme was to have the subject read the slides consecutively until he came to one which required him to respond and then, after his response, jump him over a varying number of slides to continue the program. This was intended to simulate a computer-branching technique for the subject, while in actuality the materials were being presented in a straight linear fashion as in the original booklet version.

2) Computer Programmed-History Group (Computer-PH)

The history computer programs are designed to utilize the unique interaction-game potentialities of the computer (cf., Leonard and Wing, 1967) by allowing active practice in working on various thought problems. The main rationale of these history materials is to provide computer programs which require the active continuous responding of the subject in order for him to arrive at the problem solution.

The training materials for this group consisted of five computer programs which were especially developed for this thesis. Since one of the aims of the thesis was to demonstrate the feasibility of using the computer-assisted instruction techniques within a social sciences curriculum, the subject matter of these computer programs concerns real and fictitious historical settings and problems. The general theme of the first unit (CAMERIA) is of a fictitious historical problem of the mysterious gathering and disappearance of a vast number of people which happened during the time of the Western Expansion period in the imaginary land of Cameria-- a land whose past is similar in many ways to our own American history. The second and third units (THE SILVER FOX LETTER-Parts I and II) deal with a made-up historical problem about the complex and unknown factors involved in the start of the American Revolution. The fourth unit (SAN VALLENDO) is again concerned with the Wild West in the land of Cameria as the problem of how to get a wagon train across the dangerous frontier faces the subject; coupled with this lesson is a short, explicit review of the thinking skills which the subject has implicitly been using in working through the problems. The fifth lesson (SLAVERY AND THE CIVIL WAR) presents to the subject the task of resolving conflicting accounts of the slavery issue written by different people who lived during the time of the Civil War. The complete set of these computer program materials, including the COMPUTEST coding, is reproduced in Appendix B.

Each unit centers around the presentation of pretended or actual historical information which confronts the subject with a complex and interesting problem to be solved. Initially the subject is shown different ways to approach the problem; then, as the problem develops, he is systematically led to think of different and unusual ideas which might explain the puzzling and discrepant facts of the story. Then he is posed with the problem of evaluating these possibilities against the additional information he receives. Also at the same time, he has to revise and develop other possible solution hypotheses on the basis of this new information until he finally arrives at a satisfactory solution to the problem.

As an example of these materials, the content and format of the SILVER FOX LETTER will now be examined in a little more detail. Dr. Hogan, a fictitious historian, finds a letter written in French in an old estate house of the Warren Family which is up for auction in the Boston Area. The Warren estate dates back to the pre-revolutionary era during which Charles Warren was quite an active figure. After the historian and student have done some initial sleuthing and fact finding, they discover that the letter implies that R. Newman (who was Paul Revere's lantern man in the church tower on the night of his famous ride) had been set up by some French officials or merchants as a possible person to be bribed who would arrange an incident inciting the colonists and the British to attack each other and thus precipitate a war. The motive is later discovered that the Frenchmen involved would have reaped the benefits of renewed fur trade with their former possessions in the new frontier. In this program the subject is led to discover ways of determining the authenticity of the letter, to make hypotheses about the meaning of the letter from its truncated content, to arrive at the identification of the persons implicated in the letter (discovery of R. Newman does not occur until the end of the first lesson), to think of possible ways of obtaining pertinent information, and to think of the possible implications of this historical event. More information is given to the subject as he works through the problem so that he can test his hypotheses and eventually arrive at a satisfactory conclusion of the problem based on the given facts.

A simplified view of the computer training situation is as follows. The subject sits at the computer console typewriter above which is the screen for the random-access slide projector. The subject first types his name on the computer typewriter which responds by giving him information to start a new lesson, or to continue his previous lesson by presenting the last sequence of information he was working on. Eventually this information sequence ends in some form of question which the student answers by typing his reply on the typewriter. The computer responds by branching the subject to an appropriate segment of the program on the basis of his present and past pattern of responses. In this new segment of the program, the subject receives new information and another question to work on. This constant active game-interaction between the subject and the computer continues until the student has arrived at a satisfactory solution to the problem.

3) Combined Computer General Problem-Solving and Computer Programmed-History Group (Combined-Computer)

This group alternately received the training lessons of <u>both</u> the computer-GPS and the Computer-PH group within their respective experimental formats. This is designed to provide the maximal instructional benefit of the computer training materials.

B. Comparison Subjects

4) Booklet General Problem-Solving Group (Booklet-GPS)

The subjects in this group received the same eight selected lessons from the <u>General Problem Solving</u> series[1] as the Computer-GPS group received. These materials were presented in the standard booklet form to the subjects and were administered on an individual basis with the subjects writing their responses on separate answer sheets.

5) Booklet History-Reading Group (Booklet-HR)

These subjects received the content of the computer history programs which were rewritten into five continuous stories, each one being approximately twenty double-spaced pages. (See Appendix C for the complete set of materials.) These stories were intended to provide reading materials directly related to the story-line of the computer history programs in order better to equate the motivational content features of the history reading materials with the computer history programs. At the end of each story, the subjects were required to write answers to five "end-of-the-chapter" questions about the material they had just read. Each unit required approximately one class period to complete.

(6) Booklet History-Reading Group with Thinking Guides (Booklet-HR + Guides)

This group received the same materials as the Booklet-HR group with the addition of selected appropriate thinking skill guides which the subject read before answering the questions at the end of the story. This comparison group is intended to be analogous to the "Rules Only" group used by Covington and Crutchfield (1965) which showed a modest beneficial effect over an untutored group of subjects. (Appendix D shows the thinking guides which were appended to each of the five history reading lessons.)

7) Untutored-Posttest Group (Untutored-PT)

The subjects in this group, intended as the control subjects in the usual use of the term, did not receive any type of training materials during the study, and only participated in this experiment when they took the posttests along with members of the other groups.

2. FACTORS DETERMINING MAXIMUM NUMBER OF TRAINING PROGRAMS

The fact that five lesson units of training materials were used for all the groups (except the Untutored-PT group) was not arrived at arbitrarily, but after much consideration of several important factors affecting this experiment. First, of course, was the issue of what was the minimum number of lessons that could be given which would still show differential treatment effects. This was not begging the question, but rather establishing the absolute minimum conditions. From the assumptions mentioned in this thesis, and those of Crutchfield (1966), it was felt that even a modest amount of training would be sufficient to bring about performance and attitude changes. In fact, in the initial Covington and Crutchfield studies (1965) subjects working on the training materials significantly surpassed the control subjects after taking only the first five lessons of the General Problem Solving series. It seemed feasible to expect that one could find differential effects after five lessons of training materials which would take from five to eight periods to complete.

However, even more training materials might have been utilized had it not been for the three additional factors which set limits on the number of training lessons at approximately five and the number of subjects at approximately twenty-five per training group. First, the time period chosen to run the subjects through the experiment was the period after the winter vacation and before the spring vacation, a period eleven weeks long consisting approximately of 440 class periods. It was decided to run all the subjects within this time period without a major two-week vacation interrupting the sequence of the experiment. Also, the time period before the winter vacation was needed for developing and testing the materials, while the time after the spring vacation was felt to be too involved with graduation (these were eighth graders) and uncompleted school work to permit as optimal a testing situation as during the period chosen.

The second factor had to do with the large inter-subject variability found in most educational research. Owing to this, it was felt that at least twenty to twenty-five subjects would be the minimum number possible per training group since each training group was also to be subdivided into two levels of sex and two levels of IQ. This would then mean only five to six subjects per cell, a number it was felt which could not be lower for meaningful statistical analysis.

The third factor, which was the most important in the final determination of the actual number of lessons and subjects used,was the limitation that there was only a single terminal connected to the IBM 1620 computer on which to run the computer trained subjects. To meet the other two requirements of time and statistical reliability, a choice had to be made in the experimental design between two computer groups having five training lessons lasting approximately five to eight periods, or one computer group with more training lessons. Based on the encouraging results of the studies run by Covington and Crutchfield (1965) and the fact that the initial overall experimental design had been conceived when it had been anticipated that there would be two remote computer terminals connected to a larger IBM computer, it was decided to follow the original design and run two computer groups. This decision also set the pattern of five training lessons for the other groups.

3. PLANNED COMPARISONS

The primary purpose of this thesis is to evaluate the effectiveness of a computer-assisted instructional approach for the training of productive problem-solving skills. Three different "computer" groups are utilized: one involving the General Problem Solving materials in a simulated computer training situation, another involving specially developed history materials in a real computer training situation, and third, a Combined-Computer group which receives the training materials of both the other two computer groups. It is reasonable to assume that if the computer training materials do produce an increment in productive problem-solving skills the Combined-Computer group subjects will perform the best of all the three computer groups, as well as the other comparison groups.

As mentioned before, to permit a more appropriate test of the effectiveness of the computer training materials, it is believed that subjects who participate in some form of experimental materials will be better comparison subjects than just "empty" control subjects who take only the posttest materials. Accordingly, besides the "empty" posttest control subjects, three additional different groups of comparison subjects are utilized in the experimental design who participate in the experimental training period by reading and working on various booklet materials as described above.

It must be stressed that the purpose of these three comparisons groups is not to make a direct comparison between two different methods of training--a computer versus a booklet format. If such a direct comparison of two teaching methods were intended, it is quite obvious that the booklet materials would have to have been given in a much different form, namely a branching programmed format, in order to allow any conclusions to be made about the relative merits of the two different types of teaching methods. But this approach is fraught with intrinsic methodological dilemmas and unavoidable pitfalls which vitiate any meaningful interpretations (Lumsdaine, 1962). In the first place, if one were trying to make such a direct comparison of two teaching methods, it would be naive to assume that within five units of training materials one would be able to get reliable results indicating the relative merits of the two different instructional methods involved. And even if one could assume differential effects between the instructional methods within that short a time period, the problems involved in directly equating and comparing the differences of the two methods such as branching, feedback, reinforcement, etc., are so complex as to preclude any meaningful evaluation of the relative merits of the two teaching methods.

Cognizant of these pitfalls in the attempt directly to compare two types of instructional approaches, our aim is not to make a comparison between two different training methods, but rather to determine whether computer-assisted training for productive thinking skills can succeed. To achieve a better measure of the effectiveness of the computer training methods, rather than through just a comparison with an "empty" untutored control group, several different comparison groups of subjects are utilized. These other comparison subjects are experientially involved in the training period through various experimental materials and serve as baseline performance and attitude subjects against whom the computer subjects are contrasted. But the reader must note that the relative merits or demerits of the reading materials are not being compared to the computer materials. The reading materials are intended as materials to experientially involve those subjects during the experimental training period; while, however, the computer materials are directly organized as training materials designed to increase those subjects' performance abilities and attitudes relating to productive problem-solving. This general evaluative approach of the computer training materials seems quite appropriate and indeed warranted by the paucity of actual data in this area (Atkinson, 1967a). Also it is consonant with the approach taken by other investigators in the area of computer-assisted instruction who have previously tried the method of direct comparison between various aspects of different teaching methods and have not found this approach to be very fruitful (Carter and Silberman, 1965; Licklider, 1962) and have turned instead to general experimental approaches which are aimed at analyzing and improving various computer-assisted instructional programs and methods.

The actual statistical method and rationale of our comparisons among groups will be treated in detail in the section below on Statistical Procedures for Analyses of Results; however, some discussion at this point of how and why the different computer and comparison groups will be contrasted will give the reader a better grasp of the design and rationale of this exploratory experiment. For clarification, suffice it to say that although there are many possible sets of comparisons that will fit the statistical requirements of the planned comparisons method of analysis, once the initial comparisons are chosen the number and type of other possible comparisons which can be included without becoming statistically redundant in that set becomes quite limited. Thus, in actual practice, these secondary analyses are usually not as important as the initial comparisons. The same is true in this case: There are two direct critical tests of the effectiveness of the computer-assisted instructional materials, and four subsidiary, or secondary, statistical comparisons which are made between the various computer and comparison subjects in order to aid in the further interpretation and understanding of the effects of the computer-assisted training materials.

One of the two major comparisons involves the evaluation of the three computer groups as contrasted with the three comparisons groups who took booklet materials. This is the main direct attempt to answer the basic question of this thesis: Do the computer training materials lead to an increment in performance measures and a positive change in attitudes concerning productive problem-solving skills? As mentioned above, it was felt that the use of these comparisons groups who were experientially involved in the training period offered a more appropriate baseline group of subjects for evaluating the effects of the computer training materials than just an "empty" untutored group of subjects.

The second major comparison involves the evaluation of the Combined-Computer group against the other two computer groups. It is expected that the Combined-Computer subjects will perform better than all the other experimental groups owing to the beneficial interaction effect of the two types of computer training materials received and the lengthened time of involvement in the training lessons. Accordingly, the other two computer groups can be most appropriately used as the baseline subjects against whom to test any increments in performance and positive attitudes regarding productive thinking skills achieved by the Combined-Computer subjects.

The next four planned comparisons are made with the express purpose of aiding the interpretation and understanding of the results of the two major comparisons described above. These secondary, or subsidiary, comparisons are not intended to be analyzed or interpreted in isolation from the two main comparisons described above and, indeed, are defensible only as related to, and in the context of, these two major questions. These four comparisons are arranged in the order of their relative importance in aiding in the overall interpretation of the main results.

The most important of these subsidiary comparisons concerns the comparison between the two separate computer groups. Even though there are a number of incidental factors which cannot be completely equated, there is a main difference between the two types of materials which can be looked at and evaluated to aid in the understanding of the effectiveness, if any, of the computer training materials. This difference is that, in this thesis, the General Problem Solving materials provide direct explication of the performance skills and attitudes necessary for successful problem-solving without providing an interactive responding situation. While, on the other hand, the specially developed history materials have as their main emphasis the interactive-game responding between the student and the history programs without much direct explication of the problem-solving performance skills and attitudes. This comparison will give subsidiary information which is important in two ways: First, it will indicate the relative effectiveness of these two different types of training materials in an informal way, and second, it will enable a more meaningful analysis of the Combined-Computer group's performance to be made from the analysis of each of the two types of training materials when taken separately.

The next comparison involves the analysis of the Untutored-Posttest subjects to determine the baseline performance and motivation of subjects who are not involved in the experimental training period at all, but who simply take the posttests given at the completion of the instructional materials. (Note: although this group is second in importance in the subsidiary comparisons, it is the first comparison reported in the Results and Discussion section in order to maintain a logical order of the six planned comparisons.) This group constitutes a classic "empty" control group of subjects which is used to evaluate any extra-experimental variables which might have an influence on the experimental performance of the involved subjects.

The last two analyses involve less important comparisons which are incidentally made in the attempt to glean some information which might be helpful in the understanding of the other results obtained. The more important of these two analyses contrasts the programmed method of direct teaching of thinking skills in the General Problem Solving group with the reading materials of the other two history groups. This is informally done to see if any major performance or attitude effects can be grossly attributed to the differences in the training materials. The last comparison is made to see if the history reading group which also reads thinking skill guides will perform better than the other history group without any thinking guides.

It is realized that in these secondary analyses the differences between groups cannot be solely attributed to any single differentiating factor; there is a basic confounding of story content and mode of responding. However, looking at the differences from an overall viewpoint, it is still felt that some meaningful interpretation can be made which will help future research in this area. The basic pedagogical question is still-- Can computer-assisted instructional materials be used for training students in productive thinking skills and attitudes so that they can perform better than other comparison groups who do not have such computer instruction? If the answer is in the affirmative, then future studies can begin to explore the relevant variables more minutely and plan more stringent control groups. At this stage of research, the computer training materials and methods are not that sophisticated to make such comparisons of particular hypotheses against stringent control groups very meaningful.

4. POSTTEST MATERIALS

The posttests were designed for two fifty-minute periods and consisted of two problem performance tests (WHITE HORSE INDIANS and KASKIA) being given in the first period, and an attitude questionnaire and another problem performance test (PANJA) being given in the second period. The posttests were paper-and-pencil tests which were administered within the classroom setting. The complete set of posttest materials is found in Appendix E.

The three performance problems were specially and entirely developed for this thesis. They were written within the general rationale and format of the posttest materials for productive thinking originally developed by the Crutchfield and Covington project (Covington, in press; Covington and Crutchfield, 1965).

These three problems varied in their emphasis on the three different performance measures contained in each one. All three problems had their own particular open-ended questions concerning either the setting up of the problem, the noticing of puzzling facts in the story information, or the seeking of additional information through appropriate questions. The second performance measure was that of idea-generation, while the third measure concerned the evaluation of possible solution-ideas against new information given in the problem.

The WHITE HORSE INDIANS problem had its main emphasis on idea-generation; while the problem on PANJA had its primary emphasis on idea-evaluation. KASKIA had an equal emphasis on both idea-generation and idea-evaluation. All three problems had the same measure of problem-evaluation which was given at the end of the problem and consisted of a five variable rating sheet with completion sentences.

The WHITE HORSE INDIAN problem presented an initial page of information of some pertinent facts concerning that particular mythical tribe's way of life and also their relations with the white settlers who had come into the area. The information set up the puzzling problem that the Indians and settlers had lived in harmony for three years when suddenly the Indians attacked and drove the settlers out of the area and then moved away themselves. The subject first works on discovering the puzzling information in the story and then he is asked to think of the possible reasons why the Indians might have gone on the warpath, and then why they left the area. A set to write down original ideas is induced by the direct instruction to put down different and unusual ideas which might explain these happenings.

KASKIA concerns the problem of a small group of settlers on their way from Kaskia to Fort Hall who had completely disappeared and had never been heard from again. The subject first has the chance to write down questions on which he would like to have more information, then he is asked to think of possible reasons which would explain the settlers' disappearance. A map is provided to help give some clues to other possibilities. Next, the student is given a list of eight possibilities which might explain why the settlers disappeared, and he is asked to evaluate these possibilities on the basis of new information which he receives.

PANJA is about a small mythical country which has existed peacefully for 200 years and which has suddenly erupted into violence and civil war. The subject is to take the part of a newspaper reporter and cover the story. He first states how he would start his job and then he is given more information about the country. After that he is asked to put down possible reasons why he thinks the violence has broken out. Following that, he is given five possibilities of the causes of the civil war and also new information which he is asked to use in evaluating these possibilities. Again, some new, different information is given and he is again asked to evaluate the possibilities. Finally, the critical information for the problem solution is given the subject for him to consider and evaluate.

The questionnaire was divided into two main parts. The first part dealt with the student's self-evaluation as a problem solver and was adapted from an attitude inventory by Covington (1966b). This part consisted of two sections both written in a dichotomous yes/no item choice format. The first section consisted of ten items concerned with various cognitive approaches used to solve problems; the second section comprised seventeen items which were more subjective in nature concerning how the student felt about his own thinking ability and his performance on problems.

The second part of the questionnaire was about the student's evaluation of the training materials. The major section was a numerical rating scale of five different dimensions concerning the training materials; while the other section consisted of open-ended questions about what the subjects felt they had learned from the materials and how they thought the materials could be improved. Although the Untutored-PT group did not take any special training materials, they were asked to answer these questions as if they were evaluating the school materials they used in their regular history class.

Extensive changes had to be made in the idea-evaluation measures of the problems of KASKIA and PANJA after it was discovered that the initial open-ended questions were not leading to discriminably different responses among the subjects of the first two participating classes. Therefore, this performance measure was revised into a dichotomous choice with completion-answer format. This necessitated dropping the first two classes from the final analysis on this idea-evaluation measure, as initial statistical analyses did not show any compatibility between the results of the first two classes with the later four classes which would permit any meaningful pooling of data. The testing time for the other parts of the posttests remained the same after the revision since more time was added to the posttests to insure this compatibility.

The first two homeroom classes of subjects answered a more extensive number of specific history-problem questions which also did not give any discriminable or useful information. Two of these questions were retained in the posttest battery given to the other four classes but even these were not used in the final analysis. The deleted history questions were replaced by other questions about the training materials which, it was hoped, would provide useful information for developing future computer-training materials. However, these questions too did not enter into the final statistical analysis.

5. COMPUTEST: ITS BASIC DESIGN AND USAGE

As stated in the introduction, one of the primary purposes of this thesis is to examine the efficiency of developing computer-assisted instructional (CAI) materials for the training of problem-solving skills-- this, as we have seen in the section on author languages, is directly related to the computer author language being used. This thesis was proposed, in fact, as an actual test of whether someone without experience or knowledge of computer programming (i.e., the author) could design, develop and code training materials for productive thinking using the COMPUTEST language. Since this technological aspect is such an important part of the thesis, it is felt appropriate to include here a short section on how actually to use COMPUTEST in its simplest basic form, a form which can be understood by persons who are interested in the development of CAI programs but have had no programming experience.

COMPUTEST is a computer source language that acts as an interpretive translator of the simplified COMPUTEST code letters which are actually used by the person writing a CAI program. The entire COMPUTEST translator deck (which is itself written in another computer language called SPS) is stored in the computer memory bank and various segments of it are used temporarily to interpret the simplified COMPUTEST code letters that are contained in the CAI programs. This interpretation consists of translating the COMPUTEST code letters into meaningful sequences of machine commands for the computer to follow as the subject works on the CAI program on the typewriter. Thus there are two computer programs involved in the use of the COMPUTEST system: (1) the COMPUTEST translator deck which is stored in the computer memory and remains unchanged and which interprets the simplified COMPUTEST code letters into meaningful sequences of machine commands; (2) the CAI program actually written by an instructor which uses the simplified COMPUTEST code letters with the verbal text so that in the course of the program the student can appropriately be given information, asked questions, have his answers evaluated, and be branched into a selected future segment of the program.

The central feature of COMPUTEST is that the technicalities of giving the computer the necessary commands to perform instructional functions have been minimized so that the person need only use extremely simple code letters to specify what he wants to do in his CAI program. Persons untrained in computer programming are able successfully to write basic programs after only a few minutes instruction. Certainly this fulfills Zinn's (1967) first criterion of user feasibility.

In order to write CAI programs which can successfully question and answer a student in an interactive situation, the instructor must have a computer author language which can handle four essential main functions. These are:

(1) The presentation of text information and questions to the student.

(2) The recognition of the answers which the student submits back to the computer.

(3) The scoring and tabulating of these answers so that future decisions can utilize this past information.

(4) The branching of the student into the optimal segment of the program so that the student receives the most appropriate feedback which is based not only on his present answer, but also his appropriate prior response patterns.

COMPUTEST is organized to provide the necessary simplified code letters which will enable the instructor to write CAI programs that fulfill all of these four functions to their full capabilities. The actual COMPUTEST code letters will now be examined, but first, to aid in the understanding of programming format, it is noted that only one COMPUTEST code letter is used for each IBM card and it is punched in column 1. The information, comments, questions, correct answers, etc., which the instructor wants in the CAI program are punched in columns 7 to 72; columns 3 to 6 are used for entry labels for branching within the program.

First will be given the simplest sequence of COMPUTEST code letters which can be used to form a question and answer sequence. The underlined letter in the margin is the COMPUTEST code letter which is punched in column 1 and which informs the computer via the interpretive function of the stored COMPUTEST translator deck what to do with the programmed material punched in columns 7 to 72 on the same card.

<u>C</u> Whatever is punched in columns 7 to 72 will be typed immediately in that sequence as output from the computer.

<u>R</u> Whatever is punched in columns 7 to 72 will be stored as the right answer. This card contains the information which the computer uses in judging whether the input from the student is correct (i.e., matching) or not.

<u>G</u> Good comment information is typed back only if a right (matching) answer is found in the input.

<u>B</u> Bad comment information is typed back only if the computer fails to find a matching answer in the input.

<u>A</u> Accept an answer from the typewriter. This card can contain a variety of answer recognition options and scoring options which instruct the computer to compare and score the input in various ways. (These will be explained in more detail below as they form the core of the recognition and scoring procedures.) Also this card constitutes the end of a program segment so that the computer will not proceed any further in the CAI program until the student types something at the typewriter.

These code letters form the core of the COMPUTEST system. TABLE 1 shows an actual program segment written in this fashion by an elementary school child with an accompanying printout of how it appears when taken by another student.

The fact that children in elementary school have been able to write programs basically composed around this framework which provide valid learning situations for their classmates (Starkweather, 1965; Hodge, 1966) clearly demonstrates the ease with which the COMPUTEST system can be initially used.

It is obvious from the foregoing that the COMPUTEST language amply fulfills its stated goal of user simplicity both in theory and in actual practice. As noted before, Zinn (1967) considers this to be one of the two essential features of any author language which is to be suitable for developing CAI programs. But, what of Zinn's second essential of a computer language--"...the experienced author should be able to use the capabilities of the computer to the fullest, for as complex a procedure as he can construct" (1967, p. 81)? It is clear that if the code letters listed above were the only COMPUTEST commands one could use, the system would not be effective in fulfilling the four required functions for writing CAI programs: (1) appropriate text presentation, (2) sophisticated answer recognition procedures, (3) elaborate scoring techniques, and (4) complex branching arrangements. However COMPUTEST, of course, contains other commands and options which go far beyond the restrictions of the simplified skeletal system illustrated above and do permit highly sophisticated programming techniques. Initial attempts had been made (Starkweather, et al, 1967) to develop more sophisticated programs, such as sample interviewing situations with clinical patients, but no extensive systematic attempt had been made to develop a series of computer-assisted instructional programs which would be sophisticated enough to result in a valid learning experience for students. This was the second primary goal of this thesis-- to see if it was feasible to use the COMPUTEST language for developing CAI programming materials which would be sophisticated enough actually to train for productive problem-solving skills. In theory, the COMPUTEST language was sufficiently developed to handle such sophisticated programming, but it had not yet been tried out in a systematic series of programs before this thesis.

It would be beyond the intent of this section to attempt to explain or give examples of all the other COMPUTEST code letters and options available to the user, since the main purpose is to acquaint the programming-naive reader with a simplified general rationale behind the COMPUTEST language. However, to give the reader a feel for the possibilities which are available to implement the essential four functions of any computer author language which are listed above, a brief outline of some of the other COMPUTEST code letters and options are listed. The reader is advised to consult the COMPUTEST User's Manual, Appendix F, for the full list of programming code letters and options which exist in the COMPUTEST system and which are illustrated by appropriate examples. Now we shall consider, in turn, each of the four functions which a computer language must fulfill in order to permit the user to write sophisticated CAI programs.

(1) The presentation of text information and questions is handled through the C, G, and B code letters mentioned above. In actual writing of more complex programs, most of the textual information is presented on C cards with the appropriate C card segment being chosen and branched to by means of other evaluative and branching commands executed on the student's answer.

(2) The recognition of correct or matching responses from the student's answers can be handled in many ways with varying degrees of exactness. This is possible through the "rescan" option which permits the student's answer to be re-evaluated as many times as is necessary using different recognition techniques until an appropriate matching answer is found. Thus the author can start with a very stringent test of the student's answer, and if that fails

to find a matching response, he can re-evaluate the answer again with less exacting recognition procedures until a matching response is found and then branch to an appropriate feedback response which takes account of the exactness of the match that has been made. For example, various recognition options can be used to evaluate the response to the question, "Who is the most famous rabbit in the world?" We could make the most stringent test with the "string" option which packs the entire answer together and would only consider the response "BUGSBUNNY" correct. If that failed, a "match" option would consider the answer "BUGS BUNNY" correct if it found these two words in that order in the answer. If a matching answer was still not found, then a "group" option would consider "BUGS BUNNY" correct if these two words were found in any order in the answer. If that also failed, the single word "BUGS" could then be used as the correct answer (if the student only put down "BUNNY", he might be talking about the Easter one). If that failed we could use the "compare" option to test for the first three characters to see if any match with "BUG" (since the student might have spelled it "BUGGS"). If even that failed, then we type out the response, "You sure don't know your American heritage very well!" and proceed to the next question. The reader can get the flavor of the possibilities that exist for sophisticated recognition procedures.

(3) The scoring and tabulating of these answers are handled by "tally" and "mark" options which can each keep active cumulated scores in twenty-six different positions. The "mark" option score is either on or off, while the "tally" option can keep any score in each position ranging from -99 to +99. The possibilities that exist for elaborate and differential weighting of responses are readily apparent.

(4) The branching of the student to different segments of the program is handled in a basic way by the "jump", "rjump", and "wjump" options if the instructor wants to consider only the present answer for branching. However, much more complicated and sophisticated branching techniques are possible which take into account not only the present answer, but also the appropriate accumulated scores of other past responses. These branching possibilities are handled through means of T cards which have the options to test the designated "tally" and "mark" positions to see if they are equal to, less than, exceed, etc., a designated preset value, and based on these decisions, branch the student to a prescribed program segment for the most appropriate feedback.

This is a brief, oversimplified account of the other options and code letter commands which are available in the COMPUTEST language serving to fulfill the four functions of a computer language and permitting the user to develop sophisticated CAI programs which can utilize the full capabilities of the computer. Although the actual number of code commands and options is not large, the various combinations, patterns, and sequences of these codes and options give the user full flexibility and possibilities for developing sophisticated CAI programs which are limited only by his ingenuity.

In theory, it appeared that COMPUTEST gave the instructor full flexibility and possibilities for programming CAI materials. However, it was a moot point whether, in fact, this was true for the type of CAI problem-solving programs which were to be developed in the course of this thesis. It was anticipated that the programming of this type of training materials would raise some very sophisticated computer coding problems in each of the four necessary coding functions of a computer author language (as described more fully above), especially in the area of input-recognition of the students' responses to open-ended questions. The input-recognition constituted the central coding problem of these materials since these computer training programs were expressly to be developed so that the computer subjects would have an interactive-game training situation. This required that these subjects be allowed to respond freely in their own words during their interaction with the computer training programs.

This was the second goal of the thesis-- to examine the efficiency of the COMPUTEST author language in developing CAI programs for productive thinking which could sufficiently fulfill the four necessary functions of sophisticated CAI programming: (1) appropriate text presentation, (2) sophisticated answer recognition, (3) elaborate scoring techniques, (4) complex branching arrangements.

6. APPARATUS

The equipment used for the training procedures in this experiment consisted of two different systems: (1) the actual IBM 1620 Computer system which was used for the presentation of the Computer Programmed History materials, and (2) the computer-simulated Teletype system which was used for the presentation of the "Computer" General Problem Solving materials.

The IBM 1620 computer system consisted of the following equipment:

IBM 1620 Data Processing Unit, Model 1, 20k core, with console typewriter

IBM 1311 Disk Drive Unit, Model 3

IBM 1622 Card Read/Punch Unit

These three units formed the computer instructional system for the project. The CAI programs were punched on cards and then processed through the card read/punch unit to be stored in the disk drive unit. The input/output channel was through the IBM computer console typewriter.

The computer-simulated Teletype unit consisted of the following equipment:

Teletype Corp. -- Model 15 KSR Page Printer

Teletype Corp. -- Model 14 Tape Distributor

Teletype Corp. -- Model 14 KSR Tape/Punch

The student worked at the keyboard of the Teletype page printer which was controlled by the prepunched paper tape on the connected tape distributor. Students were told that this was a second typewriter terminal which had been connected to the computer for this study. Elaborate care was taken to foster this deception by having similar initial computer responses printed, a similar feedback output format, appropriate time delays before printing the feedback, etc. Apparently the deception worked well, as no subject gave any indication in the course of the experiment that he realized the terminal was not connected to the computer (even though three of the subjects in the Combined-Computer group, who also worked on the real computer, did comment on the simple feedback from the simulated computer).

The two equipment systems were located in a special classroom which had been assigned solely to the Dixie Computer Project (see below) for its two year duration. The two typewriter terminals were situated in two corners of the room and each was enclosed within a specially built three-sided wooden cubicle which was five feet high and extended four feet in the back of the student. The wall in back of the typewriter which faced the subject had a thin 1/16 inch opaque plexiglass screen 15 by 20 inches which served as the rearview projection screen for the random-access slide projector (Kodak Model AV900). (This had been modified into a uni-directional random access projector unit by the Decision Systems Corporation.)

7. SUBJECTS

The Dixie School District of Marin County in California had been chosen as the site of the Dixie Computer Project prior to this thesis because of the initial enthusiasm of the school administration about the goals of the computer project and the fact that Dr. John Starkweather, the director of the computer project, was also an elected trustee of the school district. The Vallecito Junior High School was chosen as the actual school which would house the project for three reasons: One was the initial desire of the computer project to work with junior high school students, another was the central location of the Vallecito school to the other district schools, and the third was the availability in that school of classroom space to house the computer project for two years.

There were four factors which led to the choice of the eighth grade class in the Vallecito school as the pool of subjects for this thesis. First was the immediate accessibility of the students to the computer facility which eliminated the almost insurmountable problems which would have developed if students had to be transported on a systematic basis from other district schools. The second factor was the prior decision to develop the materials within a social science context; the eighth grade curriculum included an American History course for both semesters which offered the best hope of implementing this feature of the thesis over the seventh grade curriculum. A third important factor was the fact that the eighth grade students had already become familiar with the computer in the previous year when they informally participated in various short computer programs which were written by interested classmates. This was considered important in order to lessen any initial "Hawthorne" novelty effect which one would expect to find prevalent in the seventh grade students who had not yet worked on the computer. The fourth factor was that the principal of the Vallecito school felt the administrative logistics of the student scheduling, which would arise over the extended period of time of the experiment, could best be dealt with by using the eighth grade class.

The eighth grade was composed of eight homeroom classes each containing approximately an equal number of boys and girls within each class, and within each sex of each class there was an approximately equal number of students above and below the eighth grade mean IQ of 112. The IQ scores were based on the California Test of Mental Maturity (Form S, 1957) which had been administered to the entire class that year. The initial statistical analysis of the IQ scores of the entire eighth grade showed that the distribution approximated a normal distribution save for a deviant cluster of scores at the low end of the scale which produced a bimodal distribution. The eighth grade IQ mean was 112.0 with a standard deviation of 16.0. In order to approximate a normal distribution more closely, it was decided to exclude those students from the original pool of subjects who fell outside the 99% interval of IQ scores which ranged from 70.8 to 153.3. Therefore ten students who fell below the lower limit (no student was above the upper limit) were not included in the original pool of subjects for this thesis. The overall mean of the remaining students was now 112.9, with a mean of 112.0 for the males and 113.8 for the females.

Due to the fortuitous district assignment of students to homeroom classes based on sex and IQ, it was possible to use the homeroom class as a separate factor for the statistical analysis. This was considered highly desirable and important for the reduction of the error variance by controlling for the extraneous effects of the different homeroom teachers and other factors which would tend to affect the classes as a whole. The decision to utilize the classroom as a block to control for experimental error, rather than pick students at random from the entire eighth grade, led to the further decision randomly to select six of the eight classes for participation in the experiment. Six classes were chosen since this would give the appropriate number of subjects to fulfill the experimental design matrix which had been arrived at according to the limiting experimental factors already discussed: (1) limited training time available, (2) availability of only one computer terminal, (3) number of subjects required per statistical cell, and (4) number of training lessons. Each class was divided into a completely crossed 7 x 2 x 2 matrix of seven training groups, two categories of sex, and two levels of IQ-- above and below the adjusted class mean of 113. Thus there were 28 students from each class who were chosen to participate in the experiment-- one student in each of the 28 (7 x 2 x 2) statistical cells. A diagram of this experimental design is seen in TABLE 2.

Although the below-average IQ group of subjects is of course lower than the high IQ subject group, it is felt that the term "Low IQ" group might be misleading to the reader since the mean of this below-class-mean group is not "low" in absolute terms, being 104.4, with the median even higher at 108.0. For sake of greater compatibility with the usage of the terms low, middle, and high IQ in other studies, it is felt that the term "Middle IQ" group would be a more appropriate designation of these below-class-mean IQ subjects, while the term "High IQ" would be a fit designation for the above mean IQ students since they have a mean of 124.4 (and a median of 123.0). This use of the terms "Middle IQ" and "High IQ" would be in close compatibility with the terminology of the Crutchfield and Covington studies (Covington, 1965; Covington and Crutchfield, 1965) in which the subjects were divided into "low", "middle", and "high" IQ groups with the respective means of 91, 107, and 123.

The assignment of students within each class to the seven different groups followed a special fixed procedure designed to lessen the possibility of excessive inequality of IQ score distributions between the seven training conditions which might occur if the subjects were assigned solely by chance. Within each class (excluding the students below the lower limit of 70.8) seven students of each sex below the mean IQ of 113, and seven of each sex above the mean were randomly chosen as the 28 students from that class who would take part in the experiment. Next, the low IQ students of each sex, as well as the high IQ students of each sex, were rank-ordered according to their IQ and given a rank-order score from 1 to 7. The seven subjects in each of the four blocks of Male Low IQ, Male High IQ, Female Low IQ, Female High IQ were then randomly assigned to the seven training conditions, one student to each training condition. The method that was arrived at to control for the inequality of the IQ distribution in each of the seven training conditions was to add the rank scores of all four subjects in that training condition and see if the total sum fell within the arbitrarily designated limits of 14 to 19. Necessary adjustments were made for those training groups whose total sum fell outside these limits by switching students from those training groups which had too high a total sum to those which were too low, until the seven training groups for that class had acceptable total sum rank scores.

The optimal procedure would have been to assign the students from all the six classes at once so that not only the total sum rank score of the different training groups could be equalized for each class, but also over the entire six classes. Unfortunately the problem of minor class changes of students, as well as the moving of some students from the school over the three month period did not permit such a procedure to be strictly followed.

Compounding this problem was the inevitable slight differences in the class IQ scores which precluded a strict, complete overall procedure based on ranking of these scores. TABLE 3 shows the median IQ scores for the six classes on the basis of sex and IQ level and the overall median for the entire class.

The actual median IQ scores for each of the six planned comparisons involving different combinations of the training groups are listed in TABLE 4. Nonparametric median statistical tests (Siegel, 1956) were performed between each of the two groups of the six planned comparisons for Low IQ, High IQ and All Subjects partitions. The results of these are found in TABLE 5. Two of the tests indicate a statistically significant difference in the IQ median scores for the High IQ subjects between the two groups of the planned comparison; however, it should be noted that in both cases it is the group which would be expected to perform better that has the lower IQ median. This then would not weaken any interpretation which might be placed on these results as it actually causes a more stringent test of the hypothesis involved.

The Dixie School District is located in a suburban housing development environment twenty miles north of San Francisco. The general socio-economic conditions are characteristic of a middle to upper-middle semi-professional and professional class. All of the 168 students involved in the experiment were Caucasians, except for three Oriental students and one Negro student.

8. SCHEDULING OF SUBJECTS

A. General Plan

Even though the consideration of the various limiting factors had resolved itself into the experimental design choice of 168 subjects, the problem remained of exactly how to schedule all these subjects for their training sessions and posttests. The crucial factor here was in keeping the class homeroom as a block factor for partitioning and reducing the error variance; therefore the procedural question became how many classes to run simultaneously. To run all six classes at once would have meant that each subject would only work on his training materials once a week; this obviously would not have been sufficient to achieve a continuity of the materials, or to bring about a sense of participation in the experiment. On the other hand, if only one class was run at a time, the subjects in the Combined-Computer group would have had to work on materials twice a day, and the pressure from absences, equipment failure, and slower students would have made the time factor too critical to insure the successful running of the experiment.

It was finally decided to run two of the six classes at a time. Subjects of the Combined-Computer group would be working on their materials each day, while the other computer subjects would be working on their training materials every other day.

It was felt that the time period was short enough to insure continuity and a sense of participation in the study, and also that the time was long enough so that student absences, slower students, and any computer breakdown could be feasibly handled.

The general format for conducting the experiment was for two classes to be randomly chosen from the pool of eight or fewer remaining classes. These students were then assigned to the various training groups by the method described above and worked on their respective training materials over the next two and a half week period. The scheduling method was flexible enough so that all subjects finished working on their materials in the last two days of the training period. The two posttests were then given to each of the two classes as a whole over the next two days, one posttest on each day. The next two classes were then randomly chosen and followed the same procedure, and then finally the last two chosen classes participated.

B. Scheduling of Computer Subjects

The cooperation of the teachers involved made it possible to arrange a very flexible method of scheduling subjects simply by posting a daily list of subjects who were to come to the computer room and of periods they were to come. During each class period two students were released from their regular class to work at the computer room; one worked on the real computer, the other on the simulated one. This method turned out very well with only a few students forgetting their appointments; those that did forget were called over the intercom system within three minutes after the period started. This flexible scheduling method permitted finding replacements for absent subjects, fast makeup lessons for those who had missed lessons, and the scheduling of more periods for the slower students on the basis of their actual ongoing performance so that all students finished their training materials in the last two days before the posttests were given.

Teachers were notified which students would be in the computer room for the different periods of that day by means of the school's daily mimeographed teacher's bulletin which they received before the second period. Any teacher who did not want a particular student out of class simply notified the computer room and the necessary changes were made; in fact, fewer than ten changes were made in the entire experiment. Similarly, if a student had a valid reason to have his scheduled period changed for that day, it was arranged. However, to alleviate this problem beforehand, the schedule was prearranged so that each subject would not miss more than one class in the same school subject over his three-week period; thus it turned out only four student requests were made for schedule changes.

C. Scheduling of Booklet Subjects

Although it would have been the ideal to have the booklet subjects on a similar schedule as the computer subjects in order to equate for various incidental factors, such as individual release from class, the logistics of the school situation did not permit this scheduling method. For one thing, the noise distractions of the computer room precluded having subjects working on the booklets there at the same time that other students were working on the computer. For another thing, there was no other classroom space regularly available where students working on booklets could be scheduled throughout the day. The only feasible method, therefore, was to have all the students who were working on booklet materials come to a designated classroom twice a week to work individually on their respective training materials. Individual makeup times were scheduled for absent students.

Taking the booklet subjects out of their regular homeroom class as a group twice a week also served another purpose of allowing two special study periods each week for the computer subjects who remained in the homeroom. The computer subjects were required to make up any school subject lessons they missed when working in the computer room even though they had been excused from that class. These special study periods were considered an important experimental feature designed to compensate the computer subjects and to prevent possible complaints from them.

9. ORIENTATION TALKS

Three general orientation talks were given before the experiment began: (1) a talk with the teachers to explain the general outline of the experiment and to arrange the scheduling of the students; (2) a talk with all students to explain to them in general terms what was going to take place and to request their cooperation in the study; (3) a talk with the computer subjects to explain how to operate the different mechanical devices. A more detailed description of each talk follows.

A. Teacher Orientation Talk

Two weeks before the winter vacation began, a meeting was held with all the eighth grade teachers to explain to them that the study, which would begin the week following resumption of classes after the holidays, was concerned with examining different types of booklet and computer training materials designed to increase performance of productive thinking skills. The teachers were told that the author would give a general orientation talk to the students during the first week of resumed classes and requested the teachers not to discuss the experiment with their students, and to refer any questions which the students might have about the study during the course of the experiment to the author. General procedural questions which the teachers had were then answered.

B. Student Orientation Talks

A week before the actual beginning of the experiment, a talk was given by the author within each of the eighth grade classes explaining to the students in a very general way what was going to happen over the next three months. This talk was considered quite important in gaining the students' confidence and cooperation since the school was not a university experimental school and the students had not taken part before in any systematic large-scale experimental study. They were told that the project was interested in examining how effective different types of training materials were in teaching them to think better. It was clearly stated that it was the materials that would be evaluated and not the students. They were assured that the results would not go down on their report card or any other official record; however, to give them the feeling the materials had some relevancy to their scholastic standing, thus to insure a positive motivational set, the students were told that their final results would be shown to their homeroom teacher in an unofficial way.

The general nature of an experimental study was briefly covered in explaining to them that different students would be chosen at random to work on several different types of training lessons and that the amounts the various groups learned from taking these materials would be compared at the end of the study.

Since this study obviously was directly connected with the computer project, it was anticipated that all the students would expect to work on the computer. To prevent the feeling among the booklet students that they were being "cheated" of work on the computer, it was stressed that those who did start work at the computer had been selected by chance, and that the students who were working on other materials would have their chance later on to work on the computer.

Students were allowed to ask any general questions about the study as a whole. They were then requested not to talk about or compare the different types of materials that they would receive, but simply to work on their own materials as well as they could. Finally they were told that more specific instructions concerning the schedule and other relevant information would be given to them when it came their turn to participate in the experiment.

C. Computer Subjects Talk

The instructing of the computer subjects in the use of the computer and Teletype equipment as well as the random-access slide projector was done in a preliminary 40 minute session with special sample materials. This pretraining on the equipment was given to the first two classes only since it was found that the operation of the computer typewriter and random-access slide projector panel was quite simple and obvious. What was needed, however, was a way to call the student's attention to several mechanical details, such as the difference between the letter "o" and zero on the keyboard. This was accomplished by having printed instructions taped to the appropriate trouble spots of the equipment illustrating the correct method of operation.

Instead of having a group demonstration on sample materials for the other classes, the students were instructed individually when they came in for their first lesson on the computer or Teletype. The author was always available to help the students if they had misunderstood the instructions and were having mechanical problems, but this assistance was only occasionally needed during the first lesson, and in fact was never requested by any student after the second lesson. At the eighth grade level, typing did not present any problem for the students, and even those who had never typed before were able to perform sufficiently well so that no outside help was needed.

10. PROCEDURES IN RUNNING SUBJECTS

A. Computer Subjects

In the first session, the subject received the mechanical instructions on the operation of the typewriter and slide projector as he began work on his respective training materials-- either the programmed history materials on the IBM computer, or the General Problem Solving materials on the computer-simulated Teletype. The subject continued working on his materials independently and at his own pace until the end of the forty minute class period. The typewriter sheet which contained both the computer output of the program and the subject's input responses was collected at the end of the period and filed in the subject's individual progress folder. The next time the subject came in, he received this print-out sheet from his last session and the computer program or Teletype tape, as well as the slide projector, had been previously set up so that he could continue where he had left off in the program. When the subject finished one program he began working on the next one, unless it was within five minutes of the end of the period in which case he waited until the next session. Individual performance record sheets, as well as material folders, were kept on all students as they worked through the training materials to insure satisfactory completion of all training materials and to allow for the scheduling of extra work periods for the slower subjects based on their actual ongoing performance records.

The running of the computer subjects through the training materials went very much as planned except that time pressure was more critical than had been anticipated. Although the original intention had been to run subjects only during class periods and not during lunch period or after school, it became necessary to ask willing students to volunteer to work on their materials during one or two lunch periods during their three week participation. Occasionally, a few students volunteered to work after school for a half hour to an hour on their training materials. No student was in any way required to do this, but it was considered advisable to have those willing students come in during lunch and after school rather than to request the slower students to work faster or to shorten the lessons. Each student, no matter how slowly he went, worked at his own rate of speed. Owing to the flexible scheduling method, students received enough work periods so that they all finished at least through the first half of the final lesson of their respective materials before the posttests were given to the class.

The author was always available in the computer room to assist in remedying any mechanical or computer problem or breakdown. Only two minor computer breakdowns actually did occur in the three months of computer operation and both of these were quickly repaired with a loss of only seven class periods in the entire experiment.

B. Booklet Subjects

The booklet subjects came to the designated classroom as a group and each subject picked out his individual folder containing the training materials to be individually worked for that period. One lesson unit was planned for each period and those students who finished early did not start a new lesson, but were advised to reread the present materials. Those subjects who did not finish within the period continued on the materials in the next session before starting the succeeding lesson. The subjects wrote their answers to the various questions on separate answer-sheets. All materials were collected at the end of each period. The answer-sheets were checked over and recorded on the individual performance records before the next session to insure that the subject had put down some form of answer to each of the questions; those subjects who had not finished all the questions were required to complete them before starting on a new lesson.

Slower subjects and those who had been absent were individually scheduled to complete any unfinished materials before the posttests were given.

C. Posttests

The posttest materials consisted of pencil-and-paper tests personally administered by the author to each of the homeroom classes as a whole, with the homeroom teacher being present to assist in the distribution and collection of the test materials. Since each posttest required fifty minutes of working time, this necessitated a prearranged carryover of ten minutes into the following class period. The posttests of the first day were composed of two performance problems-- WHITE HORSE INDIANS (30 minutes) and KASKIA (20 minutes); the second day consisted of the Attitude Questionnaire (25 minutes) and other performance problem -- PANJA (25 minutes). The three performance problems were timed tests with three minutes working time generally being allotted per page. Subjects were allowed only to work on the current page and were not permitted to go ahead or go back. When one half minute was left at the end of the three minute period per page, the author told the students to finish up what they were writing. At the beginning of each problem the students were told that complete sentences were not necessary and only a few words would be sufficient if they described the idea plainly enough. They were told that the important thing was to put down all the ideas they had without being overly concerned about the spelling or grammar; in the last half minute of each time period those students who were still working were told just to put down a few words to complete their idea or ideas. The questionnaire was not timed and the students were allowed to work on the entire section as a unit; however, the elapsed time was reported at appropriate intervals and indications were given as to approximately how far the subject should be so as to insure that all subjects finished within the allotted time.

Although two classes were being run at a time, each class was separately tested in its own homeroom each day of the posttests. Makeup tests were arranged as soon as possible for the absent students with the attempt made to administer the makeup tests in a group test situation of all the absentee students.

III. SCORING AND STATISTICAL PROCEDURES

1. SCORING PROCEDURES

The first step in the preparation of the posttest data for scoring was the assignment by a secretary of random numbers to all the subjects. These code numbers were unknown to the author, who was the sole scorer of the data, in order to insure a blind impartial scoring of the results. Next, all the subjects' responses were keypunched verbatim onto IBM cards with only the identification of the question and the subjects' code number. The format of the computer printout sheets from which the actual scoring was done was arranged so that each posttest question was separately printed with all the 168 subjects' responses in the numerical order of their code numbers.

In the actual scoring of the posttests, each question was worked on at one time for the entire set of responses of all 168 subjects. All responses were read and a tentative list of criterion categories was set up based both on the empirical data and on the conceptualization of the problem-solving functions that question was intended to test. The entire set of responses for that question was then reread and this time each response was scored in a tentative category; any response which did not fit into the basic list of existing categories was considered a new category and added to the list. After this second examination of the question, the basic list with its new categories was revised to eliminate superfluous categories and to make the final remaining ones theoretically more distinct and meaningful. Finally, the entire set of responses for the question was read a third time and rerated using this final category list.

The data fell into two major conceptual divisions, each containing three main variable groupings:

The three performance measures of: (A) problem-formulation, (B) idea-generation, and (C) idea-evaluation.

The three attitude measures of: (D) problem-evaluation, (E) training material evaluation, and (F) self-evaluation.

The scoring procedures which were developed and used in the final statistical analyses for each of these measures will now be examined in more detail.

A. Problem-Formulation

Owing to the differences in open-ended questions asked about the problem-formulation in each of the three performance tests, it was not possible to use the same scoring procedure on all three tests; instead, specific measures were arrived at for each problem on this performance variable. The final measures were: WHI-2,expression of puzzlement and interest in the story (the number, in this instance of the WHITE HORSE INDIANS problem, refers to the posttest page number on which the posttest question is found); WHI-3, total number and quality of puzzling facts; KASKIA-II, total number, quality, and originality of information-seeking questions (see below for originality procedure); PANJA-8, listing of efficient information-collecting strategies and total number of specific information-seeking questions.

B. Idea-Generation

Three common measures of quantity, quality, and originality were used to score the following questions in the three problems concerned with idea-generation: WHI - 4 & 5, 6 & 7; KASKIA - 12 & 14; PANJA - 9 & 10.

The quantity measure for each question was the sum of the different categories that the responses of the subject fell into for that particular question.

The quality measure was the sum of the quality ratings that had been assigned to each of the response categories within the question. A two-point quality scale was used in evaluating WHI, a three-point scale for KASKIA; the response categories of PANJA did not lend themselves to any differential quality ratings (see below).

A quality rating of two for WHI questions was assigned to those response categories which gave both a sufficient motivating force for the Indians' behavior and a sufficient explanation for the suddeness of their actions; a quality rating of one was given if only one of the conditions was met. A quality rating of three for the KASKIA questions was given to those response categories which fulfilled the three following conditions: a sufficient reason for the disappearance, a sufficient explanation why no wagons or remains were ever found, and a sophisticated integration of the preceding two conditions in the answer. A quality rating of two was assigned to those response categories which fulfilled only the first two conditions, while a quality rating of one was given to the categories which only fulfilled the first or second condition. All the response categories for the PANJA question were given a value of two since it turned out that all the response categories (except two) fulfilled the scoring conditions of both a sufficient reason for the war, and a sufficient reason for the suddeness of the violence.

Originality ratings for each of the response categories were based on the infrequency of subjects responding in that category. Since in each problem two of the questions formed a set which was basically the same question, the tabulating of the frequencies was summed over both the questions in the set, namely, WHI-4 & 5, WHI-6 & 7, KASKIA-12 & 14, and PANJA-9 & 10. As the posttests had been especially developed for this thesis, no already established comparison norms were available; therefore, the originality measures were empirically based on the preliminary analysis of the data. Graphs were plotted for each question based on the total number of subjects who used each of the response categories. Since there appeared to be a break in the distribution of the data around the 10% point, it was decided to consider as original only those response categories which had been put down by fewer than 10% of the 168 subjects. Consequently, the following numerical rating scale was arbitrarily devised: response categories used by 0-2% of the subjects were given an originality rating of 5; 3-4% a rating of 4; 5-6% a rating of 3; 7-8% a rating of 2; 9-10% a rating of 1. Two originality measures were used in the final statistical analyses: the sum of the originality ratings, and the total number of response categories having a rating of 4 or 5.

C. Idea-Evaluation

The same method of scoring all the posttest questions concerned with idea-evaluation was used for the three problems: WHI-9, KASKIA-16 & 17, PANJA-12 & 13, PANJA-14, PANJA-15. This method was to categorize the subjects' evaluation of the possible solution-ideas listed in the posttest problems on the basis of whether their idea evaluations: (1) made full use of the given problem-information, (2) made only partial use of the problem-information, (3) completely disregarded the problem-information. (Note: As discussed more fully above, only the data of the subjects in the last four classes are utilized in this idea-evaluation performance measure.)

D. Problem-Evaluation

It will be remembered that at the end of each of the performance problems (WHI-10, KASKIA-18, PANJA-16) the subjects rated each problem on the five dimensions of: easy, fair, like, interesting, fun. The format of these posttest questions required the subject first to circle the one of the multiple choices on each of the five dimensions which best described his attitude toward the entire performance problem. Then he filled in an accompanying completion sentence for each dimension in order to explain his particular choice. Accordingly, two types of scoring procedures were used on the data. In the first procedure, after the multiple choices had been converted into numerical scores, the mean scores which the subjects made on each of the five attitude dimensions for each of the three problems were tabulated. In the second scoring procedure the reasons the subject gave for his rating of each of the five dimensions were coded into three categories: (1) those based on the problem-solving nature of the posttest problems, (2) those based on the facts or story content of the problems, (3) those based on the interest and excitement in working on the problems.

E. Training Materials Evaluation

The data for these questions in the questionnaire (Que-3, 4, 5, 6) fell into two main groupings: the attitude rating scale of the training materials (Que-6); and the open-ended questions dealing with the subjects' evaluation of the training materials (Que-3, 4, 5). The rating scale was scored by the same procedures used in the Problem-Evaluation section described above, namely, the tabulating of the mean scores for each of the five dimensions, and also the categorization of the reasons for these ratings using the same three categories as above.

The responses to the open-ended evaluation questions of the training materials, especially the two major questions on pages Que-3 and the bottom of Que-5, were categorized according to whether the subject answered: (1) that he learned how to solve problems, (2) that he learned some facts or a new story, (3) that he did not learn anything.

Since the main purpose of the questions on Que-4 and the top of Que-5 had been to give the author useful information concerning future revisions of the training materials, they were not included in the final composite variables. The two questions on Que-7 were intended to indicate the amount of the students' awareness of the utilization of problem-solving skills in the historian's work. However, since they did not show any meaningful differential results after a preliminary analysis, they were not included in the final analysis.

F. Self-Evaluation

The 27 items in the self-evaluation section on Que-1, 2, 3 were all in the form of dichotomous yes/no answers which the subject had circled for each item. These data were then converted into numerical scores (no=0, yes=1) for each item in preparation for the cluster and additive statistical analyses described below.

2. FINAL COMPOSITE VARIABLES DERIVED FROM CLUSTER ANALYSIS AND ADDITIVE PROCEDURES

It was now necessary to condense the separate data into relatively independent composite variables within each of the three performance and three attitude sections in order to simplify the vast amount of posttest information and hence to arrive at more meaningful interpretations of the results obtained. It was decided to approach this problem in two ways-- by using as the principal method the cluster analysis techniques[3] developed by Tryon (Tryon and Bailey, 1966) while also utilizing, when appropriate, composite additive scores to get at important significant information not fully conveyed by the derived cluster variables.

The procedure used in the method of cluster analysis was to make initial "open" cluster runs on all the relevant individual variables within each of the six performance and attitude sections. In this "open" run the cluster analysis program produced the initial set of clusters from the data which were based on the product-moment correlations among the items in each pool. This initial set of cluster variables was then examined and revised according to the summary pattern of results on four accompanying statistical measures of: (1) the raw intercorrelation matrix of the individual items within each composite cluster; (2) the reliability coefficient of the composite cluster definers (Spearman-Brown); (3) the reproducibility of the mean of the squared correlations among the individual items ($\bar{r}^2$) within each composite cluster; (4) the intercorrelations between the composite clusters.

After initial statistical redefining of the composite cluster variables, the clusters were further revised by the deletion or addition of individual items with the aim of achieving the most meaningful final composite clusters. The final selection of individual items based both on statistical and theoretical considerations was then subjected to the "preset" second cluster analysis resulting in the 16 final composite variables over the six performance and attitude sections. These cluster variables were produced as the unweighted composites of the standardized scores for the defining items with a mean of 50 and a standard deviation of 10.

As briefly mentioned above, it was felt that in some cases the composite cluster analysis variables did not completely convey all the information from the data which the author considered to be theoretically important in trying to assess the differential effects of the training materials. Accordingly, eight additive composite variables were also used in the final statistical analyses which offered either a different theoretical view of the data, as in the Idea-Generation and Evaluation of the Training Materials sections, or supported cluster composite variables which were statistically weak, namely, the Self-Evaluation cluster variables.

The following six sections report the item composition of the cluster and additive composite variables with accompanying statistical information and the overall interpretation of these variables.

A. Problem-Formulation

The two final clusters which were obtained for this performance measure (TABLE 6) were both problem specific. Cluster C1 was composed of the following three items from question WHI-3: (a) total number of puzzling facts, (b) total sum quality rating of the response categories, and (c) total number of responses with a quality score of two. (Note: The small letters enclosed in parentheses refer to the item letters found in the appropriate TABLES.) This cluster variable can be interpreted as indicating the skill in picking out odd and discrepant facts and clues from initial fragmentary problem-information.

The second cluster, C2, was composed of three items from the question KASKIA-11: (a) total quality of information-seeking questions, (b) total sum of originality scores for those questions, and (c) total number of questions with an originality rating of 4 or 5. This composite variable reflects not only the skill in asking relevant questions, but also the ability to ask original questions seeking unusual information which could lead to a unique solution to the problem, e.g., the question, "Was any war going on at the time the settlers disappeared?"

The test items relating to the problem-formulation question of PANJA-8 did not lead to any meaningful composite cluster either within themselves or with other items and were therefore not included in the final analysis.

B. Idea-Generation

The intercorrelations between total quantity of ideas and total quality of ideas for each of the seven idea-generation questions in the posttest were so high (above .90 in every case) that the initial cluster run produced seven trivial doublet clusters, each one centering around one of the questions. Since this did not lead to any type of theoretically meaningful compositing of the data,only the total quality rating for each of the seven questions and the total originality rating for each of the three problems were used for the final cluster runs.

Three final composite clusters (TABLE 7) were obtained for this performance measure. They were considered to be theoretically meaningful enough to be kept separate even though the statistical intercorrelation between the three clusters was slightly high. Each of the three problem-specific clusters contained the total sum of the quality ratings for each question in the problem as well as the total originality score for the problem. Cluster C3 of the WHI problem was comprised of the total quality scores of (a) WHI-4, (b) WHI-5, (c) WHI-6, (d) WHI-7, and (e) total originality score of the whole problem. Cluster C4 consisted of the total quality score of (a) KASKIA-12, (b) KASKIA-14, and (c) total originality score of the whole problem. Cluster C5 consisted of (a) the total quality score of PANJA-9 & 10, and (b) the total originality score of the problem.

The interpretation of these three composite variables is quite obvious, namely, that the total quality of ideas generated as well as their originality rating tend to be highly intercorrelated within each of the specific problems.

Three additive composite variables were also included in the final analysis of this performance measure to see if there was an overall general ability on the part of the subject to generate high quality ideas and original ideas regardless of the particular problems involved. While there was some indication from the intercorrelation matrix of such a general ability pertaining to quality of ideas, the intercorrelations among the total originality scores on all three problems were surprisingly low (.29 between WHI and KASKIA, .16 between WHI and PANJA, and .17 between KASKIA and PANJA). It appeared quite plain that these additive composite variables would give new nonredundant information about the data in this performance measure beyond the cluster composite variables. The first of these additive scores was variable A1, which comprised the total quality ratings of all the ideas generated in all three of the problems. Variable A2 was the total originality scores for all ideas in all three problems. Variable A3 contained the total number of ideas which had an originality rating of 4 or 5.

C. Idea-Evaluation

The two final clusters obtained (TABLE 8) cut across the specific problem boundaries and indicated a general consistency concerning the utilization of problem-information for the evaluation of ideas. Cluster C6 comprised the items indicating the total number of times the subject used the new information to evaluate the possibilities on (a) KASKIA-16 & 17, (b) PANJA-12 & 13, and (c) PANJA-14. Cluster C7 was based on the variables indicating the subject did not use any of the new information for evaluating the ideas in (a) KASKIA-16 & 17, (b) PANJA-12 & 13, and (c) PANJA-15. Although at first glance the two variables seem to give negatively redundant information, this is not the case, since these were not mutually exclusive categories.

D. Problem-Evaluation

Three of the four final cluster variables (TABLE 9) for this attitude measure were based on the rating scores and were specific to each of the three problems, while the fourth cluster had to do with the basis of these ratings being the problem-solving nature of the problem. Cluster C8 was specific to the WHI-10 problem and composed of the ratings on the factors of (a) interesting, (b) fun, and (c) like. Similarly cluster C9 was specific to the KASKIA-18 problem and was based on the same factors of (a) interesting, (b) fun, and (c) like. Variable C10 for PANJA-16 was also based on (a) interesting, (b) fun, and (c) like. The interpretation of these three clusters is straightforward; they represent an overall positive attitude toward working on these and similar problems. But this positive attitude, while showing some degree of intercorrelation between the three problems, is problem-specific to a surprising degree.

It is important to have a measure which indicates the extent to which the subjects have rated the posttest problems on the dimensions of: (a) interesting, (b) fun, (c) like, owing to the problem-solving nature of the posttest materials. Cluster C11 measures this.

E. Training Materials Evaluation

Two clusters (TABLE 10) similar in nature to the two types obtained in the Problem-Evaluation section above were finally arrived at. The first cluster C12 was based on the ratings put down on the rating form (Que-6) for the dimensions of (a) interesting, (b) fun, and (c) like, as well as (d) the negative category of dullness in response to the open-ended question of Que-3. The second composite variable C13 came from the two response categories of question Que-5 (bottom) which concerned the positive item of (a) the learning of thinking skills, and the negative item of (b) the learning of facts. The first cluster, C12, plainly indicates a general positive attitude toward the training materials and a willingness to work on future similar materials. The second cluster, C13, is important for showing the subjects' awareness about what they have learned from the different training materials. A high score on this variable indicates that the subjects are cognizant of the thinking skills being taught in the instructional materials.

However, beyond the information contained in C13, it was considered important for supporting the interpretation of the results to know he particular reasons the subjects put down for rating the training materials on each of the five dimensions listed on page Que-6. These categories indicating the reasons for the subjects' ratings had not been included in any of the final composite cluster variables; therefore, three theoretically important categories were made into three composite additive variables by totaling each of them over the five dimensions of easy, fair, interesting, fun, and like. The three resultant additive variables are: variable A4, the stated reason for the subjects' rating being the story-content or facts of the training materials; variable A5, the stated reason for the subjects' rating being the problem-solving nature of the materials; and variable A6, the stated reason for the subjects' rating being the interest in the materials.

F. Self-Evaluation

Three composite cluster variables (TABLE 11) derived from the 27 dichotomous item list on Que-1, 2 & 3 lend themselves very readily to theoretical interpretations. This was the major factor in including them as final composite cluster variables even though the various statistical criteria for their choice were generally weak. The rationale was that with a longer and more refined items inventory these clusters would become statistically more respectable; in any case, these three clusters show important tentative differences in the attitudes of the subjects resulting from the different training conditions.

In order to avoid the redundancy of writing out each inventory item, only the code letters will be given here; the interested reader can refer back to the posttest materials (Appendix E) to see the full inventory items. Cluster C14 was comprised of three variables, all from the first objective section (Que-1) of the self-evaluation form, namely, (a) item 4, (b) item 8, and (c) item 10. The other two composite cluster variables were entirely from the second subjective section (Que-2 & 3) of the self-evaluation form. The first of these clusters was C15 composed of (a) item 2, (b) item 3, and (c) item 16. The final cluster C16 was composed of five items: (a) item 5, (b) item 7, (c) item 8, (d) item 14, and (e) item 17.

As mentioned above these three composite variables lend themselves very readily to theoretical interpretations. C14 plainly can be defined as the student's belief that the best approach to working on a problem is to get one main idea and stick with it. Dealing now with the more subjective measures of attitude, C15 indicates an attitude of perseverance on the part of the subject to continue working on problems even if they are difficult, while C16 shows the self-confidence the subject has in his own problem-solving abilities.

As mentioned above, two additive composite variables were included in order to support the statistically below-average cluster composite variables obtained. Since the self-evaluation inventory was divided into an objective and a subjective part, and since in the cluster analysis runs none of the items from the first part clustered with the second part, it was decided to make an additive composite score for each part rather than have just one grand-total variable. Even though the second part of the inventory was subjective, each item had a "correct" answer that would be expected to be chosen by subjects who were good problem solvers. Therefore the two additive variables simply are the sum total of the correct answers for each of the two parts of the inventory. Variable A7 is the total number of correct answers for the ten-item objective part of the inventory on page Que-1, and variable A8, the total number of "correct" answers for the seventeen-item subjective part of the inventory on pages Que-2 & 3.

3. STATISTICAL PROCEDURES FOR ANALYSES OF RESULTS

There were two major statistical problems involved in this thesis. The first, as discussed above, had to do with the reduction of the data into meaningful composite variables through appropriate statistical methods; this was handled by means of the Tryon Cluster Analysis procedure. The second problem concerned the appropriate statistical method to be used in testing for differences in results between the various training groups.

A. Planned Comparisons

The experimental design of completely-crossed factors of training-group, sex, and IQ within each of the homeroom classes permitted the basic analysis of variance design to be used. However, the particular experimental questions which this thesis set out to answer permitted a much more powerful and appropriate statistical analysis to be made, rather than simply testing to see if there were any difference among all the group means by an overall F-statistic of the analysis of variance. Certain a priori comparisons between different combinations of the various experimental groups had been theoretically decided and planned on before the start of the experiment. If the a priori questions that this thesis was concerned with could fulfill the statistical requirements of the Planned Comparisons method, then this method would clearly be the most powerful and appropriate statistical test to use on the data.

The maximum number of possible planned comparisons for any given set of $\underline{J}$ independent sample means is $\underline{J}$-1 comparisons, if each comparison is to be independent both of the grand mean and of each of the others. However, not just any set of $\underline{J}$-1 comparisons can be put together; the set of $\underline{J}$-1 comparisons must fulfill certain preconditions to insure that the comparisons will be independent of each other. Given normal population distributions with equal variances, the determination of the independence of the various comparisons depends only on the weights of the comparisons involved and not on any of the means actually observed. Thus one can and should plan the comparisons that will be independent before the data are collected. The J-1 set of comparisons must be independent of each other and of the grand mean, and this is achieved through the appropriate specifying of weights to the various sample means so that the two following conditions concerning the weights are fulfilled: (1) the sum of weights across the treatment conditions for each comparison equals zero, and (2) the cross-products of the weights across columns equal zero over all possible combinations of comparisons. If these two criteria are satisfied by the weights chosen, then the comparisons are said to be orthogonal and the information provided by each comparison is actually nonredundant with and unrelated to the information provided by the other comparisons (Hays, 1963, pp. 466-467).

There are numerous sets of J-1 comparisons that can be made on any given set of J treatment conditions and it is on the basis of the a priori theoretical questions that the most appropriate set of orthogonal comparisons is chosen. Several sets of weights were developed and considered to see which one set would most appropriately answer the largest number of initial a priori questions about the data. The set of weights that was chosen is shown in TABLE 12; this set of weights fulfilled the two requirements of orthogonality and also led to the most meaningful interpretation of the data for this thesis. As discussed more fully in the introduction, comparisons II and III are considered the most important for this thesis. The other comparisons, especially the last two, are considered of secondary importance.

The specific planned comparisons arrived at are as follows:

I. Groups 1-6 versus group 7. Do the experimental groups do better than the "empty" posttest controls?

II. Groups 1-3 versus groups 4-6. Do the computer groups do better than the comparison-booklet groups?

III. Group 3 versus groups 1-2. Does the Combined-Computer group do better than the other computer groups which have only one type of material?

IV. Group 2 versus group 1. Does the group working on the computer-programmed history materials do better than the group working on the computer General Problem Solving materials?

V. Group 1 versus groups 5-6. Does the Booklet-GPS group do better than the history-booklet groups?

VI. Group 6 versus group 5. Does the booklet-history group with thinking guides do better than the booklet-history group without guides?

The statistical formula (cf. Hays, 1963, pp. 462-468) used for the one-way planned comparisons analysis is the following:

$$F = \frac{(\hat{\psi})^2}{\text{MS error } \left(\sum_j c_j^2/n_j\right)}$$

with 1 and N-J degrees of freedom;

where $\hat{\psi} = c_1M_1 + \ldots + c_JM_J = \sum_j c_jM_j$, c is the weight, and M is the sample mean. The MS error term is derived from the basic one-way analysis of variance procedure.

B. Preliminary Analysis of Variance

Since the total experimental design included the completely-crossed factors of sex and IQ across the seven experimental conditions, the question arose as to the best method of doing the actual planned comparisons of the seven experimental conditions on the data. That is, the question was whether to group all the subjects together within the seven experimental conditions regardless of sex and IQ and do one analysis, or to do two analyses on males and females, or two analyses on Middle IQ and High IQ subjects. It was not felt that it would be statistically feasible to do four analyses on the fully crossed factors of sex and IQ since there would only be six subjects per cell across the seven experimental conditions. It also would not be possible, because of the obvious statistical redundancy, to run two analyses on sex and then another two analyses on IQ. The question of doing one overall planned comparisons analysis, or two analyses based on either sex or IQ was resolved by a preliminary four-way analysis of variance on the data (including the class-homeroom factor) to see if either the effects of sex or of IQ warranted doing two planned comparisons analyses instead of one overall analysis on the experimental conditions. It should be noted that the only purpose of doing this analysis of variance was to decide whether to do one general planned comparisons analysis or two analyses based on either sex or IQ. This test was not intended as a preliminary overall F test of the experimental conditions since the planned comparisons method does not require an a priori overall significant F statistic as does the post-hoc comparisons method.

The results of the preliminary analysis of variance clearly showed that IQ had to be considered as a significant single factor influencing the results of two of the three performance sections and also having a significant effect, although not nearly as great, on the attitude measures. Also the several significant interactions with the experimental-group factor gave more weight to its being considered as a significant separate factor. The factor of sex on the other hand showed only sporadic significant effects over the performance and attitude measures, and the overall picture did not show it as being nearly as important in affecting the results as the IQ factor. This preliminary analysis clearly showed that the planned comparisons analysis should be done separately on both the Middle IQ and High IQ subjects with the factor of sex being collapsed. The effect of the class-homeroom factor was eliminated by changing all scores into deviation scores around the appropriate classroom means on each composite variable before the planned comparisons analyses were performed. Since the composite cluster scores were originally standardized scores with a mean of 50 and SD of 10, the constant 50 was added to all these deviation scores to facilitate the readibility of the results, while a constant of 10 was considered more appropriate to add to the composite additive scores to eliminate the negative deviation numbers.

IV. RESULTS AND DISCUSSIONS

In the first part of this section the results of each of the six planned comparisons between the specified combinations of experimental groups will be presented separately, first for the Middle IQ subjects and then for the High IQ subjects. Following that will be an overview of the main findings of these six comparisons concerning the effectiveness of the computer-assisted instructional materials. The final part of this section will present the results and discussion pertaining to the second main question of this thesis on the feasibility and efficiency of specially developing problem-solving CAI programs utilizing the COMPUTEST computer language.

I. EFFECTIVENESS OF TRAINING FOR PRODUCTIVE PROBLEM-SOLVING SKILLS AND ATTITUDES

A. Comparison I: All Experimental Groups vs. the "Empty" Posttest Control Group

This first comparison is basically an examination of the performance of the Untutored-PT subjects in order to establish performance and attitude baselines of completely untrained subjects against whom the other experimental subjects can be checked. Although all the experimental groups did receive some form of "training" materials, it should be remembered that, as stated in the introduction, the booklet reading groups should more appropriately be looked upon as base comparison groups rather than as training groups since the two history-booklet groups received reading materials designed solely for the purpose of experientially involving them in the experiment. It should therefore be understood that, if in this comparison results are in favor of the experimental groups, the results are confounded between the computer-trained subjects and the booklet subjects, and all that can be inferred is that there is some unspecified facilitative effect arising from the experience with the experimental materials. The specific source of the effect must be searched out in the subsequent comparisons.

(1) Middle IQ Subjects (TABLE 13)

Even though there are no statistically significant differences between the two groups in the performance variables, the fact that the untutored posttest-control subjects tend to do as well or even slightly better than the experimental subjects calls for some explanation. The interpretation of these findings becomes clearer when one examines the three Problem-Evaluation variables and finds that the Untutored-PT group has rated all three of the posttest problems higher, two of them significantly, and the basis for these ratings was the problem-solving nature of the posttest materials. This positive evaluation of the posttests by the untutored control subjects strongly supports the interpretation that the experimental design has led to an artifactual positive motivational set of novelty and escape from the daily school routine for the Untutored-PT students. This results from the fact that this was their first and only participation in the study in any form within the three weeks of involvement for their class. A similar artifactual result of this kind of spurious motivational set for the posttest-control students was found in the Racine study of the <u>General Problem</u>

Solving materials and a similar interpretation of a novelty set for the controls was argued (Olton, et al, in press).

On the other hand, within the experimental groups there seems to be another factor working here, namely an initial negative motivational set for the computer-trained subjects since they now had to work on pencil-and-paper posttests within a relatively unstimulating classroom environment whereas before they had worked on the training materials in an active feedback situation. A strong degree of support is given this interpretation when one considers the trend of the ratings of these three posttest problems by the Middle IQ Computer-History group which goes from a low of 53.6 for the first problem, to 55.1 for the second, and 60.6 for the third problem. Apparently the initial negative reaction to this noninteractive, nonfeedback testing situation gives way to an increasing involvement in working on the problems despite the unresponsive testing situation. In any case, these results are interesting and important for the interpretation of the other findings in this thesis when one considers the performance results of the Untutored-PT group as being the upper performance limit solely due to the effects of a highly positive motivational set, regardless of the nature or transience of that positive set.

The results of the Evaluation of the Training Materials seem to show that the experimental groups base their rating of the materials on the problem-solving nature of the materials or, at least, their interesting nature; not unexpectedly, the Untutored-PT controls rate their regular history materials on the basis of facts and story-content. The most important finding in the Self-Evaluation inventory is that the Untutored-PT controls indicate they would approach a problem more through getting just one idea and sticking with it rather than working with different possible approaches.

The other Self-Evaluation variables also favor the experimental groups, one significantly; while the fact that the Untutored-PT controls are higher on the variable of problem-persistence appears to result from the artifactual positive motivational set and thus be basically spurious.

(2) High IQ Subjects (TABLE 14)

In examining the performance variables of the High IQ groups in this comparison, we find the direction of the results reversed from that of the Middle IQ subjects. All the performance variables favor the experimental subjects, with the differences of the two performance Idea-Evaluation variables being statistically significant. It turns out that this reversal of the direction of the differences between the experimental and Untutored-PT control subjects is not due to the fact that the High IQ experimental subjects do much better than their Middle IQ counterparts, but that the High IQ Untutored-PT control subjects do so much poorer than their Middle IQ counterparts. In fact, the size of the differences between the High IQ and Middle IQ experimental subjects on almost every variable is less than one point-- a finding which is surprising enough considering that one would expect the High IQ subjects to perform better. But on the other hand, the differences between the High IQ and Middle IQ Untutored-PT subjects are startling. On every variable, except one, it is the Middle IQ Untutored-PT group which is <u>higher</u>. In the Idea-Generation section four of these differences range from two to six points, while in the Idea-Evaluation section there is a six-point difference in favor of the Middle IQ Untutored-PT group for variable C6 (which is a negative variable) and an eleven-point difference for variable C7!

It does not seem at all plausible that the Middle IQ Untutored subjects possess more of the necessary cognitive skills to solve complex problems like these. A more likely interpretation centers around the reasons for attitude and motivation differences between the Middle and High IQ subjects. It turns out that all students who had IQ scores of over 130 were involved in the gifted-student program of the state, while those students with IQ scores over 120 were also included in a special gifted-student program of the school district. Practically speaking, this meant that all those high IQ students were involved in special, independent, library-oriented research projects and other individual academic projects set up through their homeroom teachers. Therefore, while participation in this study would be something special for the Middle IQ students and hence result in a positive motivational set, it probably did not impress or motivate the High IQ students to the same extent.

It also seems that besides the High IQ Untutored-PT students' generally initial low motivation to perform, they also did not perform well because the posttests might not have presented a challenging enough task for them to get involved with and work on. This is supported by the fact that the High IQ Untutored-PT control group rated all three of the posttest problems as <u>easier</u> to work on than the High IQ experimental subjects, even though their performance is much poorer. This would seem to indicate the High IQ Untutored-PT subjects tended to consider the problems superficially and did not make as earnest an attempt to work on them.

This interpretation leads to a tentative hypothesis about the effects of the instructional materials upon the High IQ experimental subjects. If we consider the actual results of the High IQ Untutored-PT group in the performance measures as the upper performance limit of High IQ students with generally low motivation, we find that the training materials must have had some sort of positive effect on the High IQ experimental students in order for them to perform so much better in the Idea-Evaluation section and consistently better on all other performance variables. Is this due to an increased ability to use different cognitive problem-solving skills, or is it rather due to an attitude change concerning problem-solving and a more positive motivational disposition to work on such cognitive problems?

The latter interpretation of a positive attitude and motivational change seems to be better supported by the data. The High IQ experimental subjects rated two of the three posttest problems higher and stated the problem-solving nature of the materials as the basis for their positive rating. Also the variables in the Self-Evaluation section indicate a more positive attitude change toward problem-solving for the experimental subjects.

This discussion of these results would thus lead to the tentative hypothesis that in the Middle IQ subjects the primary effects of the training materials would be directly to bring about a performance change and, as a result of the increased proficiency in solving problems, a concomitant attitude change toward problem-solving; whereas in the High IQ subjects the main effects of the training materials would be reversed, that is, the training materials directly effect an attitude change in the subjects toward working on problems which in turn results in a performance increment. This would explain why, in spite of the poor showing of the High IQ Untutored-PT control group, the High IQ experimental subjects performed as well as they did; but it does not explain why as a general group they did not do better than the Middle IQ experimental subjects. This surprising result, as we shall see in Comparison III, is due to the exceedingly superior performance of the Combined-Computer group in the Middle IQ subjects.

The results on the Evaluation of the Training Materials show quite clearly that the experimental groups have a much stronger positive attitude toward the experimental materials than the Untutored-PT controls have toward their classroom history materials. It also can be seen in variable C13 that the experimental subjects consider that they have learned problem-solving skills in using the materials, while the Untutored-PT controls hardly mention anything about learning any type of thinking skill in their regular history-class materials.

The results on the Self-Evaluation form, as mentioned before, support the hypothesis that the experimental materials brought about a significant positive change in the experimental subjects with all of the differences being in the right direction, three of them statistically significant. It is clear that the Untutored-PT controls would tend to stick with one main idea when working on a problem rather than approaching it through different possible ideas. Variable C15 is important in helping explain the performance results of the experimental subjects as it shows a significant persistence in working on difficult problems which is singularly lacking in the Untutored-PT control subjects. The other three variables are supportive of these general conclusions.

B. Comparison II: Computer Versus Comparison-Booklet Groups

This individual comparison, along with Comparison III, constitute the main statistical analyses of this thesis. The comparison sets up the computer-training groups against the other appropriate comparison-booklet subjects to determine if, among the various experimental materials, the computer-training programs do lead to a significant increment in problem-solving performance and attitudes.

(1) Middle IQ Subjects (TABLE 15)

All the performance variables favor the computer-trained subjects with two of them being statistically significant; however, the results are not as impressive as had been initially expected. One plausible explanation for these small performance differences between the computer and booklet groups is that there were not enough training lessons to produce really substantial differential training effects. As discussed before, the choice of using only five training lessons was theoretically supported by the assumption that what was occurring in these training lessons was not so much a direct teaching of new cognitive skills, but rather a sensitization of problem-solving skills which the person already possessed (Crutchfield, 1965). On empirical grounds, this decision was based on the results that the trained subjects in the initial studies using the General Problem Solving materials (Covington and Crutchfield, 1965) were already surpassing the control subjects after only four training lessons.

However, new results obtained from the recent Racine study (Olton, et al, in press), which became known to the author after the subjects had already been run in this thesis, showed a somewhat different picture of the training effect than had the earlier studies. The increase in the trained subjects' performance after only five training lessons was still there; but now it was only a slight difference which gradually continued to grow larger as more lessons were given. This change in the results from an immediate to a gradually increasing difference was basically attributed to the teacher's role in the training period. In the initial studies (Covington and Crutchfield, 1965) the teacher played a general supportive role in talking about the training materials in a positive way and discussing certain problem aspects and how they fit in with the other school work. In the Racine study, on the other hand, the training materials were run under very stringent experimental conditions in which the teacher took no active supportive role concerning the training materials, not even discussing them in class except for the routine procedural instructions. The interpretation was that teacher participation, even if in modest amount, produces greater benefits from the materials (Olton, et al, in press) and this hypothesis was also supported by evidence from another study (Blount, et al, 1967) in which a moderate degree of teacher participation (e.g., writing comments on the students' materials, providing encouragement, etc.) increased the students' performance in booklet programmed materials by as much as fifty percent.

In the present experiment there was no teacher involvement of any kind with the training materials, a more stringent teacherless participation condition than even the Racine study. In view of this, it is perhaps not surprising that there were no large statistically significant differences between the computer-trained subjects and the comparison-booklet groups. The encouraging fact, indeed, is that all of the differences were in favor of the computer-trained subjects. When we look at the results of Comparison III, we will find very strong evidence that supports our general explanation, since adding only five more computer lessons to the Combined-Computer group produced statistically significant performance differences for all but one performance variable.

In the Problem-Evaluation results we find that the computer subjects have rated only the PANJA problem significantly higher. These results are surprising in the sense that one would have expected the computer groups to enjoy working with the posttest problems more since they have been working on other similar problems in an active feedback and interaction situation on the computer. The key to this perplexing result may lie in the interaction and feedback nature of the posttest situation; whereas the computer subjects had been used to working in an active problem environment on the computer which responded to their answers, in the posttest problems this was now completely changed into an inactive, nonfeedback situation. On the other hand, the booklet subjects had been working in a nonresponsive environment-- simply reading story materials and answering banal content questions at the end of the materials; for them the posttest problems were perhaps a much more interesting and challenging situation than that which they had been having in the training period.

This interpretation is of a negative set on the part of the computer subjects toward working on the posttest problems, already mentioned in Comparison I. There seem to be two possible explanations concerning the computer subjects' posttest performance, both of which receive some support from the experimental data. The first is that the negative set is a temporary situational effect which quickly dissipates after the initial involvement in the nonresponsive situation. This receives some confirmation from the results of the Computer-PH group inasmuch as their ratings of the posttest problems positively increases from the first to the last problem as discussed in Comparison I.

Besides the explanation of the nonresponsiveness of the general testing situation, a second possible explanation is this: The posttest problems themselves were not overall intrinsically interactive enough. This receives support from the fact that while none of the booklet posttest problems are set in as active an interactive response mode as in the computer situation, one of the three problems, PANJA, does intrinsically involve a greater feedback effect by the nature of its format. And it is only this posttest problem which is significantly rated higher by the computer subjects.

The problem on PANJA sets up an interactive situation in that the greater part of the problem consists of evaluating the possible ideas against the facts which are contained in the story; thus there is set up a quasi-feedback situation which is not found in the other two problems, especially WHI which is mainly concerned with the nonfeedback measure of Idea-Generation. Also in the PANJA problem the final facts of the problem lead to a closure in the problem solution, while for KASKIA, even though there is some interactive evaluation, there is no final information to provide an appropriate feedback conclusion.

This interpretation of the effect of the basically noninteractive posttest problems also receives support from the results of the High IQ subjects in that the Idea-Evaluation variables (which are derived mainly from PANJA) are the only performance variables in which the High IQ computer groups do significantly better than the booklet groups. Further support is given by the fact that even though KASKIA is the posttest problem which is most similar to the computer subjects' training materials, namely the unit on SAN VALLENDO, and would thus be expected to be rated highly, it was generally rated lower than the other two problems. Also, it would not be expected that this negative set interpretation would hold for the Computer-GPS group who simply took the basically noninteractive linear materials in the computer-simulated situation; and, in fact, the results for this group show only a steady decrease in positive ratings for all three posttest problems for both IQ groups.

Although there are no large statistically significant differences between the computer and booklet groups in the performance measures, the results of the Evaluation of the Training Materials show quite a different picture. Every variable is statistically significant in favor of the computer groups. The most critical variable is C12 which measures tne overall positive evaluation of the training materials; here we find the computer groups clearly have rated their materials significantly higher (.01 level). This is very important in considering the feasibility of the CAI method in the school situation from the standpoint of the students' motivation.

Consideration of the so-called "Hawthorne" effect (Uttal, 1962) might lead one to explain these differences in evaluation between the computer and booklet subjects as being basically a trivial artifactual outcome--a result of the computer subjects being initially enthusiastic about a new device, the computer. But if it can be shown that the computer was not such an exciting novelty for the subjects, then this positive evaluation by the computer subjects would have to be attributed to the training materials themselves and thus the objection would lose its value.

However, in fact, the subjects used in this experiment were not naive as concerned the computer system and operations; indeed, as discussed in the subject selection section above, this was one of the factors considered in using the eighth grade students over the new seventh grade students who were naive as to the computer project. The year before, when the eighth grade subjects were in the seventh grade, all these students had been involved in an informal way with the computer project both by participating in demonstrations of various short computer information programs written by the project staff and interested teachers, and also by working on the computer programs that had been written by other classmates on various topics concerning their school subjects. A majority of the students also had some experience in writing brief computer programs of their own in the COMPUTEST language (Starkweather, 1965; Hodge, 1966) and thus became quite familiar and sophisticated as regards actually working with the computer. (Note: Two students had become extensively involved in working on the computer and were not included as subjects in the experiment.) In fact, there was an initial student enthusiasm in that first year toward the computer project which did drop off as a result of a policy requiring any student developing his own computer program to work solely with simple question-and-answer sequences within specified school-subject content areas. These restrictions soon led to a decrease in the number of students working on their own programs and, consequently, a drop-off of the other students who would have participated by taking the computer programs written by their classmates. Indeed, far from there being a spurious positive "Hawthorne" set for working on the computer, the author found a slight negative reaction of the students when giving the initial student orientation talk. Questions were raised by some students in a negative way about "whether they were going to have to do the same things as last year", and they had to be reassured that the materials and their participation would be quite different from the previous year. Thus, on this point of the positive novelty of the computer, the "Hawthorne effect" objection does not seem to bear much weight.

Another point: The implication of the "Hawthorne" objection is that this positive set would be a transient effect only. Yet an analysis of

Comparison III shows that the Combined-Computer group, which worked twice as long on the computer materials, namely 12 to 16 periods, positively rated the training materials almost two points higher in the Middle IQ subjects and nearly as high for the High IQ subjects. If this effect were a transient one, we would expect this effect to show some sign of deteriorating in the comparison of those subjects who took 6 to 8 periods of computer instruction with those who took 12 to 16 periods. Also other studies (Suppes, 1967; Bitzer, et al, 1967), in which students were involved in working on CAI materials for longer extended periods of time up to four or five months, report the same general finding that the students' motivation remain at a consistently high level. The facts seem to show, therefore, that the positive rating of the instructional materials by the computer subjects is attributable to the nature of the training materials themselves and not to the artifactual effect of the novelty of the computer apparatus.

In the Self-Evaluation results we find that all the variables are in favor of the computer subjects with three of these being statistically significant. The computer subjects, to a greater degree than the booklet subjects, consider themselves to be good problem solvers and would persist in working on difficult problems; the booklet subjects would still tend to stick with one idea when working on a problem rather than approach it in different ways as would the computer subjects. The two additive composite variables are both statistically significant in favor of the computer subjects and show, in a general way, that the computer subjects' attitudes more closely approximate the attitudes of a theoretically good problem solver. This significant increment in the positive problem-solving attitudes of the computer-instructed subjects constitutes a highly important finding of this thesis.

(2) High IQ Subjects (TABLE 16)

The computer subjects perform significantly better than the comparison-booklet subjects on the two performance measures of Idea-Evaluation. The results of the other performance variables are mixed and do not show any statistically significant effects. The fact that the computer subjects do statistically better only on the two performance variables concerned with Idea-Evaluation lends support to the earlier interpretation that the computer subjects generally had a negative set toward working on the post-test problems because of their nonfeedback, noninteractive mode of presentation. This supporting evidence comes from the fact that the Idea-Evaluation questions, as compared with the Problem-Formulation and Idea-Generation measures, were the only ones in the posttests actually containing some intrinsic feedback qualities despite the basically nonresponsive character of the posttests; thus, it would be expected that the computer subjects would respond more to these performance questions than the others. This is substantiated by the fact that the High IQ computer history groups (the GPS simulated computer group is not involved in this discussion because of its basically noninteractive training format) rate the problem of PANJA the highest of all three posttest performance problems and it is this problem which is almost completely concerned with the feedback aspect of Idea-Evaluation.

Of course, another factor which may explain the overall poorer performance showing of the High IQ computer subjects is that their median IQ is significantly lower than the High IQ booklet subjects in this comparison as seen in TABLES 4 and 5. This significant IQ difference inadvertently arose out of the incidental factors affecting the subject-sampling procedures as discussed more fully in the subject-selection section above. Since in the preliminary analysis of variance IQ has been found to be a significant factor influencing the results of this study, this significant IQ difference between the two experimental groups in this comparison must be considered an important possible explanation.

In the Problem-Evaluation results three of the four variables are in favor of the computer groups, but none of the differences is very large. Even though the computer history groups (as seen from other comparisons) do rate the problem on PANJA the highest, the Computer-GPS group rates it the lowest, with the end-result that the computer groups as a whole do not have a combined rating much higher than the booklet groups.

The findings on the Evaluation of the Training Materials are striking and important. The computer subjects rate their materials significantly higher than the booklet subjects rate theirs, and by reason of their problem-solving nature. All the variables agree in this positive evaluation by the computer subjects with all of the differences being statistically significant. Indeed, the scores of the High IQ computer subjects on all of the Training Materials Evaluation variables, except one, are higher than the ratings given by the Middle IQ computer subjects. This very favorable evaluation by the High IQ computer subjects is certainly an encouraging and important finding in the consideration of the need for developing more sophisticated computer-instructional materials which would systematically extend over longer periods of time.

The Self-Evaluation variables also present some very interesting results. In every case the computer subjects have higher positive scores than the booklet subjects with four of the five attitude differences being statistically significant. (Note: For variable C14 a lower score is "better.") Even though these are all high IQ students who would be expected to already have high self-confidence in their thinking abilities, the subjects who have gone through the computer-training materials rate themselves significantly higher as better problem solvers in variable C16. Also they state that they will work longer on difficult problems and approach problems through different possible ideas rather than stick with only one main possibility as would the booklet groups. The two composite additive variables give full support to these cluster variables, both of them being statistically significant in favor of the computer-trained subjects.

Overall, it is quite clear that the computer-training materials produced a significant attitude change among the High IQ subjects. This, coupled with their strong positive evaluation of the computer materials and their superior performance on the Idea-Evaluation variables, gives positive proof of the beneficial effect of the computer-instructional materials and strong encouragement for their further revision and their development into other training areas.

C. Comparison III: Combined-Computer Versus Other Computer Groups

This individual comparison was made to see if there were any significant increase in performance and attitude measures for the Combined-Computer subjects who took both types of computer-training materials. This comparison constituted a test of the best computer-training condition and, along with Comparison II, is one of the most critical tests of the CAI method in this thesis.

(1) Middle IQ Subjects (TABLE 17)

In this comparison of the Middle IQ subjects, we find the most important results of the whole experiment. All of the performance variables are in favor of the Combined-Computer Group with all of the differences, except one, being statistically significant. These findings represent the most consistent pattern of performance results in all of the planned comparisons made for both Middle IQ and High IQ subjects-- in all three performance areas of Problem-Formulation, Idea-Generation and Idea-Evaluation the Combined-Computer group is significantly better. In every performance variable, except one, the actual mean score of this group is higher than any other training group or combination of training groups both for the Middle IQ and High IQ subjects. It is clear that the Combined-Computer group is superior in performance to the other Middle IQ computer and experimental groups, and that its performance even overshadows the performances of any of the High IQ groups.

There are two main possible interpretations of these results. The first and more obvious one is that the Combined-Computer group had a longer training period with more lessons than any of the other experimental groups and it was this that led to their superior performance. The second possible interpretation is that the combination and interaction of the two different types of computer-training materials produced this significant performance difference. These explanations are not mutually exclusive. It will not be possible, however, to determine from our results just how much each of these factors has contributed solely by itself to this superior performance since both factors are confounded in the experimental design of this group. This will have to await another experiment for solution; but we can consider each of these two possible interpretations separately and examine the conclusions that can be drawn about them.

The interpretation of the effects as due to the extra lessons and lengthened time the Combined-Computer subjects were in the training period is consonant with the Crutchfield and Covington studies in which it was found that as the experimental subjects continued working on the training materials they continued to become more and more proficient in solving problems (Crutchfield, 1966; Covington and Crutchfield, 1965). The recent results of the Racine study (Olton, et al, in press) also have shown this effect with, however, the important difference that the increase in performance is much more gradual for the trained subjects. The present study was run under teacherless experimental conditions which were very similar to that of the Racine study and essentially the same results have been found--no real significant performance differences in the posttests until about 12 to 15 periods of training have been completed.

One major implication of this interpretation of the results is that the Combined-Computer subjects perform better on the posttest problems simply because they have been conditioned to work longer on more problems. But that there is much more to it than that becomes obvious when one examines the performance results more carefully. If the sheer total number of ideas had been the only performance variable affected, then one could understand how the Combined-Computer group could have performed better simply through acquired readiness to persist. But even the first four performance measures of Idea-Generation were not just simple quantity counts but were the measures of the <u>quality</u> of ideas, presumably not so directly facilitated by sheer readiness to persist.

Moreover, this simple explanation would not explain the superior showing of this group on the other performance variables. For example, the

Combined-Computer subjects generated ideas which were also significantly more original than the other computer groups; and they learned to utilize the problem-information to a much greater extent in the evaluation of these ideas, to become aware of discrepant initial information, and to ask more appropriate information-seeking questions in the problem-formulation. The superior performance on these variables certainly demands more of an explanation than simple persistence of effort. It becomes obvious that the combination and interaction of the two different types of computer-instructional materials must also be considered for a complete understanding of the Combined-Computer group's superior performance.

An examination of the basic pedagogical makeup of the two series of computer materials reveals two different aspects which apparently have complemented each other in a very effective way. In the General Problem Solving materials there was the direct expounding of the basic rules of problem-solving while, however, there was a minimum of active problem feedback and interaction since these materials were set up to closely follow the original booklet version with its limited linear branching technique. On the other hand, in the programmed-history materials the direct explication of problem-solving rules was not done save for a short review session, but throughout the training programs there were ample opportunities for active problem-solving interaction and feedback as the subject worked through the problems. It seems evident that this combination of instructional approaches-- involving both the direct explication and teaching of thinking skills and the problem-solving practice of directly working on interactive problem situations-- is critically important for the most effective training of students in the development of the various cognitive skills necessary for successful problem-solving behavior.

This interpretation is supported by the most recent studies performed by the Crutchfield and Covington project in which they have developed supplementary workbook practice materials for the students to use in conjunction with the regular General Problem Solving booklets. The preliminary analyses show that in both the Hillside and Cragmont studies (Crutchfield and Covington, 1967) the students who worked on the supplementary practice materials in conjunction with the regular GPS booklets outperformed both the untutored students and the other subjects who took only the regular GPS booklet training materials.

We find that in the Problem-Evaluation results the Combined-Computer subjects rate all three problems higher than the other two computer groups with two of these three differences being statistically significant. The reason given by the Combined-Computer subjects for their positive rating of the posttest problems is the problem-solving nature of the tasks. Thus the combination and interaction of the two types of computer-training materials has led the Combined-Computer group not only to outperform the other computer groups, indeed any subjects of Middle or High IQ, but also to actually enjoy working on the problems more.

In the Evaluation of the Training Materials, the Combined-Computer subjects not only consider that they have learned more about problem-solving skills but also, to a significantly greater degree, state, as for the posttests, the problem-solving nature of the training materials as their reason for their positive rating of them. That the Combined-Computer subjects find their materials more interesting than do the other two computer groups is noteworthy in indicating that there is not a fast drop of initial enthusiasm in working on computer materials but an even higher positive rating of the computer-training materials which are twice as long. This augurs well for the development of future, more extensive computer-instructional materials of the same nature.

In the Self-Evaluation results we again find the Combined-Computer subjects higher in positive attitudes toward their approaches to working on problems and their own evaluation of themselves as problem solvers. They are significantly more aware of the necessity of approaching a problem from different possible perspectives rather than in just one main way, they are higher in espousing the need to persist in working on a difficult problem even when stuck, and they have more confidence in themselves as problem solvers. All in all, it is quite evident that the combinational and interactive effect of the Combined-Computer training materials has led to a positive change of attitudes in the subjects, which is no less important an effect-- even though more gradual in development and not as spectacular in degree-- as the effects on the performance measures.

(2) High IQ Subjects (TABLE 18)

As contrasted with the foregoing positive findings on the performance of the Middle IQ Combined-Computer group, the performance results for the High IQ subjects are difficult to evaluate. While there is a consistent trend of all the measures of Idea-Generation in favor of the Combined-Computer group, none of these differences is statistically significant. Furthermore, the two performance measures of Idea-Evaluation are in favor of the other two computer groups, while the two measures of Problem-Formulation are inconclusively split between the Combined-Computer group and the other two groups.

In examining the Problem-Evaluation data there again are found no statistically significant differences; the two computer groups do rate all three posttest problems higher, but by quite small differences. However, the Combined-Computer subjects rate the posttest problems on the basis of their problem-solving nature slightly more frequently than do the other two groups. In the data on Evaluation of the Training Materials and on the Self-Evaluation inventory, there are no significant differences and the outcomes are mixed with no consistent trends.

Why is there such a superior performance by the Middle IQ Combined-Computer group and such a complete dearth of performance increments in the High IQ Combined-Computer subjects? A plausible explanation has to do with a difference in the attitudes toward the training and posttest materials held by Middle and High IQ subjects. The results from a supplementary study in which two subjects, one a High IQ and the other a Middle IQ subject, worked together at the computer provide some enlightening information on how the materials are differently reacted to. A large percentage of the Middle IQ subjects in special posttest questions complained about the lack of interest of their High IQ partners in working on the computer-training materials and the "fooling around" by them which prevented the Middle IQ subjects from seriously working on the materials. Thus, it seems likely that the High IQ subjects may not have considered the training materials (or the posttests) sufficiently challenging. In short, it appears that the relevant issue here is one of motivation and attitude and these results clearly point out a crucial need of sufficiently motivating the higher IQ student through appropriate training methods and materials in order to produce a computer-assisted instructional situation in which learning will take place.

D. Comparison IV: Computer-GPS Versus Computer-PH

This statistical analysis is not, of course, intended as a direct comparison of the two different types of training materials per se, because of all the uncontrolled differences in these two instructional methods, pertaining to factors such as response recognition, branching techniques, active feedback, visual vs. verbal format, etc. The purpose of the comparison is to examine the separate effects of each of these two instructional materials taken alone, in order to arrive at some better understanding of the interactive effect of the two sets of materials on the performance of the Combined-Computer subjects in Comparison III.

(1) Middle IQ Subjects (TABLE 19)

Although there are no statistically significant differences between the two groups in the performance variables, the consistency of the direction of the results, especially in the Idea-Generation section, indicates that the General Problem Solving materials had the greater separate overall effect on the subjects.

In examining the Problem-Evaluation results, we find that the Computer-GPS group has rated two of the three problems higher, one of them significantly, and has done it on the basis of the problem-solving nature of the posttest materials. The differences between the two groups in the Evaluation of the Training Materials is small but in favor of the Computer-GPS group except for one important reversal on variable C12, which is the main variable about positive attitude toward the training materials. The Self-Evaluation inventory shows a very strong consistent trend in favor of the Computer-GPS subjects on every variable, with the rating on self-confidence as a problem solver being significantly in favor of the GPS group and three of the other four variables all showing a substantial difference in its favor.

Looking at the overall consistent trend of all the variables, therefore, it seems evident that the General Problem Solving computer materials alone produce a consistently greater positive effect on the Middle IQ subjects for both the performance and attitude variables than do the history computer materials alone. Nor is this surprising when one considers that the GPS materials have been especially developed and revised over the past five years for their pedagogical and psychological effectiveness in training for these problem-solving skills, that they give an easy to read and follow visual format, that they directly enumerate the thinking skills being taught, and that they take the student by means of carefully planned feedback through to the solution of the problem. The history materials for the Computer-PH group, on the other hand, do not directly explicate the thinking skills involved, except for one small review segment, and are more a direct test of the application of thinking skills in open and interactive problem-solving games in which the student participates. The more direct teaching of these problem-solving skills by the GPS materials, as contrasted with the predominantly practice aspect of the Programmed-History materials, appears to have a stronger impact on these Middle IQ subjects.

But this does not mean, of course, that the computer history materials have no real effect on these subjects. We have already seen above that when taken in combination with the GPS materials the history instructional materials produce a multiplied beneficial effect. It appears that the GPS materials directly teach the student the necessary thinking skills and the history materials give him a much more active chance to practice these skills.

(2) High IQ Subjects (TABLE 20)

For High IQ subjects, the findings are quite different. Although there are no statistically significant differences among the performance variables, the trend is consistently in favor of the history instructional materials over the GPS materials, except for one variable. This reversal from what was found in the Middle IQ subjects is consistent with an explanation that, while the Middle IQ subjects can benefit more from the direct teaching of the GPS materials, the High IQ subjects tend to profit more from an interactive situation involving practice in problem-solving tasks. This is not to say that the High IQ students could not also benefit from materials which were designed for direct teaching of the necessary thinking skills in an appropriately sophisticated manner: but, in this case, the direct training methods of the GPS lessons were written for the fifth and sixth grade level and thus would not be expected to have appeal to our brighter students of the eighth grade. Thus we find the history instructional materials lead to consistently better problem-solving performance in higher IQ students and also to a higher positive evaluation by them of the posttest problems on the basis of their problem-solving nature. The results for the Training Materials Evaluation are mixed and the differences are small with no consistent pattern being established; however, in the Self-Evaluation results there is a slight advantage in favor of the Computer-PH group.

E. Comparison V: Booklet-GPS Versus History Booklets Groups

It must be stressed that the intention of this analysis, and the following one, is not any type of direct comparison of the two groups. These groups are best thought of as "active control" groups rather than as training groups. Thus these analyses serve solely a subsidiary function for gleaning information from these results which may aid in the understanding of the previous main comparisons. Bearing in mind, then, all the uncontrolled differences between the two groups, this analysis can be loosely construed as an overall comparison between the programmed-instruction booklet format and the series of reading materials.

(1) Middle IQ Subjects (TABLE 21)

In looking at the Problem-Formulation and Idea-Generation performance variables, we find that the consistent trend of all differences (with one of them statistically significant) is in favor of the history reading groups. This pattern is reversed for the two Idea-Evaluation variables where the GPS subjects excel.

Although we see that the GPS booklets do not fare particularly well in this comparison, the picture looks quite different when we perform a special supplementary analysis in which the GPS reading group is compared with the group of subjects for whom the GPS materials were presented in the simulated "computer" interactive mode. The interesting finding is that these same GPS materials when put in the simulated "computer" mode, even though the materials are still completely linear, lead these Computer-GPS subjects to outperform the Booklet-GPS subjects on every performance variable except one.

In looking at the evaluation results, we find that for every single variable in all three attitude areas the results are in favor of the GPS booklet subjects over the history booklet subjects. The booklet subjects who took the GPS instructional materials thus seem to have learned the proper attitudes toward problem-solving better than did the booklet history subjects. The most interesting results are seen in the evaluation of the posttest problems in which the GPS group rates all three problems positively higher (and for reason of their problem-solving nature) even though, as we have seen above, they do not actually perform better on two of the three problems. It is noteworthy that the attitudes toward problem-solving can be changed through direct teaching in a programmed-instruction booklet format, even though it may take more than this to bring about a change in the problem-solving performance skills.

(2) High IQ Subjects (TABLE 22)

For the High IQ subjects there are again no regularly consistent differences in problem-solving performance favoring either group. In the evaluation variables, however, the Booklet-GPS subjects do tend to show consistent superiority over the history booklet subjects. These results substantiate the findings on the Middle IQ subjects, that attitudes toward problem-solving, especially positive self-evaluation, can be fairly readily changed through direct training of these attitudes in booklet materials.

F. Comparison VI: Booklet-HR Versus Booklet-HR & Guides Group

This comparison was made in order to see just what additional effect, if any, would be produced by supplying half of the subjects who read the history booklet stories with appropriate guides on thinking.

(1) Middle IQ Subjects (TABLE 23)

There are no statistical differences between the two groups on any of the performance variables. In the Problem-Evaluation results the thinking guides group rates two of the three problems higher, basing the ratings more strongly on the problem-solving nature of the posttest materials. In the Self-Evaluation results four of the five measures are in favor of the thinking guides subjects. They tend to approach a problem in more than one way; they also have more self-confidence in their ability to solve problems. Both of these differences are statistically significant.

It appears that merely giving subjects a list of thinking guides to read tends to enhance their own positive evaluation of themselves as problem solvers, while also leading to more positive attitudes towards approaching problems and to a higher positive rating of problems that are worked on. It does not appear, on the other hand, that simply reading thinking guides has effectively facilitated problem-solving performance for these experimental subjects. (Covington and Crutchfield (1965) reported finding a very small performance increase in subjects who simply read thinking skill guides when compared with untrained control subjects.)

(2) High IQ Subjects (TABLE 24)

There are no consistently significant differences found between the two groups in either the performance or evaluation variables. This paucity of significant differences, combined with the significant difference between these two groups in their median IQ scores (as discussed in the section on subject selection above), makes it unprofitable to attempt any interpretation of these data.

2. OVERVIEW OF MAIN FINDINGS ON THE EFFECTIVENESS OF THE COMPUTER-ASSISTED INSTRUCTIONAL MATERIALS

A. Performance Measures

1. The computer-assisted instructional materials and method lead to substantial performance gains in all three principal domains of problem-

solving: namely, Problem-Formulation, Idea-Generation, Idea-Evaluation.

2. However, these statistically significant gains occur only in the most nearly optimal computer-training condition used in this study, the Combined-Computer condition, in which the subjects received both the programmed-history materials and the General Problem Solving (GPS) lessons in their respective computer modes.

(a) The beneficial interaction of the direct teaching of thinking skills and attitudes provided by the GPS lessons and the problem-solving practice provided by the game-feedback format of the programmed-history materials appears to be one of the crucial factors in this marked superior performance by the Combined-Computer subjects.

(b) The lengthened time of the Combined-Computer group's involvement with these auto-instructional computer materials also appears to be an important factor, especially when these training materials are used without any teacher participation.

3. The superior problem-solving performance of the Combined-Computer group is found to be true only of the subjects of the Middle IQ group which has a median IQ of 108, as contrasted with the median IQ of 123 for the High IQ group. The High IQ subjects do not show a clear performance gain from the training with the computer materials.

(a) This may indicate that computer-training materials of this nature, no matter how sophisticated, may not provide a valid learning situation for high IQ subjects, and that only average or lower IQ students can substantially benefit from this type of computer-instructional method and materials. Thus, an even more significant training effect might be found among subjects with "low" IQs, i.e., less than 100.

(b) However, a second interpretation of the poor performance showing of the High IQ computer subjects seems more plausible. It appears that the High IQ subjects were not aroused to work as seriously as they might have on the training materials and posttests owing to their previous involvement in state academic programs for gifted students; this computer training might not have been perceived by them as instructionally "innovative" to the same degree as for less favored students. Moreover, for them the posttests may not have appeared sufficiently challenging.

4. There were no consistently significant performance differences found in the various comparisons made between the other experimental groups. For example, neither the history computer group nor the GPS computer group, taken separately, were consistently superior to the other comparison groups.

(a) The non-interactive character of the posttest situation appears to be an important factor in understanding why these two groups of computer subjects did not excel in performance. Since these computer-trained subjects were used to working in an active feedback setting, the nonresponsive posttest situation probably induced a relatively low motivational set in them. For the comparison booklet-reading subjects, on the other hand, the posttest situation probably constituted a motivationally positive interaction setting. This points out the necessity of having a computer-posttest situation for both computer and booklet subjects as an important experimental design factor in future experiments.

(b) However, it seems that a more important factor explaining this dearth of performance results is the insufficient number of training lessons taken by these computer subjects. This interpretation of the findings is supported by the results of a recent study using the GPS materials

(Olton, et al, in press) in which significant differences were found between trained and control subjects only after 15 instructional lessons, and this is also supported by the present findings in which the Combined-Computer subjects, who did have twice the number of computer lessons, performed consistently and significantly better over all other experimental groups.

B. Attitude Measures

1. The computer-assisted instructional materials produce an even greater positive effect in the attitude measures than found in the problem-solving performance variables. Every measure in the Training Materials Evaluation section shows a statistically significant positive difference in favor of the computer-trained subjects. In the Self-Evaluation inventory, the results are the same, with every variable showing a positive increment favoring the computer-instructed subjects; almost all of these differences also are statistically significant.

(a) A possible criticism that the computer-trained subjects may have rated their instructional materials so highly owing to a transient and spurious "Hawthorne" effect resulting from the novelty of the computer apparatus does not seem to be justified. The subjects had already had substantial contact with the computer in the previous year and were thus not particularly vulnerable to a sheer novelty effect.

2. These positive attitude effects are found in the Combined-Computer group to the same extent as found in the other computer-instructed groups.

(a) This finding further lessens the criticism that the computer subjects have only a very transient motivational set, since if this were true one would expect to find a diminution of this motivational effect for the Combined-Computer subjects who had twice the amount of computer-training materials.

(b) The Combined-Computer subjects' high positive evaluation of their longer computer-training materials augurs well for future development of more extended training materials of this nature.

3. These substantial positive attitude changes in the computer-instructed groups are found for both Middle and High IQ subjects.

(a) Inasmuch as the computer-training materials here used were only in their initial stage of development, and inasmuch as the General Problem Solving materials were initially designed for fifth grade use, not for eighth-graders, it is very encouraging to find that even the High IQ computer subjects rate their materials so positively. This gives encouragement for the development of future computer-assisted instructional materials which would be even more sophisticated in their pedagogical and psychological structure and content.

(b) It is noteworthy that there are large positive attitude changes in the Self-Evaluation Items for the High IQ subjects, even though for these subjects the computer materials did not lead to consistently significant problem-solving performance gains.

4. In the various statistical comparisons made between the other experimental groups, one major finding merits attention. The booklet subjects using the GPS materials show a substantial positive attitude change in the items of the Self-Evaluation inventory. Also, their evaluation of the training materials is appreciably more favorable than that of the history booklet groups.

(a) It thus appears that the direct training of problem-solving attitudes through the programmed-instruction format of the GPS materials does lead to significant attitude changes, a finding consistent with reports of earlier studies of these materials.

(b) But it should also be noted that the GPS materials used in the computer-simulated training setting are rated even more favorably than is the booklet version of the same materials. Here again we see evidence of the potential effectiveness of the interactive mode of a computer-assisted approach.

3. ANALYSIS OF THE EFFICIENCY AND EFFECTIVENESS OF CODING THE COMPUTER-ASSISTED INSTRUCTIONAL PROGRAMS INTO THE COMPUTEST LANGUAGE

The relative brevity of this section should not be construed by the reader as indicating that the findings reported here are incidental or secondary to the posttest results covered in the previous section. Quite to the contrary, this information answers the second vitally important question of the thesis concerning the practical feasibility of developing computer-assisted instructional materials for the training of problem-solving skills. This question, while intimately related to the question of the effectiveness of the CAI method explored in the previous results section, must be directly considered in its own right. For even though we have seen that the particular CAI materials used in this thesis did, under certain conditions, prove effective in training for problem-solving skills, this does not necessarily indicate that these initial computer programs can be efficiently and effectively expanded into more extensive and sophisticated CAI materials which could be used in actual school settings.

The answering of this basic question directly concerns the utility of the COMPUTEST author language and the amount of time and effort which were required to develop, code, and debug the CAI programs used in this experiment. That this is not an irrevelant or banal issue can be seen from the fact that some investigators have abandoned the CAI approach because of the programming difficulties in their particular subject matter area (Strum and Ward, 1967), and from the fact that the problem of total work-hours per program development is invariably brought up as a vital matter in judging the feasibility of CAI systems in the actual school situation (Atkinson, 1967b).

The significance of this discussion, and indeed of this entire thesis, would be sharply reduced if the findings referred only to the CAI programs which could be built through the COMPUTEST language and therefore only usable on the small IBM 1620 computer; this is not a computer hardware system feasible for any type of large-scale educational approach. However, even though these particular programs were written in COMPUTEST, this discussion also properly applies to the new analogous computer author language being developed by Starkweather called PILOT (Programed Inquiry, Learning, Or Teaching). This language is designed for the IBM 360 computer system which is practical and feasible for large-scale school use through a time-sharing system. PILOT is being built to fulfill the same requirements of user-simplicity as COMPUTEST, while at the same time providing the author-user with more sophisticated and elegant programming possibilities. In short, since PILOT will be as easy to use as COMPUTEST, the discussion here can also be considered appropriate to the use of the PILOT author language in the writing of such programs. Indeed, with slight effort the original history programs written in COMPUTEST for this thesis can be and are being converted into the PILOT language for use on the remote terminal system of the IBM 360.

There are two main aspects in evaluating the COMPUTEST language used in this experiment, centering about Zinn's two criteria of a computer author language. The first aspect concerns Zinn's first criterion, that of user-simplicity: "Certainly the user in such a system should be able to write in his own language with a minimum of restrictions the self-instructional materials he plans to use....the novice should be able to prepare materials for computer instruction after only minutes of exploration of the system language" (1967, p. 81). User-simplicity had already been substantiated in one way by the fact that elementary school children have been able to write simple question-and-answer type computer programs by themselves using the basic coding features of the COMPUTEST language (Starkweather, 1965; Hodge, 1966). However, it had not yet been determined whether it was possible for someone completely naive to computer programming (i.e. the thesis-author) to write computer-instructional programs in the COMPUTEST language which would be much more sophisticated and, hopefully, pedagogically effective in training for productive thinking skills.

Prior to the development and coding of the CAI programs used in this experiment, the thesis-author had only worked on two very short sample computer programs which were written in the COMPUTEST language; besides that, he had no programming experience in any other computer language. Thus the thesis-author had ample qualifications for being classified as a novice in computer programming. Despite this lack of programming experience, he was able to write, code, and debug each of the computer programs for this thesis within approximately 25 to 30 hours after the initial skeletal form had been developed. It is realized, of course, that many more hours of work will be necessary for the revision of these programs to make them instructionally efficient, both pedagogically and psychologically; however, the fact that these programs could be programmed in such a short time and be made operational so that students could actually take them, is an important indication of the simplicity and feasibility of COMPUTEST for use by nonexperienced teachers and educators.

Grubb (1967b) describes his experiences in developing CAI materials and the constant revising of the programs on the basis of the students' performances. He mentions the frustrating problems which resulted from having always to work through a programmer rather than directly coding and revising the CAI programs himself (Grubb, 1967a). His conclusion was that there existed a need for a simple author language which would permit a direct interface between the content specialist and the computer. To achieve this end, he and others developed the IBM Coursewriter author language. Starkweather (1967c) has gone on to develop COMPUTEST (and PILOT) in an effort to make the author language even simpler, so it could be feasibly used by regular teachers (and students) rather than only by content specialists and related personnel connected with CAI projects. On the basis of the thesis-author's work in writing these programs it appears quite evident that COMPUTEST does fulfill this goal and hence Zinn's first criterion of user-simplicity.

Zinn states his second criterion of an author language as follows: "An essential characteristic of this author language is that it be user oriented without denying the author-instructor access to any of the system capabilities....and the experienced author should be able to use the capabilities of the computer to the fullest, for as complex a procedure as he can construct" (1967, p. 81). Despite the obvious constraints arising from the very small hardware system in which COMPUTEST is used, it proved possible to develop computer-training materials which sufficiently fulfilled the four functions of a sophisticated CAI program, as discussed before: (1) text presentation; (2) appropriate answer recognition; (3) elaborate scoring procedure; (4) complex branching techniques. As Grubb mentions, CAI programs are to be looked upon more as in "a state of 'becoming' rather than ever reaching some steady-state condition" (1967b, p. 71). Certainly the computer programs used in this thesis can be considered to be in their first state of being, ready to be revised and expanded on the basis of the students' actual performances in the initial programs. However, the fact that even in their primitive state these CAI programs do lead the Combined-Computer subjects to increase their problem-solving performance so markedly over all the other training groups gives strong indication that they are providing a valid learning experience in the training of productive thinking skills and attitudes and thus presumably fulfilling the four functions of a CAI program mentioned above.

As discussed before, the critical issue in developing and coding the computer programs centered around the basic problem of response recognition for answers to open-ended questions. Such open-ended questions were felt to be an essential part of productive thinking training materials, and indeed, central to the basic rationale for developing training materials using the computer. Starkweather (1967c) had this very problem in mind in developing COMPUTEST so as to be able to handle a free dialogue situation through the response recognition of key words in the input from the subject. Although some initial attempts had been made to develop demonstration type clinical-interview programs (Starkweather, 1967b), this present study was the first attempt at using the COMPUTEST language for producing a systematic valid learning situation involving training for productive thinking skills which utilized such a free, open-ended question technique.

In the actual programming of these open-ended questions in the training materials, it was found that there was great flexibility in setting up an answer recognition scheme which could demand as stringent or as loose a criterion of a correct answer as the author desired. These answer recognition schemes ranged from the more stringent extreme of requiring enough key words in the input sentence so that it was certain the subject had the same concept as was considered correct in the program, to the less stringent extreme of requiring only that the input sentence contain a total number of words above a preset figure. The choice of answer recognition scheme depended on the type of question that would be asked and the consideration of whether the following segments of the program necessarily required that the subjects be in an appropriate predesignated conceptual framework before proceeding in the program. The response recognition options available within COMPUTEST were found amply sufficient to handle the different answer schemes needed in the programs. The other requirements of text presentation, scoring, and branching were all easily accomplished within the normal code commands and options of COMPUTEST.

The thesis-author found it extremely valuable to be able to code the training materials himself as this led to new approaches in structuring the materials which would not have been devised without this direct interface of the author with the computer. Also, the author did sometimes encounter special problems in trying to present and code materials which could not be handled by the present language. When this happened he was able to describe to the computer project programmer what he desired to build into the system, such as a method of counting the total number of words found in the input and testing that against a predesignated number. This can be considered an important outcome of having author-instructors directly involved in coding their own programs. From this would result a more practically efficient and sophisticated computer author language. For as Zinn writes, "In the evolution of such a system a computer programmer works with the authors to implement each new request for system capability, but it is his purpose to program himself out of the system by generalizing each function that might be repeated in slightly different ways and by different authors" (1967, p. 81).

Thus COMPUTEST fulfilled Zinn's second criterion of permitting the user to utilize the entire capabilities of the computer system and to develop as complex a program procedure as needed. As stated before, this finding would be of limited consequence were it not for the fact that an analogous language, PILOT, is being built for the IBM 360 system which will include all the features of COMPUTEST and still maintain the user-simplicity feature of COMPUTEST. In summary, the COMPUTEST language proved more than adequate to handle all the programming problems involved in developing effective and efficient CAI programs for training the students in productive problem-solving skills and attitudes.

V. SUMMARY

The overall purpose of this thesis was to investigate the effectiveness and efficiency of specially developed computer-assisted instructional (CAI) materials for the training of productive problem-solving skills and attitudes. The basic assumption was that such materials make possible a more interactive training situation, involving open-ended questioning and feedback, and hence have the potentiality of eventually leading to a truly individualized instructional setting for the training of these thinking skills. The basic pedagogical rationale concerning the structure and content of the training materials, and the underlying theoretical assumptions concerning the problem-solving skills to be trained for, were derived from the work of Crutchfield and Covington (1965). This formed the basis for the methed of investigating the question of the <u>effectiveness</u> of the CAI training materials developed. The question of the <u>efficiency</u> of the CAI materials developed was mainly one of determining the feasibility of utilizing a special computer language, COMPUTEST, which had been recently developed by Starkweather (1965). This question of efficiency centered around two main points of user-simplicity of the COMPUTEST language and its potentiality for permitting full utilization of the computer system.

The experimental design consisted of seven different training conditions, completely crossed over the two categories of sex and two levels of IQ (above and below the class mean of 113). This design was completely replicated within six randomly chosen eighth grade classes with a final total of 168 subjects. Of the seven groups, three groups constituted the major computer-training conditions. One computer group worked on selected lessons from the General Problem Solving series (Covington, Crutchfield, and Davies, 1966) in a computer-simulated teletype situation. A second computer group worked on specially developed computer programs consisting of "fictitious" American history materials which were presented on an IBM 1620 computer. A third computer group received the computer-training materials of both the preceding two groups. The other four groups consisted of three "active control" comparison booklet groups and an untutored group of subjects. The training period lasted for approximately three weeks and consisted of eight to ten 40-minute periods of instruction for most subjects. Two periods of specially developed posttests yielding measures of problem-solving performance and measures of various relevant attitudes were given at the completion of the training period.

Statistical analyses were carried out by means of the planned comparisons method, with separate analyses being performed on so-called Middle IQ subjects (mean=104) and so-called High IQ subjects (mean=124). The most consistent and statistically significant performance finding was that the Middle IQ subjects who worked on both types of computer-training materials outperformed every other experimental group of both the Middle and the High IQ subjects in the three main problem-solving functions of Problem-Formulation, Idea-Generation, and Idea-Evaluation. These subjects also rated the posttest problems as positive to a significantly greater degree. This superior performance was primarily attributed to the beneficial interaction effects of the two different types of computer-training materials; a possible alternative explanation, namely, the lengthened time of involvement in the training period was not sufficient to account for the results obtained.

The three computer groups, especially the computer history subjects, all positively rated their training materials significantly higher than did the other comparison booklet groups. This was true for both the Middle and High IQ subjects. The basis of this favorable rating was the problem-solving nature of the training materials. This was interpreted as resulting from the intrinsically attractive interactive nature of the computer-training materials; an alternative explanation in terms of a "Hawthorne" effect of the novelty of the computer was not substantiated by the facts of the study. Positive changes in problem-solving attitudes having to do with approaching a problem through more than one idea, persistence in working on difficult problems, and higher self-evaluation of oneself as a problem solver were almost all statistically significant in favor of the computer-instructed subjects for both the Middle and High IQ subjects.

The second major question of the study on the efficiency and effectiveness of using the COMPUTEST (and the analogous PILOT) computer language for the development of these CAI productive thinking materials was considered to be positively answered on two main grounds: First, the ease of development of the history computer programs by the author who had no previous programming experience; second, the fact that these computer programs were sophisticated enough to establish a valid instructional situation, as evidenced by the performance and attitude gains on the posttests for certain of the computer-instructed groups.

VI. FOOTNOTES

1. The complete 16 lessons of The Productive Thinking Program, Series One: General Problem Solving (Covington, M.V., R.S. Crutchfield and L.B. Davies, 1966) may be purchased from Educational Innovation, P.O. Box 9248, Berkeley, California.

2. COMPUTEST has been originally developed for use on the 20K IBM 1620 Models I or II with a 1311 Disk Drive, Model 3. Development is currently nearing completion in building an anlogous author language called PILOT (Programmed Inquiry, Learning Or Teaching) for use on the IBM 360 system. PILOT will fulfill the same basic programming functions as COMPUTEST and will also permit the use of much more elegant and sophisticated programming techniques while, at the same time, not losing the user-simplicity features found in COMPUTEST.

3. All cluster analyses performed in connection with this thesis utilized the BC TRY system of computer programs developed under the direction of R.C. Tryon and made available for use by the Computer Center, University of California, Berkeley.

VII. REFERENCES

Atkinson, R.C. Computerized Instruction and the Learning Process. Technical Report No. 122. Stanford, Calif.: Psychology Series: Institute for Mathematical Studies in the Social Sciences, Stanford University, September, 1967 (a).

Atkinson, R.C. CAI-learning aspects. In R.W. Gerard (Ed.), Computers and Education: A Workshop Conference at University of California, Irvine. New York: McGraw-Hill, 1967, 11-64 (b).

Bitzer, D.L., E.R. Lyman and J.A. Easley. The uses of PLATO: a computer controlled teaching system. Audiovisual Instruction, 1966, 11, 16-21.

Bitzer, D.L., B.L. Hicks, R.L. Johnson and E.R. Lyman. The PLATO system: Current research and developments. IEEE Trans. Human Factors in Electronics, 1967, HFE-8, 64-70.

Blount, N.S., H.J. Klausmeier, S.L. Johnson, W.C. Fredrick and J.G. Ramsay. The Effectiveness of Programed Materials in English Syntax and the Relationship of Selected Variables to the Learning of Concepts. Technical Report No. 17. Madison, Wisconsin: Research and Development Center for Learning and Re-education: University of Wisconsin, January, 1967.

Braunfeld, P.G. and L.D. Fosdick. The use of an automatic computer system in teaching. IEEE Trans. on Education, 1962, E-5, 156-167.

Carter, L. and H.F. Silberman. The Systems Approach, Technology and the School. Document SP-2025. Santa Monica, Calif.: System Development Corporation, April, 1965.

Cogswell, J.F. and D.P.Estavan. Explorations in Computer-Assisted Counseling. Document TM-2582. Santa Monica, Calif.: System Development Corporation, August, 1965.

Coulson, J.E. A computer-based laboratory for research and development in education. In J.E. Coulson (Ed.), Programmed Learning and Computer-Based Instruction. New York: Wiley, 1962, 191-204.

Coulson, J.E. Present Status and Future Prospects of Computer-Based Instruction. Document SP-1629. Santa Monica, Calif.: System Development Corporation, April, 1964.

Coulson, J.E., D.P. Estavan, R.J. Melaragno and H.F. Silberman. Effects of Branching in a Computer Controlled Auto-Instructional Device. Document TM-617. Santa Monica, Calif.: System Development Corporation, May, 1961.

Coulson, J.E., R.J. Melaragno and H.F. Silberman. Non-program Variables in the Application of Programmed Instruction. Document TM-2176. Santa Monica, Calif.: System Development Corporation, July, 1965.

Covington, M.V. The effectiveness of training for problem-solving efficiency and creative thinking as a function of differing ability levels among children. Paper read at Western Psychological Association, Honolulu, June, 1965.

Covington, M.V. Programmed instruction for the intellectually gifted: Experiments in productive thinking. In Special Education: Strategies for Educational Progress, Proceedings of the International Council for Exceptional Children, Toronto, April, 1966, 9-17 (a).

Covington, M.V. A childhood attitude inventory for problem solving. J. educ. Meas., 1966, 3, 234 (b).

Covington, M.V. New directions in the appraisal of creative thinking potential. J. educ. Meas., in press.

Covington, M.V. and R.S. Crutchfield. Facilitation of creative problem solving. Programed Instruction, 1965, 4(4), 3-5, 10.

Covington, M.V., R.S. Crutchfield, and Lillian B. Davies. The Productive Thinking Program, Series One: General Problem Solving. Berkeley, California: Brazelton Printing Co., 1966, 633 pp.

Crutchfield, R.S. Summary Report for the Period May 1961 - June 1964 to the Carnegie Corporation of New York on Research Activities under Its Grant for a Study of Auto-Instructional Methods and Creative Thinking. Berkeley, Calif.: Psychology Dept., Univ. of Calif., September, 1964. (Mimeographed)

Crutchfield, R.S. Instructing the individual in creative thinking. In New Approaches to Individualizing Instruction. Princeton, New Jersey: Educational Testing Service, 1965, 13-26.

Crutchfield, R.S. Creative thinking in children: Its teaching and testing. In Brim, H., R.S. Crutchfield, and W. Holtzman, Intelligence: Perspectives 1965. New York: Harcourt, Brace and World, 1966, 33-64.

Crutchfield, R.S. and M.V. Covington. Programed instruction and creativity. Programed Instruction, 1965, 4(4), 1, 2, 8-10.

Crutchfield, R.S. and M.V. Covington. Annual Report to the Carnegie Corporation of New York on Research Activities under Its Grant for a Study of Auto-Instructional Methods and Creative Thinking. Berkeley, Calif.: Psychology Dept., Univ. of Calif., June, 1967. (Mimeographed)

Duncker, K. A qualitative (experimental and theoretical) study of productive thinking (solving of comprehensible problems). J. genet. Psychol., 1926, 33, 642-708.

Duncker, K. On problem solving. Psychol. Monogr., 1945, 58 (Whole No. 270), 1-113.

Fenton, E. Teaching the New Social Studies in Secondary Schools: An Inductive Approach. New York: Holt, Rinehart, and Winston, 1966.

Grubb, R.E. CAI-technical aspects. In R.W. Gerard (Ed.), Computers and Universities. [Initial version of later revised edition published as Computers and Education, R.W. Gerard (Ed.)], 1967 (a).

Grubb, R.E. CAI-technical aspects. In R.W. Gerard (Ed.), Computers and Education: A Workshop Conference at University of California, Irvine. New York: McGraw-Hill, 1967, 65-110 (b).

Guilford, J.P. The structure of intellect. Psychol. Bull., 1956, 53, 267-293.

Hays, W.L. Statistics for Psychologists. New York: Holt, Rinehart, and Winston, 1963.

Hodge, C. A chatty computer in room B-3: COMPUTEST on IBM 1620 used in a California grade school. EDUCOM, 1966, 1 (No. 3), 3-7.

Leonard, J.M. and R.L. Wing. Advantages of using a computer in some kinds of educational games. IEEE Trans. Human Factors in Electronics, 1967, HFE-8, 75-81.

Licklider, J.C.R. Preliminary experiments in computer-aided teaching. In J.E. Coulson (Ed.), Programmed Learning and Computer-Based Instruction. New York: Wiley, 1962, 217-239.

Lumsdaine, A.A. Some theoretical and practical problems in programmed instruction. In J.E. Coulson (Ed.), Programmed Learning and Computer-Based Instruction. New York: Wiley, 1962, 134-151.

Lyman, E.R. A Descriptive List of PLATO Programs. Report R-296 June, 1966 (Revised July, 1967). Urbana, Ill.: Coordinated Science Laboratory, Univ. of Illinois, 1967.

Maier, N.R.F. Reasoning in humans. I: On direction. J. comp. Psychol., 1930, 10, 115-143.

Maier, N.R.F. An aspect of human reasoning. Brit. J. Psychol., 1933, 24, 144-155.

Mitzel, H.E., G.L. Brandon, K.H. Wodtke, D.C. Bjorkquist, D.A. Gilman, and J.K. Ritchey. Experimentation with Computer-Assisted Instruction in Technical Education: Semi-Annual Progress Report. Project No. 5-85-074. University Park, Penna.: The Pennsylvania State Univ. Computer Assisted Instructional Laboratory, 1966.

Moncrieff, B. The Sumerian Game: teaching economics with a computerized P.I. program. Programed Instruction, 1965, 4, 10-11.

Olton, R.M., J. Wardrop, M.V. Covington, W. Goodwin, R.S. Crutchfield, H.J. Klausmeier, and T. Ronda. The Development of Productive Thinking Skills in Fifth-grade Children. Technical Report. Madison, Wisconsin:

Research and Development Center for Cognitive Learning, The University of Wisconsin, in press.

Osborn, A.F. Applied imagination (3rd ed.). New York: Scribner's, 1957.

Parnes, S.J. Can creativity be increased? In S.J. Parnes and H.F. Harding (Eds.), A Source Book for Creative Thinking. New York: Scribner's, 1962, 185-191.

Parnes, S.J. Research on developing creative behavior. In C.W. Taylor (Ed.), Widening Horizons in Creativity. New York: Wiley, 1964, 145-169.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

Starkweather, J.A. COMPUTEST: a computer language for individual testing, instruction, and interviewing. Psychol. Repts., 1965, 17, 227-237.

Starkweather, J.A. COMPUTEST IID: User's Guide. Technical Report. San Francisco, Calif.: Dixie Computer Learning Project, Univ. of Calif. Medical Center, 1967 (a).

Starkweather, J.A. Computer-assisted learning in medical education. Canad. Medical Assoc. Journal, 1967, 97, 733-738 (b).

Starkweather, J.A. Statement made in open discussion period during Session II: CAI: technical aspects. In R.W. Gerard (Ed.), Computers and Education: A workshop conference at University of California, Irvine. New York: McGraw-Hill, 1967, 95-96 (c).

Starkweather, J.A., M. Kamp, and A. Monto. Psychiatric interview simulation by computer. Methods of Information in Medicine, 1967, VI, 15-23.

Stolurow, L.M. Computer-Based Instruction. Technical Report No. 9. Urbana, Ill.: Training Research Laboratory, Univ. of Illinois, 1965.

Strum, R.D. and J.R. Ward. Some comments on computer-assisted instruction in engineering education. IEEE Trans. on Education, 1967, E-10, 1-3.

Suchman, J. Inquiry training in the elementary school. The sci. Teacher, 1960, 27, 42-47.

Suchman, J.P. Inquiry training: Building skills for autonomous discovery. Merrill-Palmer Quart., 1961, 7, 147-169.

Suppes, P. Computer-Assisted Instruction in the Schools: Potentialities, Problems, Prospects. Technical Report No. 81. Stanford, Calif.: Psychology Series: Institute for Mathematical Studies in the Social Sciences, Stanford Univ., October, 1965.

Suppes, P. Tomorrow's education? Computer-based instruction in the elementary school. Education Age, 1966, 2 (3), (a).

Suppes, P. The uses of computers in education. Scient. Amer., 1966, 215, 206-221 (b).

Suppes, P. On using computers to individualize instruction. In D.D. Bushnell and D.W. Allen (Eds.), The Computer in American Education. New York: Wiley, 1967, 11-24.

Suppes, P., M. Jerman, and G. Groen. Arithmetic drills and review on a computer-based teletype. The arith. Teacher, 1966, 13, 303-309.

Taylor, C.W. (Ed.). Creativity: Progress and Potential. New York: McGraw-Hill, 1964 (a).

Taylor, C.W. (Ed.). Widening Horizons in Creativity. New York: Wiley, 1964 (b).

Taylor, C.W. and F. Barron (Eds.). Scientific Creativity: Its Recognition and Development. New York: Wiley, 1963.

Torrance, E.P. Guiding Creative Talent. Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Torrance, E.P. Education and the Creative Potential. Minneapolis, Minn.: University of Minnesota Press, 1963.

Torrance, E.P. Rewarding Creative Behavior: Experiments in Classroom Creativity. Englewood Cliffs, N.J.: Prentice-Hall, 1965.

Tryon, R.C. and D.E. Bailey. The BC TRY computer system of cluster and factor analysis. Multivariate behavior Research, 1966, 1, 95-111.

Uttal, W.R. On conversational interaction. In J.E. Coulson (Ed.), Programmed Learning and Computer-Based Instruction. New York: Wiley, 1962, 171-190.

Wertheimer, M. Productive Thinking. New York: Harper, 1945.

Zinn, K.L. Computer assistence for instruction: A review of systems and projects. In D.D. Bushnell and D.W. Allen (Eds.), The Computer in American Education. New York: Wiley, 1967, 77-107.

TABLE 1

Sample Segment of Question-and-Answer Sequence in COMPUTEST Language Program Written By an Elementary School Pupil; With A Sample Protocol Printout Sheet

Actual COMPUTEST program with computer code letters

```
C    Who discovered America. . . .
R    Eric, Red, Viking
G    You sure know your history.
B    Nope, you forgot about the Vikings.
A

C    Who is the father of our country. . .
R    George Washington
G    Of course.
B    Were you really born in this country. . .
A    match 2, rjump to XI
```

(Note: match 2 is a special option which requires the student to put down both words to be considered a correct answer; rjump is a COMPUTEST computer code which will jump the student to segment XI (below) if he has a right answer.)

```
C    Come on...Put down his first and last name.
R    George Washington
G    That's better.
B    You must have just come over on the boat, it's George Washington--
B    you Dumbell.
A

CXI  When did the American Revolution begin. . .
```

(etc.)

Sample Printout Protocol

```
Who discovered America. . . .
          I think it was Chris Columbus RS
Nope, you forgot about the Vikings.
Who is the father of our country. . .
          Good old george...RS
Were you really born in this country. . .
Come on. . .Put down his first and last name.
          George Washington, I cannot tell a lie.  RS
That's better.
When did the American Revolution begin. . . .
```

(etc.)

TABLE 2

Experimental Design Diagram of the Completely-Crossed Factors of Experimental Groups (7 levels), IQ (2 levels), and Sex (2 levels)

Note: This design is completely replicated within each of the six class homerooms used in this experiment, therefore there are 6 subjects in each cell block.

Experimental Groups \ IQ / Sex	Below-Mean		Above-Mean		No. Ss.
	Males	Females	Males	Females	
COMPUTER-GPS	6	6	6	6	24
COMPUTER-PH	6	6			24
COMBINED-COMPUTER	6				24
BOOKLET-GPS					24
BOOKLET-HR				6	24
BOOKLET-HR + GUIDES			6	6	24
UNTUTORED-PT	6	6	6	6	24
No. Ss.	42	42	42	42	168

TABLE 3

Overall and Individual Class Medians of IQ Scores for: All Subjects, Middle IQ, High IQ, Males, and Females

		All Ss.	Middle IQ	High IQ	Males	Females
CLASS	1	112.5	106.0	124.0	112.5	113.0
	2	117.0	104.5	125.5	119.0	114.0
	3	113.0	109.5	122.0	115.0	112.5
	4	115.0	107.5	120.5	114.0	115.0
	5	113.0	108.0	126.5	114.0	113.5
	6	114.5	107.5	124.0	116.0	114.5
OVERALL		114.0	108.0	123.0	115.0	113.5

TABLE 4

Median IQ Scores for Different Combinations of Experimental Groups in the Planned Comparisons Design for: All Subjects, Middle IQ, and High IQ

ALL SS

Group A		Group B	Group A	Group B	Average
1-6	Vs.	7	114.5	113.0	114.0
123	Vs.	456	115.0	114.0	114.5
12	Vs.	3	113.5	116.5	115.0
1	Vs.	2	112.5	114.0	113.5
4	Vs.	56	115.5	113.5	114.0
5	Vs.	6	115.5	113.5	113.5

MIDDLE IQ

Group A		Group B	Group A	Group B	Average
1-6	Vs.	7	108.0	103.5	108.0
123	Vs.	456	108.5	105.0	108.0
12	Vs.	3	108.0	110.0	108.0
1	Vs.	2	108.0	109.0	108.0
4	Vs.	56	102.5	105.0	105.0
5	Vs.	6	95.5	107.5	105.0

HIGH IQ

Group A		Group B	Group A	Group B	Average
1-6	Vs.	7	123.0	125.5	123.0
123	Vs.	456	121.0	126.0	123.0
12	Vs.	3	121.5	120.5	121.0
1	Vs.	2	120.5	123.0	121.5
4	Vs.	56	124.0	127.0	126.0
5	Vs.	6	129.5	119.5	127.0

TABLE 5

Results of Nonparametric Median Tests on IQ Scores of The Planned Comparison Combinations of Experimental Groups for: All Subjects, Middle IQ, and High IQ

Groups			ALL SS.	MIDDLE IQ	HIGH IQ
1-6	Vs.	7	.25	.16	.16
123	Vs.	456	.03	.51	2.73*
12	Vs.	3	.01	.13	.35
1	Vs.	2	.08	.67	.17
4	Vs.	56	1.74	.13	.68
5	Vs.	6	.08	1.50	4.29**

*Significant at .10 level

**Significant at .05 level

TABLE 6

Composite Cluster Variables for Problem-Formulation

(See text for explanation)

CLUSTER C1: Reliability = .95, $\overline{r}^2$ = .42*

		a)	b)	c)
Items	a)	--	.81	.72
	b)	.81	--	.94
	c)	.72	.94	--

CLUSTER C2: Reliability = .96, $\overline{r}^2$ = .59

		a)	b)	c)
Items	a)	--	.81	.68
	b)	.81	--	.79
	c)	.68	.79	--

The intercorrelation of the raw cluster scores is .33

*Note: Reliability = the reliability of the composite of the cluster definers (Spearman-Brown); $\overline{r}^2$ = reproducibility of the mean of the squared correlations among items.

TABLE 7

Composite Cluster Variables for Idea-Generation

(See text for explanation)

CLUSTER C3: Reliability = .82, $\overline{r}^2$ = .36

		a)	b)	c)	d)	e)
	a)	--	.47	.43	.42	.44
	b)	.47	--	.38	.44	.52
Items	c)	.43	.38	--	.46	.47
	d)	.42	.44	.46	--	.53
	e)	.44	.52	.47	.53	--

CLUSTER C4: Reliability = .71, $\overline{r}^2$ = .58

	a)	b)	c)
a)	-	.28	.51
b)	.28	--	.39
c)	.51	.39	--

CLUSTER C5: Reliability = .76, $\overline{r}^2$ = .75

The intercorrelation of the two items is .58

Raw cluster score correlation matrix

	C3	C4	C5
C3	--	.43	.39
C4	.43	--	.52
C5	.39	.52	--

TABLE 8

Composite Cluster Variables for Idea-Evaluation

(See text for explanation)

CLUSTER C7: Reliability = .88, $\bar{r}^2$ = .42

		a)	b)	c)
	a)	--	.57	.66
Items	b)	.57	--	.70
	c)	.66	.70	--

CLUSTER C8: Reliability = .78, $\bar{r}^2$ = .54

		a)	b)	c)
	a)	--	.49	.35
Items	b)	.49	--	.41
	c)	.35	.41	--

The intercorrelation of the raw cluster scores is .24

TABLE 9

Composite Cluster Variables for Problem-Evaluation

(See text for explanation)

CLUSTER C8: Reliability = .84, $\bar{r}^2$ = .41

		a)	b)	c)
Items	a)	--	.65	.57
	b)	.65	--	.62
	c)	.57	.62	--

CLUSTER C9: Reliability = .85, $\bar{r}^2$ = .34

		a)	b)	c)
Items	a)	--	.66	.60
	b)	.66	--	.63
	c)	.60	.63	--

CLUSTER C10: Reliability = .90, $\bar{r}^2$ = .36

		a)	b)	c)
Items	a)	--	.77	.66
	b)	.77	--	.78
	c)	.66	.78	--

CLUSTER C11: Reliability = .73, $\bar{r}^2$ = .28

		a)	b)	c)
Items	a)	--	.40	.49
	b)	.40	--	.47
	c)	.49	.47	--

Raw cluster score correlation matrix

	C8	C9	C10	C11
C8	--	.31	.30	.33
C9	.31	--	.46	.36
C10	.30	.46	--	.47
C11	.33	.36	.47	--

TABLE 10

Composite Cluster Variables for Training Materials Evaluation

(See text for explanation)

CLUSTER C12: Reliability = .86, $\bar{r}^2$ = .71

		a)	b)	c)	d)
Items	a)	--	.69	.56	-.55
	b)	.69	--	.64	-.60
	c)	.56	.64	--	-.55
	d)	-.55	-.60	-.55	--

CLUSTER C13: Reliability = .67, $\bar{r}^2$ = .21

The intercorrelation is -.46

The intercorrelation of the raw cluster scores is .33

TABLE 11

Composite Cluster Variables for Self-Evaluation

(See text for explanation)

CLUSTER C14: Reliability = .58, $\bar{r}^2$ = .26

		a)	b)	c)
	a)	--	.21	.22
Items	b)	.21	--	.43
	c)	.22	.43	--

CLUSTER C15: Reliability = .75, $\bar{r}^2$ = .26

		a)	b)	c)
	a)	--	.65	.36
Items	b)	.65	--	.40
	c)	.36	.40	--

CLUSTER C16: Reliability = .68, $\bar{r}^2$ = .40

		a)	b)	c)	d)	e)
	a)	--	.38	.37	-.33	.18
	b)	.38	--	.22	-.29	.28
Items	c)	.37	.22	--	-.15	.32
	d)	-.33	-.29	-.15	--	-.36
	e)	.18	.28	.32	-.36	--

Raw cluster score correlation matrix

	C14	C15	C16
C14	--	-.06	-.17
C15	-.06	--	.27
C16	-.17	.27	--

TABLE 12

Weights Used for Planned Comparisons Statistical Analyses

PLANNED COMPARISONS	EXPERIMENTAL GROUPS 1	2	3	4	5	6	7
I	1/6	1/6	1/6	1/6	1/6	1/6	-1
II	1/3	1/3	1/3	-1/3	-1/3	-1/3	0
III	1/2	1/2	-1	0	0	0	0
IV	1	-1	0	0	0	0	0
V	0	0	0	-1	1/2	1/2	0
VI	0	0	0	0	1	-1	0

TABLE 13

Comparison I: All Experimental Groups Versus the "Empty" Posttest-Control Group: Middle IQ Subjects.

Group Mean Scores of Composite Cluster and Additive Variables, F Statistics of Planned Comparisons, and Significance Levels are Presented.

(1 and 77 degrees of freedom)

Performance Variables	All Exp. Groups	Posttest Controls	F Statistics
Problem-Formulation			
C1 $	50.22	48.68	<1
C2	49.91	50.52	<1
Idea-Generation			
C3	49.44	53.38	1.62
C4	49.49	53.05	1.82
C5	49.92	50.50	<1
A1	9.54	12.76	1.12
A2	10.05	9.69	<1
A3	10.00	10.01	<1
Idea-Evaluation			
C6 $$	50.07	49.55	<1
C7 $$	49.66	52.04	<1
Attitude Variables			
Problem-Evaluation			
C8	48.99	56.07	6.29**
C9	49.52	52.86	1.12
C10	49.21	54.74	3.92*
C11	49.24	54.56	3.77*
Evaluation of Training Materials			
C12	50.09	49.45	<1
C13	51.05	43.71	8.49***
A4	9.90	10.58	5.73**
A5	10.02	9.88	<1
A6	10.09	9.49	4.02**
Self-Evaluation			
C14	49.18	54.95	4.39**
C15	49.74	51.55	<1
C16	50.08	49.52	<1
A7	10.15	9.13	3.98**
A8	10.05	9.73	<1

TABLE 13 (Continued)

$ See section III.2 for explanation of these alphanumeric variable code names

$$ These two variables have 1 and 49 degrees of freedom

*	Significant at .10 level with F statistic 2.78
**	Significant at .05 level with F statistic 3.97
***	Significant at .01 level with F statistic 7.02
****	Significant at .001 level with F statistic 11.80

#	Significant at .10 level with F statistic 2.82
##	Significant at .05 level with F statistic 4.04
###	Significant at .01 level with F statistic 7.21
####	Significant at .001 level with F statistic 12.32

TABLE 14

Comparison I: All Experimental Groups Versus the "Empty" Posttest-Control Group: High IQ Subjects.

(1 and 77 degrees of freedom)

Performance Variables	All Exp. Groups	Posttest Controls	F Statistics
Problem-Formulation			
C1 $	50.14	49.14	<1
C2	50.05	49.73	<1
Idea-Generation			
C3	50.57	46.61	2.52
C4	50.19	48.89	<1
C5	50.07	49.58	<1
A1	10.15	9.08	<1
A2	10.46	7.23	1.26
A3	10.11	9.32	2.30
Idea-Evaluation			
C6 $$	49.22	54.67	5.43 ##
C7 $$	51.60	40.41	14.55 ####
Attitude Variables			
Problem-Evaluation			
C8	49.94	50.39	<1
C9	50.37	47.79	<1
C10	50.18	48.94	<1
C11	50.32	48.11	<1
Evaluation of Training Materials			
C12	50.92	44.45	5.40 **
C13	52.29	36.73	50.46 ****
A4	9.98	10.13	<1
A5	10.07	9.61	2.99 *
A6	10.13	9.25	8.04 ***
Self-Evaluation			
C14	48.91	56.52	7.16 ***
C15	51.24	42.55	9.48 ***
C16	50.10	49.38	<1
A7	10.15	9.11	3.69 *
A8	10.18	8.93	1.60

(Refer to TABLE 13 for explanation of symbols)

TABLE 15

Comparison II: Computer Versus Comparison-Booklet Groups: Middle IQ Subjects.

(1 and 77 degrees of freedom)

Performance Variables	Computer Groups	Booklet Groups	F Statistics
Problem-Formulation			
C1 $	50.75	49.69	< 1
C2	50.88	48.94	< 1
Idea-Generation			
C3	51.28	47.59	2.49
C4	51.30	47.68	3.30 *
C5	51.40	48.53	1.97
A1	11.71	7.37	3.53 *
A2	11.22	8.89	1.09
A3	10.26	9.73	1.71
Idea-Evaluation			
C6 $$	49.94	50.21	< 1
C7 $$	51.46	47.86	2.39
Attitude Variables			
Problem-Evaluation			
C8	49.90	48.08	< 1
C9	48.96	50.08	< 1
C10	51.95	46.47	6.75 **
C11	50.19	48.30	< 1
Evaluation of Training Materials			
C12	53.57	46.62	10.91 ***
C13	54.90	47.20	16.32 ****
A4	9.69	10.11	3.76 *
A5	10.27	9.77	7.44 ***
A6	10.29	9.88	3.42 *
Self-Evaluation			
C14	48.18	50.17	< 1
C15	51.83	47.65	3.85 *
C16	51.66	48.50	2.00
A7	10.69	9.60	7.96 ***
A8	10.87	9.23	5.57 **

(Refer to TABLE 13 for explanation of symbols)

TABLE 16

Comparison II: Computer Versus Comparison-Booklet Groups: High IQ Subjects.

(1 and 77 degrees of freedom)

Performance Variables	Computer Groups	Booklet Groups	F Statistic
Problem-Formulation			
C1 $	49.43	50.86	<1
C2	48.94	51.15	<1
Idea-Generation			
C3	49.83	51.30	<1
C4	50.50	49.87	<1
C5	49.01	51.13	<1
A1	9.53	10.78	<1
A2	9.81	11.12	<1
A3	10.16	10.07	<1
Idea-Evaluation			
C6 $$	47.73	50.72	2.86 #
C7 $$	53.91	49.29	4.34 ##
Attitude Variables			
Problem-Evaluation			
C8	50.62	49.25	<1
C9	49.79	50.95	<1
C10	50.78	49.58	<1
C11	50.46	50.17	<1
Evaluation of Training Materials			
C12	55.46	46.39	18.62 ****
C13	55.73	48.85	16.24 ****
A4	9.77	10.19	4.70 **
A5	10.25	9.89	3.24 *
A6	10.47	9.78	8.87 ***
Self-Evaluation			
C14	47.85	49.97	<1
C15	53.14	49.35	3.16 *
C16	52.60	47.61	4.82 **
A7	10.52	9.77	3.35 *
A8	11.18	9.18	7.18 ***

(Refer to TABLE 13 for explanation of symbols)

TABLE 17

Comparison III: Combined-Computer Versus Other Computer Groups: Middle IQ Subjects.

(1 and 77 degrees of freedom)

Performance Variables	Combined-Computer Group	Other Computer Groups	F Statistic
Problem-Formulation			
C1 $	54.57	48.84	3.82 *
C2	55.07	48.80	3.84 *
Idea-Generation			
C3	51.69	51.08	<1
C4	58.64	47.63	13.57 ****
C5	55.85	49.18	4.42 **
A1	18.35	8.39	8.29 ***
A2	15.36	9.15	3.43 *
A3	11.01	9.89	3.44 *
Idea-Evaluation			
C6 $$	46.10	51.85	3.80 #
C7 $$	58.10	48.54	8.12 ###
Attitude Variables			
Problem-Evaluation			
C8	53.61	48.05	3.01 *
C9	52.11	47.39	1.73
C10	55.64	50.11	3.05 *
C11	55.96	47.30	7.77 ***
Evaluation of Training Materials			
C12	54.77	52.97	<1
C13	58.07	53.31	2.77
A4	9.75	9.67	<1
A5	10.88	9.96	11.11 ***
A6	10.32	10.28	<1
Self-Evaluation			
C14	44.36	50.10	3.38 *
C15	52.98	51.26	<1
C16	54.30	50.34	1.39
A7	10.96	10.55	<1
A8	12.14	10.23	3.39 *

(Refer to TABLE 13 for explanation of symbols)

TABLE 18

Comparison III: Combined-Computer Versus Other Computer Groups: High IQ Subjects.

(1 and 77 degrees of freedom)

Performance Variables	Combined-Computer Group	Other Computer Groups	F Statistic
Problem-Formulation			
C1 $	50.14	49.07	<1
C2	48.38	49.17	<1
Idea-Generation			
C3	50.25	49.62	<1
C4	51.12	50.19	<1
C5	51.28	47.89	1.04
A1	11.00	8.79	<1
A2	12.14	8.54	1.15
A3	10.66	9.91	1.60
Idea-Evaluation			
C6 $$	50.57	46.31	2.58
C7 $$	52.17	54.78	<1
Attitude Variables			
Problem-Evaluation			
C8	48.69	51.59	<1
C9	48.66	50.35	<1
C10	50.36	50.99	<1
C11	51.41	49.99	<1
Evaluation of Training Materials			
C12	55.10	55.65	<1
C13	56.34	55.43	<1
A4	9.88	9.71	<1
A5	10.19	10.27	<1
A6	10.25	10.58	<1
Self-Evaluation			
C14	47.96	47.80	<1
C15	51.09	54.16	<1
C16	53.91	51.94	<1
A7	10.36	10.61	<1
A8	11.43	11.05	<1

(Refer to TABLE 13 for explanation of symbols)

TABLE 19

Comparison IV: Computer-GPS Versus Computer-PH Group: Middle IQ Subjects.

(1 and 77 degrees of freedom)

Performance Variables	Computer-GPS Group	Computer-PH Group	F Statistic
Problem-Formulation			
C1 $	50.42	47.26	<1
C2	48.21	49.39	<1
Idea-Generation			
C3	53.41	48.75	1.32
C4	48.05	47.22	<1
C5	50.95	47.41	<1
A1	9.35	7.43	<1
A2	11.71	6.52	1.84
A3	10.26	9.51	1.15
Idea-Evaluation			
C6 $$	52.94	50.76	<1
C7 $$	48.58	47.69	<1
Attitude Variables			
Problem-Evaluation			
C8	52.46	43.63	5.69 **
C9	49.66	45.12	1.20
C10	49.62	50.60	<1
C11	49.89	44.70	2.09
Evaluation of Training Materials			
C12	52.21	53.73	<1
C13	54.40	52.23	<1
A4	9.50	9.83	<1
A5	10.13	9.80	1.10
A6	10.57	9.99	2.23
Self-Evaluation			
C14	48.51	51.69	<1
C15	53.04	49.48	<1
C16	53.91	46.77	3.40 *
A7	10.88	10.21	1.00
A8	11.23	9.23	2.76

(Refer to TABLE 13 for explanation of symbols)

TABLE 20

Comparison IV: Computer-GPS Versus Computer-PH Group: High IQ Subjects.

(1 and 77 degrees of freedom)

Performance Variables	Computer-GPS Group	Computer-PH Group	F Statistic
Problem-Formulation			
C1 $	46.53	51.61	1.35
C2	48.38	49.96	<1
Idea-Generation			
C3	49.63	49.62	<1
C4	49.05	51.33	<1
C5	47.03	48.72	<1
A1	8.08	9.50	<1
A2	6.39	10.89	1.42
A3	9.74	10.07	<1
Idea-Evaluation			
C6 $$	47.14	45.48	<1
C7 $$	53.31	56.25	<1
Attitude Variables			
Problem-Evaluation			
C8	50.32	52.85	<1
C9	50.29	50.41	<1
C10	47.95	54.02	2.77
C11	47.17	52.81	2.44
Evaluation of Training Materials			
C12	55.09	56.20	<1
C13	56.96	53.90	1.07
A4	9.71	9.71	<1
A5	10.02	10.52	2.07
A6	11.08	10.08	6.13 **
Self-Evaluation			
C14	48.69	46.91	<1
C15	53.08	55.24	<1
C16	51.96	51.92	<1
A7	10.44	10.77	<1
A8	10.68	11.43	<1

(Refer to TABLE 13 for explanation of symbols)

TABLE 21

Comparison V: Booklet-GPS Versus History-Booklet Groups: Middle IQ Subjects.

(1 and 77 degrees of freedom)

Performance Variables	Booklet-GPS Group	History-Booklet	F Statistic
Problem-Formulation			
C1 $	48.46	50.30	<1
C2	47.66	49.58	<1
Idea-Generation			
C3	45.55	48.72	<1
C4	45.05	49.00	1.72
C5	47.50	48.90	<1
A1	5.93	8.10	<1
A2	5.86	10.40	1.84
A3	9.01	10.10	3.19*
Idea-Evaluation			
C6 $$	48.36	51.14	<1
C7 $$	48.43	47.57	<1
Attitude Variables			
Problem-Evaluation			
C8	48.61	47.81	<1
C9	52.28	48.99	<1
C10	47.67	45.87	<1
C11	49.85	47.52	<1
Evaluation of Training Materials			
C12	50.11	44.87	2.76
C13	53.18	44.21	9.84 ***
A4	9.92	10.21	<1
A5	10.13	9.59	3.88 *
A6	10.07	9.78	<1
Self-Evaluation			
C14	46.23	52.14	3.58 *
C15	50.23	46.37	1.46
C16	49.94	47.78	<1
A7	10.13	9.34	1.88
A8	10.39	8.64	2.82 *

(Refer to TABLE 13 for explanation of symbols)

TABLE 22

Comparison V: Booklet-GPS Versus History-Booklet Groups: High IQ Subjects.

(1 and 77 degrees of freedom)

Performance Variables	Booklet-GPS Group	History-Booklet	F Statistic
Problem-Formulation			
C1 $	54.01	49.29	1.55
C2	49.03	52.22	<1
Idea-Generation			
C3	53.42	50.42	1.27
C4	51.84	48.89	<1
C5	51.32	51.04	<1
A1	12.42	9.96	<1
A2	11.06	11.14	<1
A3	9.99	10.11	<1
Idea-Evaluation			
C6 $$	51.52	50.31	<1
C7 $$	49.64	49.11	<1
Attitude Variables			
Problem-Evaluation			
C8	50.35	48.70	<1
C9	50.18	51.34	<1
C10	47.78	50.47	<1
C11	54.72	47.89	4.78 **
Evaluation of Training Materials			
C12	47.85	45.65	<1
C13	53.73	46.41	8.16 ***
A4	10.13	10.21	<1
A5	10.52	9.57	10.15 ***
A6	9.42	9.96	2.40
Self-Evaluation			
C14	44.30	52.81	6.97 **
C15	49.61	49.21	<1
C16	50.99	55.92	2.21
A7	11.11	9.11	10.57 ***
A8	10.68	8.43	4.04 **

(Refer to TABLE 13 for explanation of symbols)

TABLE 23

Comparison VI: Booklet-HR Versus Booklet-HR & Guides Group: Middle IQ Subjects.

(1 and 77 degrees of freedom)

Performance Variables	Booklet-HR Group	Booklet-HR & Guides	F Statistic
Problem-Formulation			
C1 $	49.41	51.19	<1
C2	50.90	48.26	<1
Idea-Generation			
C3	51.25	45.98	1.69
C4	49.50	48.49	<1
C5	48.81	48.99	<1
A1	7.60	8.60	<1
A2	12.94	7.86	1.73
A3	10.43	9.76	<1
Idea-Evaluation			
C6 $$	53.75	48.52	2.35
C7 $$	44.71	50.44	2.01
Attitude Variables			
Problem-Evaluation			
C8	45.73	49.89	1.26
C9	51.29	46.69	1.24
C10	42.64	49.10	3.13 *
C11	46.10	48.93	<1
Evaluation of Training Materials			
C12	43.28	46.46	<1
C13	44.99	43.43	<1
A4	10.25	10.17	<1
A5	9.55	9.63	<1
A6	10.07	9.49	2.23
Self-Evaluation			
C14	56.98	47.29	7.23 ***
C15	48.87	43.86	1.84
C16	44.12	51.43	3.56*
A7	8.88	9.80	1.89
A8	8.48	8.81	<1

(Refer to TABLE 13 for explanation of symbols)

TABLE 24

Comparison VI: Booklet-HR Versus Booklet-HR & Guides Group: High IQ Subjects.

(1 and 77 degrees of freedom)

Performance Variables	Booklet-HR Group	Booklet-HR & Guides	F Statistic
Problem-Formulation			
C1 $	47.76	50.82	<1
C2	54.35	50.08	1.11
Idea-Generation			
C3	51.44	49.03	<1
C4	47.11	50.67	1.05
C5	50.87	51.21	<1
A1	8.92	11.00	<1
A2	11.39	10.89	<1
A3	10.32	9.91	<1
Idea-Evaluation			
C6 $$	50.43	50.20	<1
C7 $$	48.54	49.68	<1
Attitude Variables			
Problem-Evaluation			
C8	50.35	47.04	<1
C9	54.89	47.79	3.81 *
C10	51.91	49.03	<1
C11	49.53	46.25	<1
Evaluation of Training Materials			
C12	48.84	42.47	3.06 *
C13	46.94	45.89	<1
A4	10.13	10.30	<1
A5	9.69	9.44	<1
A6	10.17	9.75	1.06
Self-Evaluation			
C14	52.20	53.42	<1
C15	53.09	45.34	4.40 **
C16	46.45	45.40	<1
A7	8.77	9.44	<1
A8	9.26	7.60	1.66

(Refer to TABLE 13 for explanation of symbols)