AD-785 900

COMPARISON OF TWO TEACHING STRATEGIES IN COMPUTER-ASSISTED INSTRUCTION

BOLT BERANEK AND NEWMAN, INCORPORATED

PREPARED FOR
OFFICE OF NAVAL RESEARCH
ADVANCED RESEARCH PROJECTS AGENCY

SEPTEMBER 1974

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: Technical Report No. 3	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER: AD 785900
4. TITLE (and Subtitle): COMPARISON OF TWO TEACHING STRATEGIES IN COMPUTER-ASSISTED INSTRUCTION		5. TYPE OF REPORT & PERIOD COVERED: Semi-Annual (1 Feb 74 - 31 Jul 74)
		6. PERFORMING ORG. REPORT NUMBER: BBN Report No. 2885
7. AUTHOR(S): Allan M. Collins		8. CONTRACT OR GRANT NUMBER(S): No. N00014-71-C-0228
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Bolt Beranek and Newman Inc. 50 Moulton St., Cambridge, Mass.		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61153N RR042-06;PR042-06-01 NR154-330
11. CONTROLLING OFFICE NAME AND ADDRESS: Personnel and Training Research Programs Office of Naval Research (Code 458) Arlington, VA 22217		12. REPORT DATE: September 1974
		13. NUMBER OF PAGES: 42
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report): Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

To be submitted to Educational Journal

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

CAI, teaching methods, education, learning theory

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Three experiments were run using the SCHOLAR CAI system to teach geography to high-school students The experiments compared a method of teaching derived from analysis of human tutors (Tutorial Mode) vs. a method derived from programmed instruction (Block-Test Mode). In the three experiments, Block-Test Mode was systematically converged toward Tutorial Mode in order to pinpoint what aspects of teaching strategy affected

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (cont).
student's learning. Tutorial Mode was significantly more effective in the first two experiments, and nonsignificantly in the third. The results indicated that the major factor affecting student's learning was the strategy that tutors use of reviewing the material in greater depth on a second pass. Allowing the students to ask questions, and the tutorial strategy for relating new material to the student's previous knowledge contributed only a small amount to the differences found in the first two experiments.

COMPARISON OF TWO TEACHING STRATEGIES IN COMPUTER-ASSISTED INSTRUCTION

Allan M. Collins

Bolt Beranek and Newman Inc.

Cambridge, Massachusetts 02138

Contract No. N00014-71-C-0228, dated March 1, 1971
Amendment Modification No. P00002, dated January 1, 1973
Expiration Date, June 30, 1975
Total Amount of Contract $586,000
Principal Investigator, Allan M. Collins [(617) 491-1850]

Sponsored by: Office of Naval Research
Contract Authority No. NR 154-330
Scientific Officers: Dr. Marshall Farr and
Dr. Joseph Young

and

Advanced Research Projects Agency
ARPA Order No. 2284, dated August 30, 1973
Program Code No. 61101E

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This research was sponsored by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-71-C-0228, Contract Authority Identification Number, NR No. 154-330. Reproduction in whole or in part is permitted for any purpose of the United States Government.

I would like to thank our contract monitors, Dr. Marshall Farr and Dr. Joseph Young, who have contributed many ideas to our research on teaching methods. I would also like to thank Joseph J. Passafiume, Nelleke Aiello, Eleanor H. Warnock, and Susan M. Graesser who programmed the various teaching strategies in SCHOLAR, ran the subjects, and analyzed the data. Finally, I would like to thank Dr. Thomas J. Triggs who developed the experimental design and Dr. Richard W. Pew for his comments on a draft of this report.

ABSTRACT

Three experiments were run using the SCHOLAR CAI system to teach geography to high-school students. The experiments compared a method of teaching derived from analysis of human tutors (Tutorial Mode) vs. a method derived from programmed instruction (Block-Test Mode). In the three experiments, Block-Test Mode was systematically converged toward Tutorial Mode in order to pinpoint what aspects of teaching strategy affected student's learning. Tutorial Mode was significantly more effective in the first two experiments, and nonsignificantly in the third. The results indicated that the major factor affecting student's learning was the strategy that tutors use of reviewing the material in greater depth on a second pass. Allowing the students to ask questions, and the tutorial strategy for relating new material to the student's previous knowledge contributed only a small amount to the differences found in the first two experiments.

INTRODUCTION

When Carbonell (1970) developed the SCHOLAR CAI system, he created a new type of generative CAI system, capable of carrying on a tutorial dialogue with students. In SCHOLAR knowledge about geography is stored in a semantic network (Quillian, 1968) of facts and concepts, structured like human memory (Collins and Quillian, 1972a). Dialogue with the student is handled by a variety of procedures that utilize the knowledge stored in the semantic network. Different procedures can be used to ask the student questions, evaluate his answers, correct his errors, answer his questions, and present him with new information. Because the information in SCHOLAR is distinct from the procedures used for teaching that information, it is possible to vary the teaching strategy while holding the information constant. This makes it possible to compare different teaching methods in a systematic manner.

Such an evaluation can be made by attempting to teach students the same domain of information with different versions of SCHOLAR, and measuring their learning by comparing pre-test and post-test scores on the material covered. In the three experiments reported here, we compared two different methods of teaching in SCHOLAR, one derived from analysis of tutorial dialogues (called Tutorial Mode) and one derived from programmed learning (called Block-Test Mode).

In order to implement a Tutorial Mode in SCHOLAR, we studied actual human tutoring of South American geography to see how teachers adapt their teaching to the individual student (Collins, Warnock, and Passafiume, 1975). Several important findings from this earlier study related to the ways that tutors select topics, interveave questioning and presentation, and review previous material. Together with the ability to answer student questions, these aspects are basic to the way Tutorial Mode operated in the three experiments.

The topic selection strategy used by the best tutors produced a structure of topics and subtopics like an outline for a course. For example, the tutor might start off with a question like "Do you know any geographical features of South America?" If the student mentioned the Andes, for example, then the tutor would discuss various aspects of the Andes for a while, perhaps including as a subtopic the highest mountain in the Andes, named Aconcagua. After exhausting the important information under the Andes, the tutor would usually ask about other geographical features, like the Amazon or Cape Horn. Each of these would be discussed for a while until the major geographical features were covered, at which point the tutor would pick a new topic such as regions or countries. Thus the topics and subtopics form a nested outline structure, with the tutor probing a little way into each subtopic, and then popping up to the previous level when the important information is exhausted at the lower level.

The way the tutor interweaves questioning and presentation is the essence of how the tutor relates his teaching to the individual student. The dialogues showed that the tutors' questions occur at the top-level and beginning topics in the outline. This is because the tutor starts out asking questions to find out what the student already knows, and then presents new material that is related to the student's previous knowledge. The object seems to be to tie as much information as the student can assimilate into the structure of his previous knowledge (Collins and Quillian, 1972b; Norman, 1973).

The other important aspect of tutoring that we have implemented in SCHOLAR is reviewing. In the dialogues we analyzed, the better tutors went over the material on a second pass, asking about things the student didn't know the first time through, and adding more detail to the structure of information built up on the first pass. The tutorial method as a whole reflected a strategy Norman (1973) refers to as web teaching, where the teacher first tries to establish a framework of basic knowledge and then fills in more and more detail on subsequent passes, much like a spider spinning a web.

In contrast to the tutorial method of teaching, the strategy used in programmed instruction involves presenting small amounts of information, and then asking questions about the information

just presented. We built a version of this strategy into Block-Test mode, which first presents a block of information followed by questions about that block. Block-Test Mode follows the same topic selection strategy as Tutorial Mode, but unlike Tutorial Mode the student can not ask questions.

The programmed instruction strategy in various forms is used widely in CAI. But there is a fundamental problem to this method of instruction: in order for a student to answer most of the questions he is asked, he ends up half parroting back something he read a little earlier in slightly different form. There is at least some evidence in the psychological literature that this leads to little long-term retention. In particular Craik (1971) and McCabe and Madigan (1971) found that in list learning, information recalled from the end of the list is unlikely to be remembered later, because it is recalled out of short term memory without the effort necessary to retain it longer. A similar effect in paired-associate learning underlies the advantage of the anticipation method (where presentation of the first member of the pair precedes presentation of the pair) over the prompting method (where presentation of the pair precedes presentation of the first member). This is because the prompting method involves repeating back what one has just read. These results suggest that the way Tutorial Mode combines questioning and presentation may provide an advantage over Block-Test Mode. In the context of CAI such a tutorial strategy

is only possible in a generative system like SCHOLAR.

In order to explore what aspects of the two teaching strategies affect students' learning, we conducted a series of three experiments. Block-Test Mode was converged toward Tutorial Mode during the course of the experiments in order to pinpoint what differences between the modes had the major effects on students' learning.

METHOD

This section describes the way the two modes operated in the experiments and the testing procedures used. Three experiments were run altogether, each with eight high-school students. Each student learned about two South American countries in Tutorial Mode, and two countries in Block-Test Mode. Thus, students served as their own controls. Students' learning was measured by the difference between pre- and post-test scores on a pencil and paper test.

The best way to understand how the two modes operated is to compare protocols from the two modes with the part of the semantic information network on which the protocols are based. Figure 1 shows part of the entry under Chile that was used in the protocols shown in Figs. 2 and 3. Any topic stored under Chile has an entry elsewhere in the semantic network and can become a topic in its own right, as happened with the Central Valley in Figure 2. The entry for the Central Valley shown in Figure 1 illustrates how each of the things referred to under Chile are expanded elsewhere in the network.

What should be stressed about the data base is the outline structure of information, and the importance tags or I-tags (I0 to I6) associated with each fact. The structure of information in the data base parallels the nested structure of topics and sub-topics selected by the better tutors. The topic selection strategy described was therefore applied directly to the data base structure.

```
CHILE
     SUPERCONCEPT  (I0) COUNTRY
     LOCATION (I0)
          IN (I0)
               SOUTH/AMERICA (I0) SOUTHWESTERN
          BORDERING COUNTRIES (I1)
               NORTHERN (I2) PERU
               NORTHEASTERN (I2) BOLIVIA
               EASTERN (I1) ARGENTINA
                    BOUNDARY (I2) ANDES
          COAST (I1)
               WESTERN (I1) PACIFIC
                    ON (I0) HUMBOLDT/CURRENT
               SOUTHERN (I3) ANTARCTIC
     CAPITAL (I0) SANTIAGO
     REGIONS (I1) $L CENTRAL/VALLEY ATACAMA/DESERT SOUTHERN/ANDES
     PEOPLE (I2)
          LANGUAGE (I0) SPANISH
          RELIGION (I2)
               PRINCIPAL (I0) CATHOLICISM
          POPULATION (I3)
               APPROX (I0) 9,000,000
     CITIES (I3)
          PRINCIPAL (I0) $L SANTIAGO VALPARAISO CONCEPCION
               ANTOFAGASTA ARICA VINA/DEL/MAR PUNTA/ARENAS IQUIQUE

CENTRAL/VALLEY
     SUPERCONCEPT (I0) VALLEY REGION
     LOCATION (I0)
          IN (I0)
               CHILE (I0) CENTRAL
          ON (I0)
               COAST (I0)
                    OF (I0) PACIFIC
     CLIMATE (I2)
          TEMPERATURE (I0) TEMPERATE MEDITERRANEAN
     PRODUCTS (I2)
          AGRICULTURAL (I0) $L WHEAT GRAPES WINE
```

Fig. 1. Partial Entries for Chile and Central Valley in SCHOLAR's Data Base.

The I-tags reflect the varying degrees of importance of different facts in the data base, the most important having the tag I0. These guide the topic selection strategy in picking the most important information at each choice point, and in deciding when to pop up to more important topics. The $L that occurs in the data base indicates that the following elements form a list.

Tutorial Mode

In the experiments, Tutorial Mode selected topics in order of importance. It started out asking the student questions about each topic (e.g., the location of Chile or the people of Chile). As long as the student answered correctly, it went deeper into the topic until time was used up for that topic or there was no important information left (as measured by I-tags). But whenever the student made a mistake, SCHOLAR presented a few related facts and popped up to the next most important topic. In the first two experiments students were allowed to ask questions whenever they liked. In all the experiments, time was allocated between a first pass over the material and a review pass, where questions missed earlier were asked again. If there was time left over after the second pass, SCHOLAR would start on a third pass. The material was covered in more depth on each subsequent pass.

Figure 2 shows an excerpt from an actual protocol on Chile between Tutorial Mode and a high-school student in Experiment 2.

Dialogue	Annotated Comments
WHAT IS CHILE?	This is the most important property of Chile. (A tutor would suppress it on the assumption that the student knows the answer.) In evaluating the answer, it looks for "country", ignores the rest.
COUNTRY IN SOUTH AMERICA	
VERY GOOD.	
CORRECT OR INCORRECT? CHILE IS IN NORTHERN SOUTH AMERICA.	Location is the next most important attribute under Chile.
INCORRECT	The student answers correctly, so SCHOLAR will go deeper into Location.
YOU ARE CORRECT.	
CHILE IS IN SOUTHWESTERN SOUTH AMERICA.	
THE EASTERN BORDERING COUNTRY OF CHILE IS ---.	This is the most important Bordering-Country of Chile.
I DON'T KNOW	
THE ANSWER IS: ARGENTINA.	
THE BOUNDARY OF THE EASTERN BORDERING COUNTRIES OF CHILE IS THE ANDES.	Since the student has missed the question, SCHOLAR presents some related information about Bordering-Countries.
THE NORTHERN BORDERING COUNTRY OF CHILE IS PERU.	
THE NORTHEASTERN BORDERING COUNTRY IS BOLIVIA.	
WHAT IS THE CAPITAL OF CHILE?	SCHOLAR would have gone on to ask about Coasts under Location, but it has run out of time for Location, so it goes to the next attribute, Capital.
WHAT ARE SOME CITIES IN CHILE	
THE PRINCIPAL CITIES OF CHILE ARE SANTIAGO, VALPARAISO, CONCEPCION, ANTOFAGASTA, ARICA, VINA DEL MAR, PUNTA ARENAS, AND IQUIQUE.	The student gets some help, though without directly asking SCHOLAR's question.

Fig. 2a. An Annotated Transcript of the Beginning of a Dialogue Between Tutorial Mode of SCHOLAR and a High-School Student in Exp.2. (The Student's Inputs are Between Asterisks.)

NOW YOU CAN EITHER ANSWER MY QUESTION OR ASK A NEW ONE.

SANTIAGO

VERY GOOD.

Because the student is right, SCHOLAR goes into the data base entry for Santiago to ask more questions. (To go into an entry, the entry must have the main topic as a Superpart, i.e., it must be part of Chile.)

IS THE FOLLOWING TRUE OR FALSE?
SANTIAGO IS IN CENTRAL CHILE.

TRUE

THAT'S FINE.

SCHOLAR has run out of the time allotted to Capital and so goes on to Regions. Time is allocated on the basis of the number of top-level attributes and their I-tag weights.

THE REGIONS OF CHILE ARE ---.

I DON'T KNOW

THE ANSWER IS: THE CENTRAL VALLEY, THE ATACAMA DESERT, AND THE SOUTHERN ANDES.

Since the student does not know any Regions, SCHOLAR gives the correct answer, plus some information about two of the Regions.

THE CENTRAL VALLEY IS IN CENTRAL CHILE. IT IS ON THE COAST OF THE PACIFIC.

THE ATACAMA DESERT IS IN CHILE AND PERU. IT IS BETWEEN THE PACIFIC AND THE ANDES MOUNTAINS.

Language is under the top-level attribute Population. SCHOLAR goes on to Population not because of time but because the student doesn't know any Regions.

WHAT IS THE LANGUAGE OF CHILE?

SPANISH

YOU ARE CORRECT.

That's all SCHOLAR knows about Language, so it goes on to Religion, which is at the same level as Language under Population.
The rest of the information about Religion is above the I-tag level for the first pass, so no related information is printed.

WHAT IS THE PRINCIPAL RELIGION OF CHILE?

I DON'T KNOW

THE ANSWER IS: CATHOLICISM.

Fig. 2b. An Annotated Transcript of a Dialogue Between Tutorial Mode of SCHOLAR and a High-School Student in Exp. 2. (The Student's Inputs are Between Asterisks.)

(Beginning of Review Pass)

THE EASTERN BORDERING COUNTRY OF CHILE IS ---.	On the second pass, SCHOLAR skips to the first piece of information the student missed earlier.
I DON'T KNOW	
THE ANSWER IS: ARGENTINA.	Since the student misses it again, the same related information as before prints outs. SCHOLAR knows that this information was presented earlier, and we may revise SCHOLAR so that it asks about the related information here instead of presenting it again.
THE BOUNDARY OF THE EASTERN BORDERING COUNTRIES OF CHILE IS THE ANDES.	
THE NORTHERN BORDERING COUNTRY OF CHILE IS PERU.	
THE NORTHEASTERN BORDERING COUNTRY IS BOLIVIA.	
WHAT ARE THE REGIONS OF CHILE?	The student missed this the first time. A new answer evaluation module points out that the Andes is right but more specifically it is the Southern Andes.
CENTRAL VALLEY ATACAMA DESSERT AND THE ANDES	
"CENTRAL VALLEY" AND "ATACAMA DESERT" ARE FINE. YOU COULD ALSO MENTION THE SOUTHERN ANDES. I DON'T THINK "DESSERT" IS CORRECT.	SCHOLAR recognizes Atacama as the Atacama Desert and does not understand Dessert.
THE ATACAMA DESERT IS IN CHILE AND PERU. IT IS BETWEEN THE PACIFIC AND THE ANDES MOUNTAINS.	Giving related information about the Atacama Desert was an error that has been corrected.
THE SOUTHERN ANDES ARE IN SOUTHWESTERN SOUTH AMERICA. THEY ARE IN ARGENTINA AND CHILE.	
IS IT TRUE THAT THE CENTRAL VALLEY IS IN CENTRAL CHILE?	Because the student has named the Central Valley, SCHOLAR goes into the entry for Central Valley to ask questions about it. Like human tutors, SCHOLAR will go off and discuss what the student answers correctly. The dialogue continues with a whole series of questions about the Central Valley and the Atacama Desert.
YES	
VERY GOOD.	

Fig. 2c. An Annotated Transcript of the Beginning of the Review Pass Between Tutorial Mode of SCHOLAR and a High-School Student in Exp. 2.

In this experiment students were encouraged to ask questions, and one of these appears in the short excerpt shown. This excerpt also illustrates how the review pass skipped over questions answered earlier, and went into the material in greater depth than the first pass.

Block-Test Mode

Instead of starting with questions like Tutorial Mode, Block-Test Mode presented a block of information and then asked questions about the facts presented. There was a screen over the teletype in the experiments that prevented subjects from looking back at the block during the questions. The version used in the first experiment asked once about all the facts presented in a block. The version used in the second and third experiments randomly selected two of the facts presented to ask about. For each question the student missed, this latter version would ask about an additional fact that was presented. Then it would ask about each fact the student missed once more, to see if he learned the right answer. When the questioning was over for one block, Block-Test Mode presented the next block and then asked questions about it. This cycle continued until the time was used up (except in Experiment 3 where time was divided between a first pass and a review pass, as in Tutorial Mode).

Block-Test Mode selected topics in the same order of importance as Tutorial Mode. Like Tutorial Mode, it went down deeper into a topic until time was used up for that topic or there was no important information left (as measured by I-tags). Unlike Tutorial Mode, it did not pop-up out of a topic when the student made a mistake, so that Block-Test Mode went into most topics in more depth.

Figure 3 shows an actual protocol about Chile between Block-Test Mode and a high-school student in Experiment 3. This excerpt illustrates all the different features that were incorporated into Block-Test Mode in the course of the three experiments. In particular, it illustrates the presentation of a block of facts, the random selection of facts to ask questions about, the inclusion of an additional question for each question the student missed, the repeat of each question the student missed within a block, and the review pass where the facts the student answered correctly on the first pass were skipped over. Block-Test Mode in Experiment 1 did not have most of these features as will be explained later, and in Experiment 2 did not include a review pass.

Dialogue	Annotated Comments
CHILE IS A COUNTRY. IT IS IN SOUTHWESTERN SOUTH AMERICA. THE EASTERN BORDERING COUNTRY IS ARGENTINA. THE BOUNDARY IS THE ANDES.	Four facts from the data base are presented as a block of information.
WHAT IS THE EASTERN BORDERING COUNTRY OF CHILE?	SCHOLAR then asks about one of the four facts chosen randomly.
ARGENTINA	The student answers correctly.
VERY GOOD.	
WHAT IS THE BOUNDARY OF THE EASTERN BORDERING COUNTRIES OF CHILE?	SCHOLAR asks about a second fact, also chosen randomly.
PACIFIC	The student makes a mistake.
WRONG.	
THE CORRECT ANSWER IS: THE ANDES.	
WHAT IS CHILE?	SCHOLAR asks about more than two facts, because the student missed one of the two questions above.
A COUNTRY IN SOUTH AMERICA	The student answers correctly.
THAT'S FINE.	
THE BOUNDARY OF THE EASTERN BORDERING COUNTRIES OF CHILE IS ---.	SCHOLAR reasks the question the student missed earlier.
ARGENTINA	The student missed again.
THAT'S NOT RIGHT.	SCHOLAR only reviews each missed question once.
THE CORRECT ANSWER IS: THE ANDES.	

Fig. 3a. An Annotated Transcript of the Beginning of a Dialogue Between Block-Test Mode of SCHOLAR and a High-School Student in Exp. 3. (The Student's Inputs are Between Asterisks.)

THE NORTHERN BORDERING COUNTRY OF CHILE IS PERU.
THE NORTHEASTEIN BORDERING COUNTRY IS BOLIVIA.
THE WESTERN COAST IS THE PACIFIC.
IT IS ON THE HUMBOLDT CURRENT.

Having reviewed all the questions the student missed the first time, SCHOLAR presents a new block of information.

THE WESTERN COAST OF CHILE IS ---.

PACIFIC

THAT'S RIGHT.

IS THE FOLLOWING TRUE OR FALSE?

THE WESTERN COAST OF CHILE IS NOT ON THE HUMBOLDT CURRENT.

*FALSE:
:
FALSE*

SCHOLAR is confused because the student typed "False" twice. It treats it as a wrong answer, and goes on. Nothing disastrous happens.

"FALSE" IS FINE.
I DON'T THINK "FALSE" IS CORRECT.

THE NORTHERN BORDERING COUNTRY OF CHILE IS ---.

DONT KNOEA

SCHOLAR doesn't recognize "Don't Know" because it is misspelled, and treats it as an error.

WRONG.

THE NORTHERN BORDERING COUNTRY OF CHILE IS PERU.

The dialogue continued through the rest of the first pass.

. . . . . .

Fig. 3b. An Annotated Transcript of the Beginning of a Dialogue Between Block-Test Mode of SCHOLAR and a High-School Student in Exp. 3. (The Student's Inputs are Between Asterisks.)

(Beginning of Review Pass)

CHILE IS IN SOUTHWESTERN SOUTH AMERICA.
THE BOUNDARY OF THE EASTERN BORDERING COUNTRIES IS THE ANDES.
THE NORTHERN BORDERING COUNTRY IS PERU.
THE NORTHEASTERN BORDERING COUNTRY IS BOLIVIA.

On the review pass, SCHOLAR skips the facts the student answered correctly on the first pass. It repeats those facts the student missed or was not asked about.

IS IT TRUE THAT CHILE IS IN SOUTHWESTERN SOUTH AMERICA?

YES

YOU ARE CORRECT.

WHAT IS THE NORTHERN BORDERING COUNTRY OF CHILE?

PERU

THAT'S RIGHT.

THE WESTERN COAST OF CHILE IS ON THE HUMBOLDT CURRENT.
THE SOUTHERN COAST IS THE ANTARCTIC.
THE CAPITAL IS SANTIAGO.
THE REGIONS ARE THE CENTRAL VALLEY, THE ATACAMA DESERT, AND THE SOUTHERN ANDES.

SCHOLAR presents a new block because the student answered both questions correctly. Again it skips over information the student answered correctly on the first pass.

THE SOUTHERN COAST OF CHILE IS ---.

ANTARTICA

YOU MISSPELLED "ANTARCTIC".
THAT'S FINE.

IS IT CORRECT TO SAY THAT THE WESTERN COAST OF CHILE IS ON THE HUMBOLDT CURRENT?

YES

VERY GOOD.

Fig. 3c. An Annotated Transcript of the Beginning of the Review Pass Between Block-Test Mode of SCHOLAR and a High-School Student in Exp. 3.

Testing Procedure

The subjects in the experiments were high-school students paid for their services. They participated in four one-hour sessions with SCHOLAR on two successive Saturdays. In each session they learned about a different South American country. They took a pre-test covering all four countries on the first Saturday before any sessions. Post-tests on the two countries studied each Saturday were given the following Tuesday or Wednesday. Thus the post-test scores measured fairly long-term retention.

On each Saturday the student learned about one country in Tutorial Mode and one in Block-Test Mode. The order of the two modes was reversed for each subject on the two days, and counter-balanced across subjects. The four countries were Brazil, Argentina, Chile, and Venezuela. There was slightly more information stored about the first two countries than about the last two. Thus, Brazil was paired with Argentina, and Chile with Venezuela, so that if a student learned about one in Tutorial Mode, he learned about the other in Block-Test Mode. This counterbalancing by country, day, and order determined the eight different orders used for the eight subjects in each experiment.

The sessions were set to run one hour in overall time before SCHOLAR ended the session. The amount of material covered in an hour varied somewhat depending on the response time of the computer. These variations would not produce any systematic differences between the modes, because the order was the counterbalanced

across subjects. However, in the last two experiments, there may have been less material covered than in the first experiment, because of the increased load on the computer.

The post-tests consisted of 50 questions for each country which fairly exhaustively covered the information the student might have learned in the session. The post-test for each country was divided into two separately administered sections, because questions in the second part were likely to give away answers to questions in the first part. To the extent possible, questions were analogous from country to country. The pre-test consisted of 20 questions (a subset of the 50) from each of the four countries; in fact the easiest of the 50 questions. Based on the pre-test scores, none of the students knew much in advance about any of the countries.

The students were given instructions about the typewriter terminals they used and about the particular mode in which they were about to run. They were told they could not ask questions in Block-Test Mode, and in Tutorial Mode in Experiment 3. With Tutorial Mode in Experiments 1 and 2 they were told the form of some types of questions they could ask, with several examples of different kinds.

Questionnaires were given along with the pre-test and final post-test. The pre-test questionnaire was used to make sure none

of the students had extensive experience with computers or with any particular country. There were also several questions related to whether students like to control their own learning or not, to see if these answers had any predictive value with respect to how much students would learn in the different modes. The questionnaire given with the final post-test interrogated students about the effectiveness of different aspects of the two modes.

EXPERIMENT 1

In the first experiment we attempted to see if clearly different forms of the two strategies would produce a significant difference in the amount of knowledge students acquire. Thus the version of Block-Test Mode used in the first experiment was quite primitive. It proceeded along the same path whether the student answered correctly or incorrectly. This experiment then compared a highly adaptive method of teaching with a non-adaptive method.

Method

In Tutorial Mode, students were introduced to a small number of question types they could ask: the types were "Where is X?" "What is X?" and "Tell me about X." where X could be anything like the Central Valley or temperate. This type of question is useful when SCHOLAR uses a term the student doesn't know about. Students were not strongly encouraged to ask questions, and in fact did not ask many.

The version of Block-Test Mode used in Experiment 1 presented eight pieces of information rather than the four pieces shown in Fig. 3. (Less than eight were presented if it were in a subtopic like Central Valley, where there were less than eight available for selection.) It then went on to ask about each of these eight facts in the order presented. It corrected the student when he

made a mistake, but it did not review any questions the student missed. When the student had been questioned about all eight facts, Block-Test Mode presented the next block of eight facts, and so on until the hour was used up.

Results and Discussion

The average difference scores between pre- and post-tests for each subject are shown in Table 1, broken down by presentation mode. To analyze the results of the experiment, we used a three-way analysis of variance based on the difference scores with mode, order of the modes, and subjects as the three factors. Since there was only one observation per cell, we took the mean square of the triple interaction as the estimate of error variance. Of the main factors, the effect of mode was significant ($F(1,7)=17.53$, $p\leq.01$), the effect of subjects was significant ($F(7,7)=14.45$, $p\leq.01$), and the effect of order was not significant ($F(1,7)=.38$). Of the two-way interactions, the interaction between mode and order was significant ($F(1,7)=10.58$, $p<.05$), and the other two interactions were not: for subjects and mode, $F(7,7)=.73$, and for subjects and order, $F(7,7)=2.71$. The significant interaction between mode and order reflects the fact that subjects tended to remember the second country they learned about on each day better than the first country. But when Block-Test Mode was second, this effect was more than offset by the mode effect, and so the difference in retention shows up as an interaction.

TABLE 1

Difference Scores between Pre-Test and Post-Test for each Subject, broken down by Presentation Mode in the Three Experiments.

	Experiment 1		Experiment 2		Experiment 3	
	Block-Test Mode	Tutorial Mode	Block-Test Mode	Tutorial Mode	Block-Test Mode	Tutorial Mode
Subjects						
S1	7.0	11.5	14.5	19.5	16.0	17.5
S2	2.0	3.5	12.5	17.0	18.0	13.0
S3	12.0	17.0	10.0	6.0	13.5	16.0
S4	8.0	10.5	1.5	4.5	15.5	11.5
S5	16.5	17.0	3.0	8.0	1.0	7.5
S6	11.0	17.5	7.5	8.5	5.0	5.0
S7	8.5	12.0	10.5	22.0	14.0	17.5
S8	8.5	11.5	7.5	12.5	9.5	7.0
Average	9.2	12.5	8.4	12.4	11.6	11.9

The two effects we were interested in were the effect of mode, where Tutorial Mode was clearly superior to Block-Test Mode, and the lack of any interaction between mode and subjects. Taken together these two results indicate that the superiority of Tutorial Mode was common to all the students and not just to those who prefer to control their own learning. Hence it is clear that some aspects of Tutorial Mode were of general benefit to student's learning of factual knowledge.

In the first experiment, students did not ask many questions in Tutorial Mode. On a questionnaire given with the final post-test, the students commented favorably about Tutorial Mode and particularly the procedure of going over material more than once. In contrast, they said Block-Test Mode gave them too much information at once. They also felt it was very helpful to get information related to the question they missed in Tutorial Mode. Based on these comments, and the infrequency of questions by the students, the superiority of Tutorial Mode probably was due to its reviewing, to the excess of information presented at one time in Block-Test Mode, and to the difference in the way the two modes combine questioning and presentation.

EXPERIMENT 2

In the second experiment, we decided to see how a more adaptive version of Block-Test Mode would compare with Tutorial Mode. In view of the student's criticisms of Block-Test Mode, we cut the amount of information in a block to four facts, and included reviewing of questions within a block. At the same time, we strongly encouraged students to ask questions in Tutorial Mode in order to see if this would lead to greater learning in Tutorial Mode.

Method

Eight new students from the same high school were used as subjects. Tutorial Mode was improved in several small ways, but functioned essentially as in the first experiment. However, subjects were encouraged to ask questions and told about a wider variety of questions they could ask, with many more examples given. The main additional types explained to the students were "What?" questions like "What is the climate of the Central Valley?" and true/false questions like "Is Lima in Chile?"

Block-Test Mode was changed drastically in the second experiment. The number of facts presented in each block was cut to four, except

when fewer than four facts were stored with a particular subtopic. Instead of asking about all the facts presented, this version of Block-Test Mode randomly selected two of the facts presented to ask about. It asked an additional question for each one the student missed, up to the number presented. Then it would review any the student missed or did not answer the first time. After this reviewing within the block, it went on to the next block. There was no review pass through the material as there was in Tutorial Mode, nor were the students allowed to ask questions.

Results and Discussion

The results for Experiment 2 are also shown in Table 1, and again the students learned much more in Tutorial Mode. Of the main effects in the analysis of variance, the effect of mode was significant ($F(1,7)=6.18, p<.05$), the effect of subjects was significant ($F(7,7)=5.16, p<.05$), and the effect of order was not significant ($F(1,7)=3.48$). Of the two-way interactions, none of the effects were significant: for subjects and order $F(7,7)=.94$, and for subjects and order $F(7,7)=3.75$, and for order and mode $F(1,7)=.02$. There was more variability in the data for this experiment, which may be because of variability in the load on the computer.

In this experiment we eliminated the two aspects of Block-Test Mode that the students complained about; i.e., the excess of information in a block and the failure to review missed questions. However, Tutorial Mode still showed the same large advantage over Block-Test Mode. There were three differences between the two modes that might contribute to this difference in effectiveness: 1) the way they combine questioning and presentation, 2) the allocation of time between a first pass and a review pass in Tutorial Mode and 3) the option for students to ask questions in Tutorial Mode. Only the first of these is inherently different between the modes.

The students were encouraged to ask more questions with Tutorial Mode in this experiment, and in fact they did. Some asked quite a few questions and it looked as if question-asking was a detriment for some and a benefit for others. When a student went off asking questions on his own, the most important information often was not covered. But when the student used questions for clarification, the ability to ask questions was probably quite helpful. This double-edged nature of questions may also have contributed to the increase in variability of the data.

We had expected an overall increase in scores between the two experiments because of the improvements to Block-Test Mode

and the increase in student questions in Tutorial Mode. The fact that there was no improvement in Tutorial Mode may have been due to the double-edged nature of questions. In Block-Test Mode for which there was also no improvement, it may have been due to the failure to ask about every fact presented. Alternatively, it could have been due to the particular groups of subjects, or the fact that the load was slightly heavier on the computer in the second experiment, which would lead to a decrease in material covered. Any comparisons across experiments, though interesting, are somewhat tenuous.

EXPERIMENT 3

There were three factors then that could have contributed to the large difference between the two modes: 1) the way the modes combine questioning and presentation, 2) the review pass, and 3) the option to ask questions. In order to pin down the contribution of each of these factors, we ran a third experiment converging the two modes toward each other. The first factor was inherently different between the modes so this difference was unchanged. But by adding a review pass to Block-Test Mode, and by removing the option to ask questions in Tutorial Mode, we could parcel out the effect of these different factors.

Method

Eight new subjects were used from the same high school. Tutorial Mode was the same as in the previous experiment, except that the subjects were not told that they could ask questions.

Block-Test Mode was also the same except that it was revised to allocate its time between a first pass and a review pass (see Figure 3c). Thus it did not go as deeply into each topic as the previous version. In the review pass, a block of four facts was followed by a sample of two questions just as in the first pass. But the information the student answered correctly on the first

pass was skipped over on the review pass. A fact was not skipped over, however, if it was presented on the first pass but not asked about, or if the student missed the question about it the first time but answered correctly on the review question within the block. The review pass went over the material to the same depth as the first pass, but any subsequent passes (if time was left) went somewhat deeper.

Results and Discussion

The results for Experiment 3 are shown in Table 1 where they can be compared to the results of the first two experiments. Of the main effects in the analysis of variance, the effect of subjects was significant ($F(7,7)=7.67$, $p < .01$), but neither the effect of mode ($F(1,7)=.058$) nor the effect of order ($F(1,7)=.669$) were significant. None of the interactions were significant: for subjects and mode $F(7,7)=1.16$, for subjects and order $F(7,7)=1.47$, and for mode and order $F(1,7)=.058$.

It is clear that allocating time between a first pass and a review pass eliminated most of the difference between the modes, and so this must have been the major factor contributing to the difference between the modes. The residual difference between the two modes is a measure of the effect of the way the two modes combine questioning and presentation. There seems to be a slight

benefit to the tutorial strategy of questioning first to learn what the student knows, but it is not significant with such a small number of subjects.

The decrease in scores for Tutorial Mode from Experiments 1 and 2 to Experiment 3 is a measure of the effectiveness of the option to ask questions. Though comparisons across experiments are tenuous, the stability of the scores with Tutorial Mode in Experiments 1 and 2 increases confidence in the comparison with Experiment 3. The decrease in Experiment 3 is a small effect, but suggests that the option to ask questions is somewhat beneficial.

DISCUSSION

The effectiveness of the review pass for the student's learning in the experiments emphasizes the importance of Norman's (1973) notion of web teaching. By including two or more passes in Block-Test Mode, it began to approximate the web-teaching strategy of building a basic framework on one pass, to which new material is added on later passes. The second pass repeats what the student was least certain about on the first pass, so that the repetition is concentrated on the most difficult things to remember. It is a strategy particularly suited to CAI, because computers have perfect memory for what the student answered earlier.

In the experiments there was little benefit from using the tutorial strategy of finding out what the student knows first, and then adding related information. This may reflect the fact that all the students knew practically nothing about South American geography to start with. Therefore, there was little to be gained from an ability to skip over information the student already knew. The real effectiveness of the tutorial strategy is its ability to adapt to the level of the individual student (though Tutorial Mode only partly reflects the human tutor's ability), and the students in these experiments did not have enough knowledge to take advantage of this capability.

The failure of the within block reviewing using Block-Test Mode in Experiment 2 to benefit student's learning probably reflects the spacing effect often found in experimental psychology (Melton, 1970). By spacing repetitions far apart in time as in Experiment 3, the material is more likely to be doubly encoded, whereas repetitions spaced close together as in Experiment 2 are more likely to be encoded only once. Thus the material widely spaced is better remembered.

The method used here can be extended to study many different aspects of teaching methods. It can be used to investigate teaching of different kinds of knowledge other than the factual knowledge in SCHOLAR's geography data base. There are now functioning SCHOLAR systems for teaching users how to use a computer text-editing system (Grignetti, et al., 1974) and for teaching geography with maps (Collins & Warnock, 1974). Another system is planned for teaching functional relations in geography (Collins, et al, 1975). For example the method can be used with the map facility by comparing how well students learn the same material with and without maps. The effectiveness of different methods may vary with the different kinds of knowledge being taught or with different students. Beacuse of its inherent flexibility, the teaching method SCHOLAR uses can be geared to the particular kind of knowledge or even left to the student's choice. This is one of the great potential benefits of generative CAI.

The fact that SCHOLAR can be used to test particular aspects of teaching methods makes it potentially a valuable tool for educational research. The possibility of trying out single modifications in teaching strategy to see their effects on students' learning rate is unique to SCHOLAR. Human teachers of course can make such modifications in their own teaching strategies, but there is no way to control all the other factors that might vary as they changed strategy. However, any specific version of SCHOLAR is a fixed system, and so an unbiased comparison can be made using any number of subjects. In this way the accumulation of systematic knowledge about teaching methods can begin to occur.

REFERENCES

Carbonell, J. R. AI in CAI: An artificial intelligence approach to computer-aided instruction. IEEE Transactions on Man-Machine Systems, 1970, MMS-11, 190-202.

Collins, A. M., and Quillian, M. R. How to make a language user. In E. Tulving and W. Donaldson (Eds.), Organization of Memory. New York: Academic Press, 1972a.

Collins, A. M. and Quillian, M. R. Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.) Cognition in learning and memory. New York: Wiley, 1972b.

Collins, A. M. and Warnock, E. H. Semantic Networks. Cambridge, Mass.: Bolt Beranek and Newman, Report No. 2833, 1974.

Collins, A. M., Warnock, E. H., and Passafuime, J. J. Analysis and synthesis of tutorial dialogues. To appear in G. H. Bower (Ed.) The psychology of learning and motivation, Vol. 9 New York: Academic Press, 1975.

Collins, A. M., Warnock, E. H., Aiello, N., and Miller, M. W. Reasoning from incomplete knowledge. To appear in in D. G. Bobrow and A. M. Collins (Eds.) Representation and understanding. New York: Academic Press, 1975.

Craik, F. I. M. The fate of primary memory items in free recall. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 143-148.

Grignetti, M. C., Gould, L., Bell, A., Haumann, C., Passafiume, J. J., and Harris, G. A mixed-initiative tutorial system to aid users of the on-line system (NLS). Cambridge, Mass., Bolt Beranek and Newman Inc. Report, 1974.

Madigan, S. A., and McCabe, L. Perfect recall and total forgetting: A problem for models of short-term memory. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 101-106.

Melton, A. W. The situation with respect to the spacing of repetitions and memory. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 596-606.

Norman, D. A. Memory, knowledge, and the answering of questions. In R. L. Solso (Ed.), Contemporary issues in cognitive psychology: The Loyola symposium. New York: Halsted Press, 1973.

Quillian, M. R. Semantic memory. In M. Minsky (Ed.), Semantic information processing. Cambridge, Massachusetts: M.I.T. Press, 1968.