COMPUTERIZED INSTRUCTION AND THE LEARNING PROCESS [1]

RICHARD C. ATKINSON

Stanford University

IN recent years there has been a tremendous number of articles and news releases dealing with computer-assisted instruction, or as it has been abbreviated, CAI. One might conjecture that this proliferation is an indicant of rapid progress in the field. Unfortunately, I doubt that it is. A few of the reports about CAI are based on substantial experience and research, but the majority are vague speculations and conjectures with little if any data or real experience to back them up. I do not want to denigrate the role of speculation and conjecture in a newly developing area like CAI. However, of late it seems to have produced little more than a repetition of ideas that were exciting in the 1950s but, in the absence of new research, are simply well-worn cliches in the late 1960s.

These remarks should not be misinterpreted. Important and significant research on CAI is being carried on in many laboratories around the country, but certainly not as much as one is led to believe by the attendant publicity. The problem for someone trying to evaluate developments in the field is to distinguish between those reports that are based on fact and those that are disguised forms of science fiction. In my paper, I shall try to stay very close to data and actual experience. My claims will be less grand than many that have been made for CAI, but they will be based on a substantial research effort.

In 1964 Patrick Suppes and I initiated a project under a grant from the Office of Education to develop and implement a CAI program in initial reading and mathematics. Because of our particular research interests, Suppes has taken responsibility for the mathematics curriculum and I have been responsible for the initial reading program. At the beginning of the project, two major hurdles had to be overcome. There was no lesson material in either mathematics or reading suitable for CAI, and an integrated CAI system had not yet been designed and produced by a single manufacturer. The development of the curricula and the development of the system have been carried out as a parallel effort over the last 3 years with each having a decided influence on the other.

Today I would like to report on the progress of the reading program with particular reference to the past school year when for the first time a sizable group of students received a major portion of their daily reading instruction under computer control. The first year's operation must be considered essentially as an extended debugging of both the computer system and the curriculum materials. Nevertheless, some interesting comments can be made on the basis of this experience regarding both the feasibility of CAI and the impact of such instruction on the overall learning process.

Before describing the Stanford Project, a few general remarks may help place it in perspective. Three levels of CAI can be defined. Discrimination between levels is based not on hardware considerations, but principally on the complexity and sophistication of the student-system interaction. An advanced student-system interaction may be achieved with a simple teletype terminal, and the most primitive interaction may require some highly sophisticated computer programming and elaborate student terminal devices.

At the simplest interactional level are those systems that present a fixed, linear sequence of problems. Student errors may be corrected in a variety of ways, but no real-time decisions are made for modifying the flow of instructional material as a function of the student's response history. Such systems have been termed "drill-and-practice" systems and at Stanford University are exemplified by a series of fourth-, fifth-, and sixth-grade programs in arithmetic and language arts that are designed to supplement classroom instruction. These particular programs are being used in several different areas of California and also in Kentucky and Mississippi, all under control of one central

[1] Invited address presented at the meeting of the Division of Educational Psychology, American Psychological Association, Washington, D. C., September 1967.

computer located at Stanford University. Currently as many as 2,000 students are being run per day; it requires little imagination to see how such a system could be extended to cover the entire country. Unfortunately, I do not have time to discuss these drill-and-practice programs in this paper, but there are several recent reports describing the research (Fishman, Keller, & Atkinson, 1968; Suppes, 1966; Suppes, Jerman, & Groen, 1966).

At the other extreme of our scale characterizing student-system interactions are "dialogue" programs. Such programs are under investigation at several universities and industrial concerns, but to date progress has been extremely limited. The goal of the dialogue approach is to provide the richest possible student-system interaction where the student is free to construct natural-language responses, ask questions in an unrestricted mode, and in general exercise almost complete control over the sequence of learning events.

"Tutorial" programs lie between the above extremes of student-system interaction. Tutorial programs have the capability for real-time decision making and instructional branching contingent on a single response or on some subset of the student's response history. Such programs allow students to follow separate and diverse paths through the curriculum based on their particular performance records. The probability is high in a tutorial program that no two students will encounter exactly the same sequence of lesson materials. However, student responses are greatly restricted since they must be chosen from a prescribed set of responses, or constructed in such a manner that a relatively simple text analysis will be sufficient for their evaluation. The CAI Reading Program is tutorial in nature, and it is this level of student-interaction that will be discussed today.

THE STANFORD CAI SYSTEM

The Stanford Tutorial System was developed under a contract between the University and the IBM Corporation. Subsequent developments by IBM of the basic system have led to what has been designated the IBM-1500 Instructional System which should soon be commercially available. The basic system consists of a central process computer with accompanying disc-storage units, proctor stations, and an interphase to 16 student terminals. The central process computer acts as an intermediary between each student and his particular course material which is stored in one of the disc-storage units. A student terminal consists of a picture projector, a cathode ray tube (CRT), a light pen, a modified typewriter keyboard, and an audio system which can play prerecorded messages (see Figure 1).

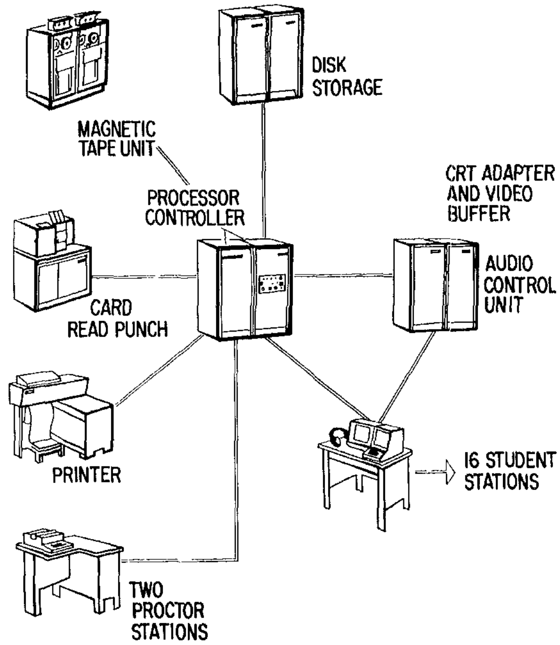

FIG. 1. System configuration for Stanford CAI System.

The CRT is essentially a television screen on which alpha-numeric characters and a limited set of graphics (i.e., simple line drawings) can be generated under computer control. The film projector is a rear-view projection device which permits us to display still pictures in black and white or color. Each film strip is stored in a self-threading cartridge and contains over 1,000 images which may be accessed very quickly under computer control. The student receives audio messages via a high-speed device capable of selecting any number of messages varying in length from a few seconds to over 15 minutes. The audio messages are stored in tape cartridges which contain approximately 2 hours of messages and, like the film cartridge, may be changed very quickly. To gain the student's attention, an arrow can be placed at any point on the CRT and moved in synchronization with an audio message to emphasize given words or phrases, much like the "bouncing ball" in a singing cartoon.

The major response device used in the reading program is the light pen, which is simply a light-sensitive probe. When the light pen is placed on the CRT, coordinates of the position touched are sensed as a response and recorded by the computer. Responses may also be entered into the system through the typewriter keyboard. However, only limited use has been made of this response mode in the reading program. This is not to minimize the value of keyboard responses, but rather to admit that we have not as yet addressed ourselves to the problem of teaching first-grade children to handle a typewriter keyboard.

The CAI System controls the flow of information and the input of student responses according to the instructional logic built into the curriculum materials. The sequence of events is roughly as follows: The computer assembles the necessary commands for a given instructional sequence from a disc-storage unit. The commands involve directions to the terminal device to display a given sequence of symbols on the CRT, to present a particular image on the film projector, and to play a specific audio message. After the appropriate visual and auditory materials have been presented, a "ready" signal indicates to the student that a response is expected. Once a response has been entered, it is evaluated and, on the basis of this evaluation and the student's past history, the computer makes a decision as to what materials will subsequently be presented. The time-sharing nature of the system allows us to handle 16 students simultaneously and to cycle through these evaluative steps so rapidly that from a student's viewpoint it appears that he is getting immediate attention from the computer whenever he inputs a response.

THE CAI READING CURRICULUM

The flexibility offered by this computer system is of value only if the curriculum materials make sense both in terms of the logical organization of the subject matter and the psychology of the learning processes involved. Time does not permit a detailed discussion of the rationale behind the curriculum that we have developed. Let me simply say that our approach to initial reading can be characterized as applied psycholinguistics. Hypotheses about the reading process and the nature of learning to read have been formulated on the basis of linguistic information, observations of language use, and an analysis of the function of the written code. These hypotheses have been tested in a series of pilot studies structured to simulate actual teaching situations. On the basis of these experimental findings, the hypotheses have been modified, retested, and ultimately incorporated into the curriculum as principles dictating the format and flow of the instructional sequence. Of course, this statement is somewhat of an idealization, since very little curriculum material can be said to have been the perfect end product of rigorous empirical evaluation. We would claim, however, that the fundamental tenets of the Stanford reading program have been formulated and modified on the basis of considerable empirical evidence. It seems probable that these will be further modified as more data accumulate.

The introduction of new words from one level of the curriculum to the next is dictated by a number of principles (Rodgers, 1967). These principles are specified in terms of a basic unit that we have called the vocalic center group (VCG). The VCG in English is defined as a vowel nucleus with zero to three preceding and zero to four following consonants. The sequencing of new vocabulary is determined by the length of the VCG units, and the regularity of the orthographic and phonological correspondences. Typical of the principles are the following:

1. VCG sets containing single consonant elements are introduced before those containing consonant clusters (*tap* and *rap* before *trap*).
2. VCG sets containing initial consonant clusters are introduced before those containing final consonant clusters (*stop* before *post*).
3. VCG sets containing check (short) vowels are introduced before those containing letter name (long) vowels (*met* and *mat* before *meat* or mate).
4. Single VCG sequences are introduced before multiple VCG sequences (*mat* before *matter, stut* before *stutter*).

More detailed rules are required to determine the order for introducing specific vowels and consonants within a VCG pattern, and for introducing specific VCG patterns in polysyllabic words. These rules frequently represented a compromise between linguistic factors, pattern productivity, item frequency, and textual "usefulness," in that order of significance.

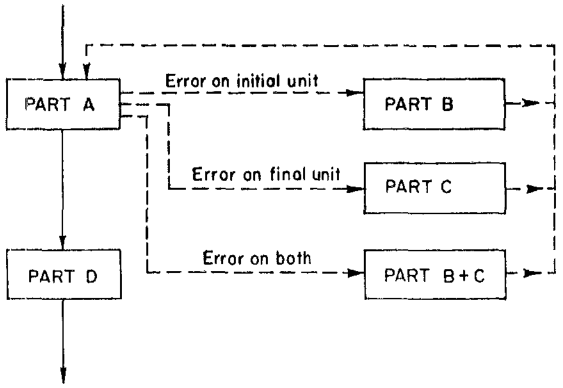

FIG. 2. Flow chart for the construction of a cell in the matrix construction task.

The instructional materials are divided into eight levels each composed of about 32 lessons.[2] The lessons are designed so that the average student will complete one in approximately 30 minutes, but this can vary greatly with the fast student finishing much sooner and the slow student sometimes taking 2 hours or more if he hits most of the remedial material. Within a lesson, the various instructional tasks can be divided into three broad areas: (*a*) decoding skills, (*b*) comprehension skills, (*c*) games and other motivational devices. Decoding skills involve such tasks as letter and letter-string identification, word list learning, phonic drills, and related types of activities. Comprehension involves such tasks as having the computer read to the child or having the child himself read sentences, paragraphs, or complete stories about which he is then asked a series of questions. The questions deal with the direct recall of facts, generalizations about main ideas in the story, and inferential questions which require the child to relate information presented in the story to his own experience. Finally, many different types of games are sequenced into the lessons primarily to encourage continued attention to the materials. The games are similar to those played in the classroom and are structured to evaluate the developing reading skills of the child.

Matrix construction. To illustrate the instructional materials focusing on decoding skills let me describe a task that we have called matrix "construction." This task provides practice in learning to associate orthographically similar sequences with appropriate rhyme and alliteration patterns. Rhyming patterns are presented in the columns of the matrix, and alliteration patterns are presented in the rows of the matrix as indicated in Figure 4.

The matrix is constructed one cell at a time. The initial consonant of a CVC word is termed the *initial unit*, and the vowel and the final consonant are termed the final unit. The intersection of an initial unit row and a final unit column determines the entry in any cell.

The problem format for the construction of each cell is divided into four parts: Parts A and D are standard instructional sections and Parts B and C are remedial sections. The flow diagram in Figure 2 indicates that remedial Parts B and C are branches from Part A and may be presented independently or in combination.

To see how this goes, let us consider the example illustrated in Figure 3. The student first sees on the CRT the empty cell with its associated initial and final units and an array of response choices. He hears the audio message indicated by response request 1 (RR 1) in Part A of Figure 3. If the student makes the correct response (CA) (i.e., touches *ran* with his light pen), he proceeds to Part D where he sees the word written in the cell and receives one additional practice trial.

In the initial presentation in Part A, the array of multiple-choice responses is designed to identify three possible types of errors:

1. The initial unit is correct, but the final unit is not.
2. The final unit is correct, but the initial unit is not.
3. Neither the initial unit nor the final unit is correctly identified.

If, in Part A, the student responds with *fan* he is branched to remedial Part B where attention is focused on the initial unit of the cell. If a correct response is made in Part B, the student is returned to Part A for a second attempt. If an incorrect response (WA) is made in Part B, an arrow is displayed on the CRT to indicate the correct response, which the student is then asked to touch.

If, in Part A, the student responds with *rat,* he is branched to remedial Part C where additional instruction is given on the final unit of the cell.

[2] For a detailed account of the curriculum materials see Wilson and Atkinson (1967) and Rodgers (1967). See also Atkinson and Hansen (1966) and Hansen and Rodgers (1965).

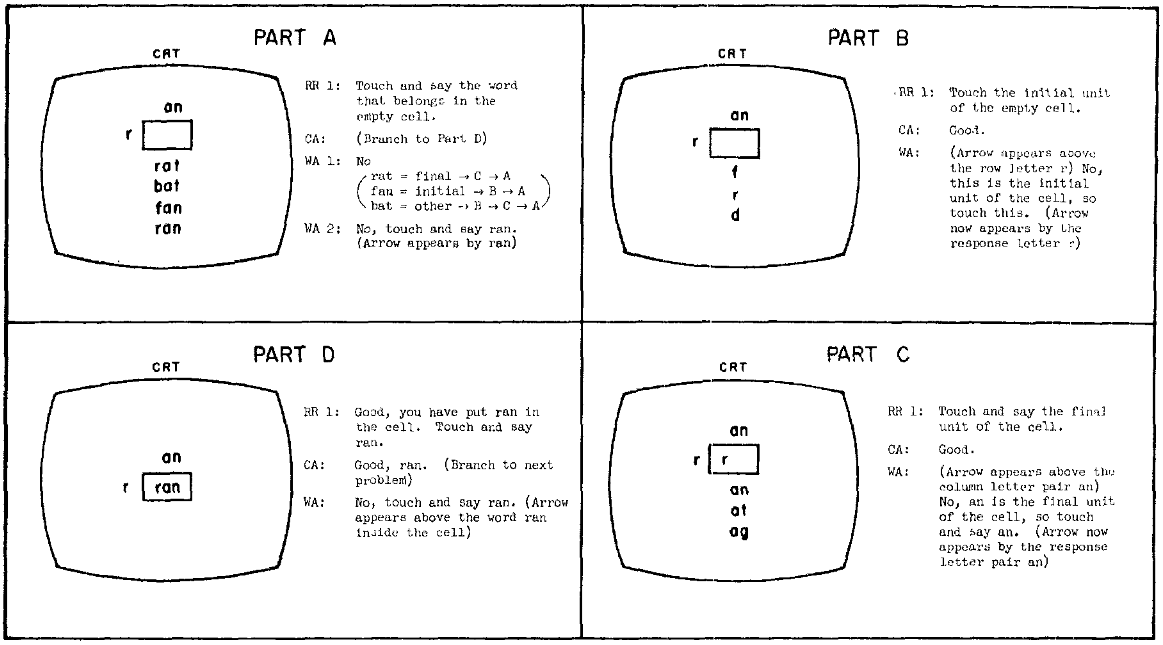

FIG. 3. First cell of the matrix construction task.

The procedure in Part C is similar to Part B. However, it should be noted that in the remedial instruction the initial letter is never pronounced (Part B), whereas the final unit is always pronounced (Part C). If, in Part A, the student responds with *bat,* then he has made an error on both the initial and final unit and is branched through both Part B and Part C.

When the student returns to Part A after completing a remedial section, a correct response will advance him to Part D as indicated. If a wrong answer response is made on the second pass, an arrow is placed beside the correct response area and held there until a correct response is made. If the next response is still an error, a message is sent to the proctor and the sequence is repeated from the beginning.

When a student has made a correct response on Parts A and D, he is advanced to the next word cell of the matrix which has a problem format and sequence identical to that just described. The individual cell building is continued block by block until the matrix is complete. The upper left-hand panel of Figure 4 indicates the CRT display for adding the next cell in our example. The order in which row and column cells are added is essentially random.

When the matrix is complete, the entries are reordered and a criterion test is given over all cell entries. The test involves displaying the full matrix with complete cell entries as indicated in the lower left-hand panel of Figure 4. Randomized requests are made to the student to identify cell entries. Since the first pass through the full matrix is viewed as a criterion test, no reinforcement is given. Errors are categorized as initial, final, and other; if the percentage of total errors on the criterion test exceeds a predetermined value, then remedial exercises are provided of the type shown in the two right-hand panels of Figure 4. If all the errors are recorded in one category (initial or final), only the remedial material appropriate to that category is presented. If the errors are distributed over both categories, then both types of remedial material are presented. After working through one or both of the remedial sections, the student is branched back for a second pass through the criterion matrix. The second pass is a teaching trial as opposed to the initial test cycle; the student proceeds with the standard correction and optimization routines.

An analysis of performance on the matrix task is still incomplete, but some preliminary results are available. On the initial pass (Part A) our students were correct about 45% of the time; however, when an error did occur, 21% of the time it in-

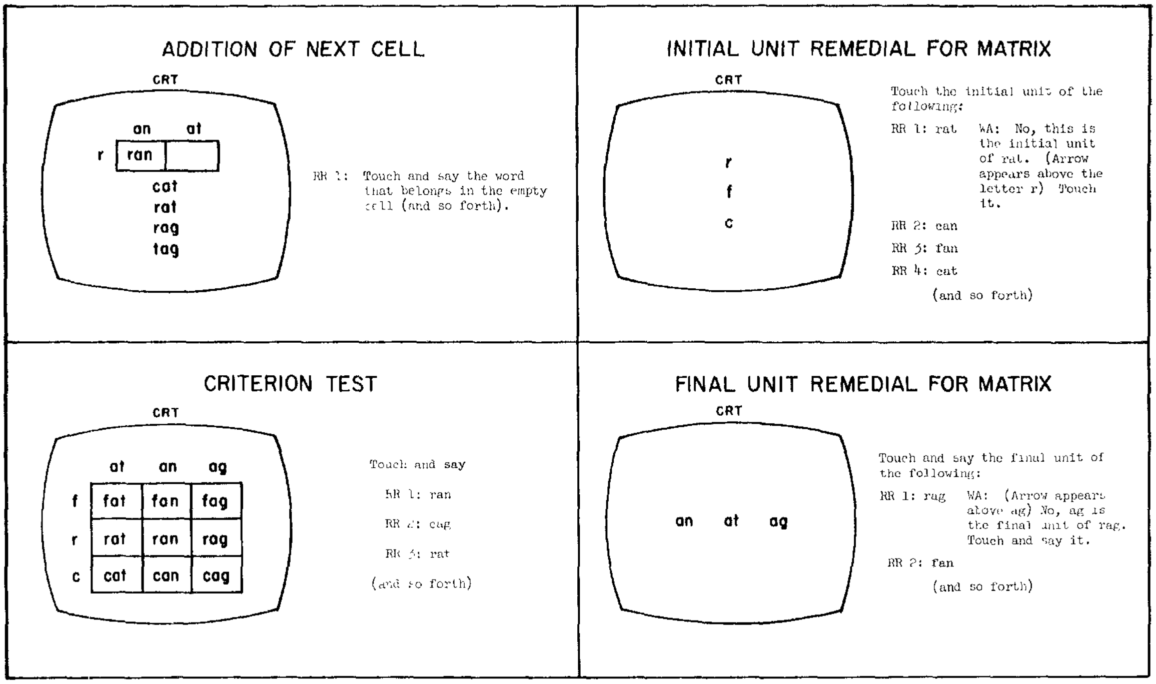

FIG. 4. Continuation of matrix construction task.

volved only the final unit, 53% of the time only the initial unit, and 26% of the time both initial and final units. The pattern of performances changed markedly on the first pass through the criterion test. Here the subject was correct about 65% of the time; when an error occurred, 32% of the time it involved only the final unit, 33% of the time only the initial unit, and 35% of the time both units. Thus performance showed a significant improvement from Part A to the criterion test; equally important, initial errors were more than twice as frequent as final errors in Part A, but were virtually equal on the criterion test.

The matrix exercise is a good example of the material used in the curriculum to teaching decoding skills. We now consider two examples ("form class" and "inquiries") of tasks that are designed to teach comprehension skills.

Form class. Comprehension of a sentence involves an understanding of English syntax. One behavioral manifestation of a child's syntactic sophistication is his ability to group words into appropriate form classes. This task provides lesson materials that teach the form-class characteristics of the words just presented in the matrix section of a lesson. The following type of problem is presented to the student (the material in the box is displayed on the CRT and below are audio messages; the child answers by appropriately placing his light pen on the CRT):

	tan	
	fat	
Dan saw the	man	hat.
	run	

Only one of the words in the column will make sense in the sentence. Touch and say the word that belongs in the sentence.

CA: Yes, Dan saw the tan hat. Do the next one.
WA: No, tan is the word that makes sense. Dan saw the tan hat. Touch and say tan. (An arrow then appears above tan.)

The sentence is composed of words that are in the reading vocabulary of the student (i.e., they have been presented in previous or current lessons). The response set includes a word which is of the correct form class but is semantically inappropriate, two words that are of the wrong form class, and the correct word. A controlled variety of sentence types is employed, and the answer sets are distributed over all syntactic slots within each sentence type. Responses are categorized in rather broad terms as *nouns, verbs, modifiers,* and *other.*

The response data can be examined for systematic errors over a large number of items. Examples of the kinds of questions that can be asked are: (*a*) Are errors for various form classes in various sentence positions similarly distributed? (*b*) How are response latencies affected by the syntactic and serial position of the response set within the sentence? Answers to these and other questions should provide information that will permit more systematic study of the relationship of sentence structure to reading instruction.

Inquiries. Individual words in sentences may constitute unique and conversationally correct answers to questions. These questions take the interrogative "Who ....? What ....? How ....?" etc. The ability to select the word in a sentence that uniquely answers one of these questions demonstrates one form of reading comprehension. The inquiry exercises constitute an assessment of this reading comprehension ability. In the following example, the sentence "John hit the ball" is displayed on the CRT accompanied by these audio messages:

Touch and say the word that answers the question.

RR 1 Who hit the ball?
CA: Yes, the word "John" tells us who hit the ball.
WA: No, John tells us who hit the ball. Touch and say John. (An arrow then appears on the CRT above John.)
RR 2 What did John hit?
CA: Yes, the word "ball" tells us what John hit.
WA: No, ball tells us what John hit. Touch and say ball. (An arrow then appears above ball.)

As in the form-class section, each sentence is composed of words from the student's reading vocabulary. A wide variety of sentence structures is utilized, beginning with simple subject-verb-object sentences and progressing to structures of increasing complexity. Data from this task bear on several hypotheses about comprehension. If comprehension is equated with a correct response to an inquiry question, then the following statements are verified by our data: (*a*) Items for which the correct answer is in the medial position of the sentence are more difficult to comprehend than items in the initial or final positions; final position items are easier to comprehend than items in the initial position. (*b*) Items for which the correct answer is an adjective are more difficult to comprehend than items in which the correct answer is a noun or verb; similarly nouns are more difficult than verbs. (*c*) Longer sentences, measured by word length, are more difficult to comprehend than shorter sentences.

These are only a few examples of the types of tasks used in the reading curriculum, but they indicate the nature of the student-system interaction. What is not illustrated by these examples is the potential for long-term optimization policies based on an extended response history from the subject. We shall return to this topic later.

PROBLEMS IN IMPLEMENTING THE CURRICULUM

Before turning to the data from last year's run, let me consider briefly the problem of translating the curriculum materials into a language that can be understood by the computer. The particular computer language we use is called Coursewriter II, a language which was developed by IBM in close collaboration with Stanford. A coded lesson is a series of Coursewriter II commands which causes the computer to display and manipulate text on the CRT, position and display film in the projector, position and play audio messages, accept and evaluate keyboard and lightpen responses, update the performance record of each student, and implement the branching logic of the lesson flow by means of manipulating and referencing a set of switches and counters. A typical lesson in the reading program, which takes the average student

TABLE 1

AUDIO SCRIP AND FILM CHIPS WITH HYPOTHETICAL ADDRESSES

Address	Message
	Audio information
A01	Touch and say the word that goes with the picture.
A02	Good. Bag. Do the next one.
A03	No.
A04	The word that goes with the picture is bag. Touch and say bag.
A05	Good. Card. Do the next one.
A06	No.
A07	The word that goes with the picture is card. Touch and say card.
	Film strip
F01	Picture of a bag.
F02	Picture of a card.

TABLE 2

COMPUTER COMMANDS REQUIRED TO PRESENT TWO EXAMPLES OF THE PROBLEM DESCRIBED IN THE TEXT

Commands	Explanation
PR	Problem: Prepares machine for beginning of new problem.
LD 0/S1	Load: Loads zero into the error switch (S1). The role of switches and counters will be explained later.
FP F01	Film Position: Displays frame F01 (picture of a bag).
DT 5,18/bat/	Display Text: Displays "bat" on line 5 starting in column 18 on the CRT.
DT 7,18/bag/	Displays "bag" on line 7 starting in column 18 on the CRT.
DT 9,18/rat/	Displays "rat" on line 9 starting in column 18 on the CRT.
AUP A01	Audio Play: Plays audio message A01. "Touch and say the word that goes with the picture."
L1 EP 30/ABCD1	Enter and Process: Activates the light-pen; specifies the time limit (30 sec.) and the problem identifier (ABCD1) that will be placed in the data record along with all responses to this problem. If a response is made within the time limit the computer skips from this command down to the CA (correct answer comparison) command. If no response is made within the time limit, the commands immediately following the EP command are executed.
AD 1/C4	Add: Adds one to the overtime counter (C4).
LD 1/S1	Loads one into the error switch (S1).
AUP A04	Plays message A04. "The word that goes with the picture is bag. Touch and say bag."
DT 7,16/→/	Displays arrow on line 7, column 16 (arrow pointing at "bag").
BR L1	Branch: Branches to command labeled L1. The computer will now do that command and continue from that point.
CA 1,7,3,18/C1	Correct Answer: Compares student's response with an area one line high starting on line 7 and three columns wide starting in column 18 of the CRT. If his response falls within this area, it will be recorded in the data with the answer identifier C1. When a correct answer has been made, the commands from here down to WA (wrong answer comparison) are executed. Then the program jumps ahead to the next PR. If the response does not fall in the correct area, the machine skips from this command down to the WA command.
BR L2/S1/1	Branches to command labeled L2 if the error switch (S1) is equal to one.
AD 1/C1	Adds one to the initial correct answer counter (C1).
L2 AUP A02	Plays audio message A02. "Good. Bag. Do the next one."
WA 1,5,3,18/W1 WA 1,9,3,18/W2	Wrong Answer: These two commands compare the student response with the areas of the two wrong answers, that is, the area one line high starting on line 5 and three columns wide starting in column 18, and the area one line high starting on line 9 and three columns wide starting in column 18. If the response falls within one of these two areas, it will be recorded with the appropriate identifier (W1 or W2). When a defined wrong answer has been made, the commands from here down to UN (undefined answer) are executed. Then the computer goes back to the EP for this problem. If the response does not fall in one of the defined wrong answer areas, the machine skips from this command down to the UN command.

Table 2—Continued

Commands	Explanation
AD 1/C2	Adds one to the defined wrong answer counter (C2).
L3 LD 1/S1	Loads one into the error switch (S1).
AUP A03	Plays message A03. "No."
AUP A04	Plays message A04. "The word that goes with the picture is bag. Touch and say bag."
DT 7,16/→/	Displays arrow on line 7, column 16.
UN	Undefined Wrong Answer: If machine reaches this point in the program, the student has made neither a correct nor a defined wrong answer.
AD 1/C3	Adds one to the undefined answer counter (C3).
BR L3	Branches to command labeled L3. (The same thing should be done for both UN and WA answers. This branch saves repeating the commands from L3 down to UN.)
PR	Prepares the machine for next problem.
LD 0/S1 FP F02 DT 5,18/card/ DT 7,18/cart/ DT 9,18/hard/	These commands prepare the display for the 2nd problem. Notice the new film position and new words displayed. The student was told to "do the next one" when he finished the last problem so he needs no audio message to begin this.
L4 EP 30/ABCD2	Light-pen is activated.
AD 1/C4 LD 1/S1 AUP A07 DT 5,16/→/ BR L4	These commands are done only if no response is made in the time limit of 30 seconds. Otherwise the machine skips to the CA command.
CA 1,5,4,18/C2	Compares response with correct answer area.
BR L5/S1/1 AD 1/C1 L5 AUP A05	Adds one to the initial correct answer counter unless the error switch (S1) shows that an error has been made for this problem. The student is told he is correct and goes on to the next problem. These commands are executed only if a correct answer has been made.
WA 1,7,4,18/W3 WA 1,9,4,18/W4	Compare response with defined wrong answer.
AD 1/C2 L6 LD 1/S1 AUP A06 AUP A07 DT 5,16/→/	Adds one to the defined wrong answer area and the error switch (S1) is loaded with one to show that an error has been made on this problem. The student is told he is wrong and shown the correct answer and asked to touch it. These commands are executed only if a defined wrong answer has been made.
UN	An undefined response has been made if the machine reaches this command.
AD 1/C3 BR L6	Adds one to the undefined answer counter and we branch up to give the same audio, etc. as is given for the defined wrong answer.

about 30 minutes to complete, requires in excess of 9,000 coursewriter commands for its execution.

A simple example will give you some feeling for the coding problem. The example is from a task designed to teach both letter discrimination and the meaning of words. A picture illustrating the word being taught is presented on the projector screen. Three words, including the word illustrated, are presented on the CRT. A message is played on the audio asking the child to touch the word on the CRT that matches the picture on the film projector. The student can then make his response using the light pen. If he makes no response within the specified time limit of 30 seconds, he is told the correct answer, an arrow points to it, and he is asked to touch it. If he makes a response within the time limit, the point that he touches is compared by the computer with the correct-answer area. If he places the light pen within the correct area, he is told that he was correct and goes on to the next problem. If the response was not in the correct area, it is compared with the area defined as a wrong answer. If his response is within this area, he is told that it is wrong, given the correct answer, and asked to touch it. If his initial response was neither in the anticipated wrong-answer area nor in the correct-answer area, then the student has made an undefined answer. He is given the same message that he would have heard had he touched a defined wrong answer; however, the response is recorded on the data record as undefined. The student tries again until he makes the correct response; he then goes on to the next problem.

To prepare an instructional sequence of this sort, the programmer must write a detailed list of commands for the computer. He must also record on an audio tape all the messages the student might hear during the lesson in approximately the order in which they will occur. Each audio message has an address on the tape and will be called for and played when appropriate. Similarly a film strip is prepared with one frame for each picture required in the lesson. Each frame has an address and can be called for in any order.

Table 1 shows the audio messages and film pictures required for two sample problems along with the hypothetical addresses on the audio tape and film strip. Listed in Table 2 are the computer commands required to present two examples of the problems described above, analyze the student's responses, and record his data record. The left column in the table lists the actual computer commands, and the right column provides an explanation of each command.

While a student is on the system, he may complete as many as 5 to 10 problems of this type per minute. Obviously, if all of the instructional material has to be coded in this detail the task would be virtually impossible. Fortunately, there are ways of simplifying coding procedure if parts of the instructional materials are alike in format and differ only in certain specified ways. For example, the two problems presented in Table 2 differ only in (*a*) the film display, (*b*) the words on the CRT, (*c*) the problem identifier, (*d*) the three audio addresses, (*e*) the row display of the arrow, (*f*) the correct answer area, and (*g*) the correct answer identifier. This string of code can be defined once, given a two-letter name, and used later by giving a one-line macro command.

The use of macros cuts down greatly the effort required to present many different but basically similar problems. For example, the two problems presented in Table 2 can be rewritten in macro format using only two lines of code: Problem 1: CM PW]F01]bat|bag]rat|A01|ABCD1]A04]-A02|A03]7]1,7,3,18|C1|; Problem 2: CM PW|-F02|card]cart]hard||ABCD2|A07]A05|A06|5]-1,5,4,18|C2|. The command to call a macro is CM, and PW is an arbitrary two-character code for the macro involving a picture-to-word match. Notice that in Problem 2 there is no introductory audio message; the "||" indicates that this parameter is not to be filled in.

The macro capability of the source language has two distinct advantages over code written command by command. The first is ease and speed of coding. The call of one macro is obviously easier than writing the comparable string of code. The second advantage is increase in accuracy. Not only are coding errors drastically curtailed, but if the macro is defective or needs to be changed, every occurence of it in the lesson coding can be corrected by modifying the original macro; in general, the code can stay as it is. The more standard the various problem formats, the more valuable the macro capability becomes. Apart from a few nonstandard introductory audio messages and display items, approximately 95% of the reading curriculum has been programmed using about 110 basic macros.

The macro command feature of the language has significant implications for psychological research. By simply changing a few commands in a particular macro, one can alter the flow of the teaching sequence whenever that macro is called in the program. Thus, the logic of an instructional sequence that occurs thousands of times in the reading curriculum can be redesigned by adding or modifying a few lines of code in a given macro. If, for example, we wanted to change the timing relations, the type of feedback, or characteristics of the CRT display in the task described above, it would require only a few lines of code in the PW macro and would not necessitate making changes at every point in the curriculum where the picture-to-word exercise occurred. Thus, a range of experimental manipulations can be carried out using the same basic program and display materials, and requiring changes only in the command structure of the macros.

As indicated in Table 2, a bank of switches and counters is defined in the computer and can be used to keep a running record on each student. There is a sufficient number of these registers so that quite sophisticated schemes of optimization and accompanying branching are possible. Thus, one is in a position to present a series of words and to optimize the number of correct responses to some stipulated criteria, for example, five consecutive correct responses for each of the words. Or one can select from an array of phrases choosing those phrases for presentation that have the greatest number of previous errors. As a consequence of these decisions, each student pursues a fundamentally different path through the reading materials.

SOME RESULTS FROM THE FIRST YEAR OF OPERATION

The Stanford CAI Project is being conducted at the Brentwood School in the Ravenswood School District (East Palo Alto, California). There were several reasons for selecting this school. It had sufficient population to provide a sample of well over 100 first-grade students. The students were primarily from "culturally disadvantaged" homes. And the past performance of the school's principal and faculty had demonstrated a willingness to undertake educational innovations.

Computerized instruction began in November of 1966 with half of the first-grade students taking reading via CAI and the other half, which functioned as a control group, being taught reading by a teacher in the classroom. The children in the control group were not left out of the project, for they took mathematics from the CAI system instead. The full analysis of the student data is a tremendous task which is still underway. However, a few general results have already been tabulated that provide some measure of the program's success.

Within the lesson material there is a central core of problems which we have termed main-line problems. These are problems over which each student must exhibit mastery in one form or another. Main-line problems may be branched around by successfully passing certain screening

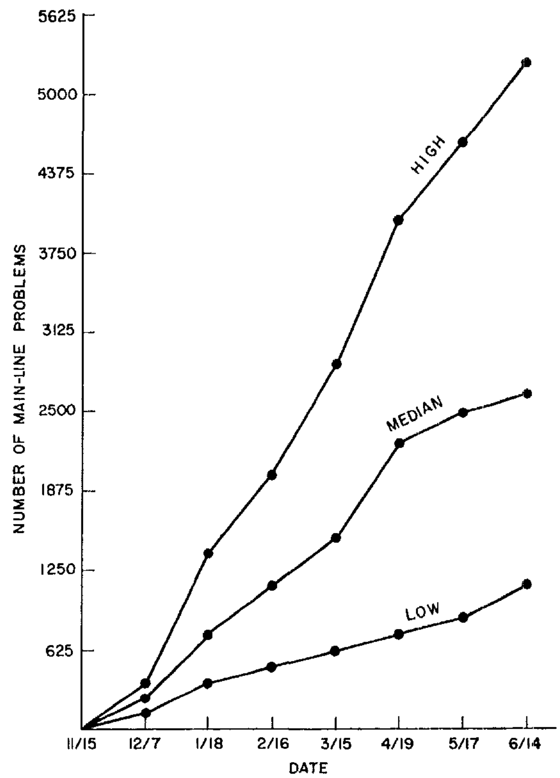

FIG. 5. Cumulative number of main-line problems for fastest, median, and slowest student.

tests, or they may be met and successfully solved; they may be met with incorrect responses, in which case the student is branched to remedial material. The first year of the project ended with a difference between the fastest and slowest student of over 4,000 main-line problems completed. The cumulative response curves for the fastest, median, and slowest students are given in Figure 5. Also of interest is the rate of progress during the course of the year. Figure 6 presents the cumulative number of problems completed per hour on a month-by-month basis again for the fastest, median, and slowest student. It is interesting to note that the rate measure was essentially constant over time for increase for the fast student.

From the standpoint of both the total number of problems completed during the year and rate of progress, it appears that the CAI curriculum is responsive to individual differences. The differences noted above must not be confused with a variation in rate of response. The difference in response rate among students was very small. The average response rate was approximately four per

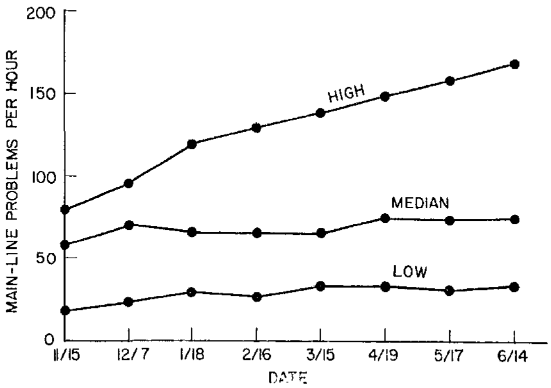

FIG. 6. Cumulative rate of progress for fastest, median, and slowest student.

minute and was not correlated with a student's rate of progress through the curriculum. The differences in total number of main-line problems completed can be accounted for by the amount of remedial material, the optimization routines, and the number of accelerations for the different students.

It has been a common finding that girls generally acquire reading skills more rapidly than boys. The sex differences in reading performance have been attributed, at least in part, to the social organization of the classroom and to the value and reward structures of the predominantly female primary grade teachers. It has also been argued on developmental grounds that first-grade girls are more facile in visual memorization than boys of the same age, and that this facility aids the girls in the sight-word method of vocabulary acquisition commonly used in basal readers. If these two arguments are correct, then one would expect that placing students in a CAI environment and using a curriculum which emphasizes analytic skills, as opposed to rote memorization, would minimize sex differences in reading. In order to test this hypothesis, the rate of progress scores were statistically evaluated for sex effects. The result, which was rather surprising, is that there was no difference between male and female students in rate of progress through the CAI curriculum.

Sex differences however might be a factor in accuracy of performance. To test this notion the final accuracy scores on four standard problem types were examined. The four problem types, which are representative of the entire curriculum, were Letter Identification, Word List Learning, Matrix Construction, and Sentence Comprehension. On these four tasks, the only difference between boys and girls that was statistically significant at the .05 level was for word-list learning. These results, while by no means definitive, do lend support to the notion that when students are removed from the normal classroom environment and placed on a CAI program, boys perform as well as girls in overall rate of progress. The results also suggest that in a CAI environment the sex difference is minimized in proportion to the emphasis on analysis rather than rote memorization in the learning task. The one problem type where the girls achieved significantly higher scores than the boys, word-list learning, is essentially a paired-associate learning task.

As noted earlier, the first graders in our school were divided into two groups. Half of them received reading instruction from the CAI system; the other half did not (they received mathematics instruction instead). Both groups were tested extensively using conventional instruments before the project began and again near the end of the school year. The two groups were not significantly different at the start of the year. Table 3 presents the results for some of the tests that were administered at the end of the year. As inspection of the table will show, the group that received reading instruction via CAI performed significantly better on all of the posttests except for the comprehension subtest of the California Achievement Test. These results are most encouraging. Further, it should

TABLE 3

POSTTEST RESULTS FOR EXPERIMENTAL AND CONTROL GROUPS

Test	Experimental	Control	p value
California Achievement Test			
Vocabulary	45.91	38.10	$<.01$
Comprehension	41.45	40.62	—
Total	45.63	39.61	$<.01$
Hartley Reading Test			
Form class	11.22	9.00	$<.05$
Vocabulary	19.38	17.05	$<.01$
Phonetic discrimination	30.88	25.15	$<.01$
Pronunciation			
Nonsense word	6.03	2.30	$<.01$
Word	9.95	5.95	$<.01$
Recognition			
Nonsense word	18.43	15.25	$<.01$
Word	19.61	16.60	$<.01$

be noted that at least some of the factors that might result in a "Hawthorne phenomenon" are not present here; the "control" group was exposed to CAI experience in their mathematics instruction. While that may leave room for some effects in their reading, it does remove the chief objection, since these students also had reason to feel that special attention was being given to them. It is of interest to note that the average Stanford-Binet IQ score for these students (both experimental and control) is 89.[3]

Owing to systems and hardware difficulties, our program was not in full operation until late in November of 1966. Initially, students were given a relatively brief period of time per day on the terminals. This period was increased to 20 minutes after the first 6 weeks; in the last month we allowed students to stay on the terminal 30 to 35 minutes. We wished to find out how well first-grade students would adapt to such long periods of time. They adapt quite well, and next year we plan to use 30-minute periods for all students throughout the year. This may seem like a long session for a first-grader, but our observations suggest that their span of attention is well over a half hour if the instructional sequence is truly responsive to their response inputs. This year's students had a relatively small number of total hours on the system. We hope that by beginning in the early fall and using half-hour periods, we will be able to give each student at least 80 to 90 hours on the terminals next year.

I do not have time to discuss the social-psychological effects of introducing CAI into an actual school setting. However, systematic observations have been made by a trained clinical psychologist, and a report is being prepared. To preview this report, it is fair to say that the students, teachers, and parents were quite favorable to the program.

Nor will time permit a detailed account of the various optimization routines used in the reading curriculum. But since this topic is a major focus of our research effort, it requires some discussion here. As noted earlier, the curriculum incorporates an array of screening and sequencing procedures designed to optimize learning. These optimization schemes vary in terms of the range of curriculum included, and it has been convenient to classify them as either short- or long-term procedures. Short-term procedures refer to decision rules that are applicable to specific problem formats and utilize the very recent response history of a subject to determine what instructional materials to present next. Long-term optimization procedures are applicable to diverse units of the curriculum and utilize a summarized form of the subject's complete response record to specify his future path through major instructional units.

As an example of a short-term optimization procedure, consider one that follows directly from a learning theoretic analysis of the reading task involved (Groen & Atkinson, 1966). Suppose that a list of m words is to be taught to the child, and it has been decided that instruction is to be carried out using the picture-to-word format described earlier. In essence, this problem format involves a series of discrete trials, where on each trial a picture illustrating the word being taught is presented on the projector screen and three words (including the word illustrated) are presented on the CRT. The student makes a response from among these words, and the trial is terminated by telling him the correct answer. If x trials are allocated for this type of instruction (where x is much larger than m), how should they be used to maximize the amount of learning that will take place? Should the m items be presented an equal number of times and distributed randomly over the x trials, or are there other strategies that take account of idiosyncratic features of a given subject's response record? If it is assumed that the learning process for this task is adequately described by the one-element model of stimulus sampling theory, and there is evidence that this is the case, then the optimal presentation strategy can be prescribed. The optimal strategy is initiated by presenting the m items in any order on the first m trials, and a continuation of this strategy is optimal over the remaining $x - m$ trials if, and only if, it conforms to the following rules:

1. For every item, set the count at 0 at the beginning of trial $m + 1$.

2. Present an item at a given trial if, and only if, its count is *least* among the counts for all items at the beginning of the trial.

3. If several items are eligible under Rule 2, select from these the item that has the smallest number of presentations; if several items are still eligible, select with equal probability from this set.

[3] More details on these and other analyses may be found in Atkinson (1967) and Wilson and Atkinson (1967).

4. Following a trial, increase the count for presented item by 1 if the subject's response was correct, but set it at 0 if the response was incorrect.

Even though these decision rules are fairly simple, they would be difficult to implement without the aid of a computer. Data from this year's experiment establish that the above strategy is better than one that presents the items equally often in a fixed order.

This is only one example of the type of short-term optimization strategies that are used in the reading curriculum. Some of the other schemes are more complex, involving the application of dynamic programming principles (Groen & Atkinson, 1966), and use information not only about the response history but also the speed of responding. In some cases the optimization schemes can be derived directly from mathematical models of the learning process, whereas others are not tied to theoretical analyses but are based on intuitive considerations that seem promising.[4]

Even if short-term optimization strategies can be devised which are effective, a total reading curriculum that is optimal still has not been achieved. It is, of course, possible to optimize performance on each unit of the curriculum while, at the same time, sequencing through the units in an order that is not particularly efficient for learning. The most significant aspect of curriculum development is with regard to long-term optimization procedures, where the subject's total response history can be used to determine the best order for branching through major instructional units and also the proper balance between drill and tutorial activities. It seems clear that no theory of instruction is likely to use all the information we have on a student to make instructional decisions from one moment to the next. Even for the most sophisticated long-term schemes, only a sample of the subject's history is going to be useful. In general, the problem of deciding on an appropriate sample of the history is similar to the problem of finding an observable statistic that provides a good estimate of a population parameter. The observable history sample may be regarded as an estimate of the student's state of learning. A desirable property for such a history sample would be for it to summarize all information concerning the current learning state of the student so that no elaboration of the history would provide additional information. In the theory of statistical inference, a statistic with an analogous property is called a sufficient statistic. Hence, it seems appropriate to call an observable sample history with this property a "sufficient history."

In the present version of the reading curriculum, several long-term optimization procedures have been introduced with appropriate sufficient histories. As yet, the theoretical rationale for these procedures has not been thoroughly worked out, and not enough data have been collected to evaluate their effectiveness. However, an analysis of long-term optimization problems, and what data we do have, has been instructive and has suggested a number of experiments that need to be carried out this year. It is my hope that such analyses, combined with the potential for educational research under the highly controlled conditions offered by CAI, will lay the groundwork for a theory of instruction that is useful to the educator. Such a theory of instruction will have to be based on a model of the learning process that has broad generality and yet yields detailed predictions when applied to specific tasks.

In my view, the development of a viable theory of instruction and the corresponding learning theory will be an interactive enterprise, with advances in each area influencing the concepts and data base in the other. For too long, psychologists studying learning have shown little interest in instructional problems, whereas educators have made only primitive and superficial applications of learning theory. Both fields would have advanced more rapidly if an appropriate interchange of ideas and problems had existed. It is my hope that prospects for CAI, as both a tool for research and a mode of instruction, will act as a catalyst for a rapid evolution of new concepts in learning theory as well as a corresponding theory of instruction.

[4] The learning models and optimization methods that underlie much of the reading curriculum are discussed in Atkinson and Shiffrin (1968), Groen and Atkinson (1966), Rodgers (1967), and Wilson and Atkinson (1967).

REFERENCES

ATKINSON, R. C., & HANSEN, D. N. Computer-assisted instruction in initial reading: The Stanford Project. *Reading Research Quarterly*, 1966, 2, 5–25.

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and* *theory.* Vol. 2. New York: Academic Press, 1968, in press.

FISHMAN, E. J., KELLER, L., & ATKINSON, R. C. Massed vs. distributed practice in computerized spelling drills. *Journal of Educational Psychology,* 1968, **59,** in press.

GROEN, G. J., & ATKINSON, R. C. Models for optimizing the learning process. *Psychological Bulletin,* 1966, **66,** 309–320.

HANSEN, D. N., & ROGERS, T. S. An exploration of psycholinguistic units in initial reading. Technical Report 74, 1965, Stanford University, Institute for Mathematical Studies in the Social Sciences.

RODGERS, T. S. Linguistic considerations in the design of the Stanford computer-based curriculum in initial reading. Technical Report 111, 1967, Stanford University, Institute for Mathematical Studies in the Social Sciences.

SUPPES, P. The uses of computers in education. *Scientific American,* 1966, **215,** 206–221.

SUPPES, P., JERMAN, M., & GROEN, G. J. Arithmetic drills and review on a computer-based teletype. *Arithmetic Teacher,* 1966, April, 303–308.

WILSON, H. A., & ATKINSON, R. C. Computer-based instruction in initial reading: A progress report on the Stanford project. Technical Report 119, 1967, Stanford University, Institute for Mathematical Studies in the Social Sciences. (To be published in H. Levin & J. Williams, Eds.), *Basic studies in reading.* New York: Harper and Row.)