J. EDUCATIONAL TECHNOLOGY SYSTEMS, Vol. 11(1), 1982-83

EFFECTIVENESS OF A COMPUTER-BASED READING COMPREHENSION PROGRAM FOR ADULTS*

STEPHEN M. ALESSI**

MARTIN SIEGEL

DOROTHY SILVER

University of Illinois at Urbana

HANK BARNES

Sheridan Correctional Center

ABSTRACT

A formative evaluation of the instructional effectiveness of twenty hours of computer-based reading comprehension instruction for adults was conducted. The lessons taught reading comprehension subskills called 'information finding' and 'paraphrasing.' Thirty-six adults studied either the reading lessons or computer-based mathematics lessons for the same length of time, about two months. The group that studied reading lessons improved significantly in pre to posttest performance relative to mathematics students. The increased test performance was retained on another posttest given a month later. The study supports the value of a complete computer-based reading comprehension curriculum following similar instructional strategies.

INTRODUCTION

This paper discusses the design and formative evaluation of a computer-based instruction program in reading for disadvantaged adults. The contents of the paper are: 1) the content and structure of the program, 2) the design of the evaluation, 3) a description of test materials used in the evaluation, 4) the evaluation procedures, 5) the results of the evaluation, and 6) our conclusions of the program's strengths and weaknesses.

* This research was supported by a State of Illinois, Department of Corrections grant to the PLATO Education Group, CERL, University of Illinois at Urbana-Champaign.

** Stephen M. Alessi is now at the University of Iowa.

THE PROGRAM

The PLATO Corrections Project began in 1975 with the mission to implement computer-based educational programs in Illinois adult correctional centers [1]. The project has installed terminals connected to the PLATO computer-based education network in five adult correctional centers and has trained teachers at each center in the use of the available computer-based lessons.

Many of the lessons made available were not written as a part of the project, but by previous curriculum development projects for elementary schools, high schools, and community colleges. The PLATO Corrections Project has, however, developed materials in reading, vocabulary, social studies, math, and spelling.

The majority of lessons the project has developed have been in reading and vocabulary. The reading lessons include about twenty hours of instruction emphasizing reading comprehension, a similar number of hours of reading practice for fluency and accuracy, and some drills on word decoding for students with very minimal reading skills. The formative evaluation to be described was for the twenty hours of reading comprehension instruction. A more complete description of that program can be found in reference [2]. An overview of the program is necessary here to understand the evaluation.

The development of the reading comprehension program [3] is still in progress, with lessons existing, underway, or envisioned on: finding information in text, paraphrasing text and answering paraphrase questions about texts, synthesis of text (identifying main ideas, themes, etc.), making inferences from text (comprehending information which is implied though never explicitly stated in a text), dealing with logical relationships in text, vocabulary development, following directions from text, skills for taking reading comprehension tests, and self-monitoring of comprehension (being aware of whether or not you are understanding and adjusting your reading strategy as a result of that awareness).

At the time this evaluation was conducted, lessons were completed and in use by students on finding information in text (henceforth called simply finding information), paraphrasing text and answering paraphrase questions about texts (henceforth called paraphrasing) and vocabulary development. We chose to evaluate only the lessons comprising the first two areas, finding information and paraphrasing.

The lessons on finding information and paraphrasing require about forty student sessions of twenty to forty minutes each. Each session uses two or three instructional formats. A format is a particular instructional pattern for teaching a subskill. There are seven formats for finding information and five formats for paraphrasing. A particular format is reused many times with different words, sentences, or paragraphs and each such unique presentation of a format we call a task.

Within a single session the student first receives and responds to a number of tasks from a particular format. The student continues to receive tasks from the

format until the student is performing well. The student then begins tasks from a second format, and perhaps a third and fourth. For example, in the very first session the student receives tasks from two formats relevant to finding information. In later sessions other (more difficult) formats are introduced. After about ten sessions formats relevant to paraphrasing are introduced, and thereafter all sessions include instruction and practice on both finding information and paraphrasing.

The earlier formats are generally cued. That is, the format is made easier in some way by giving prompts to the nature of the response. Some later formats are identical to earlier ones except that they are uncued. As uncued formats are introduced, cued formats are dropped permanently. There is also some blending of formats between finding information and paraphrasing. Once paraphrasing formats are introduced, some of the finding information formats begin to ask questions about information in sentences in paraphrased form.

We have attempted to adhere to some basic instructional principles in the instructional formats and in the program as a whole. The primary guiding principles of a format are: 1) present the concept being taught in as clear and straightforward a manner as possible; 2) make the skill being taught both objective and overt; and 3) provide effective correction and review procedures. The guiding principles of the program as a whole are: 1) that it be self-paced; 2) that instruction be individualized; and 3) that the student achieve mastery of subskills before advancing.

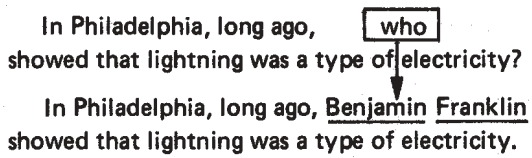
The finding information formats aim to teach students to find answers to questions asking who, what, why, where, when and how. The simplest format begins by requiring the student to underline the part of a sentence that gives a particular piece of information. For instance, given the sentence:

"In Philadelphia, long ago, Benjamin Franklin showed that lightning was a type of electricity."

the student might be asked to underline the words that tell:

"Who showed that lightning was a form of electricity?"

If the student makes a response error in this format, the computer shows the student the correct answer and how the answer given was wrong. If the student responded by underlining *Philadelphia* the computer would display "That tells *where* lightning was shown to be a type of electricity. What part tells *who* showed that lightning was a type of electricity?" If the student's error persists, a more detailed graphic correction such as the following would occur:



In this case the correction procedure seeks to discriminate the words which comprise responses to different questions.

The underlining response and these correction procedures are an example of making the concept presentation and the response clear, objective, and overt.

The tasks generated in this format use a flipbook procedure. On subsequent tasks 'long ago' might be replaced with 'in the 18th century' and 'Benjamin Franklin' might be replaced with 'an American' or other phrases. Different root sentences have the who, what, why, where, when and how components in different orders to avoid students learning misrules such as 'the *where* component always comes at the beginning.' The flipbook procedure has three purposes. First, it allows the generation of many sentences without storing a lot of sentences in the computer. Secondly, the conceptual approach to instruction implies that you often want the program to change a sentence in a very small (but important) way to illustrate a concept, such as by distinguishing between examples and nonexamples of a concept. For instance, successive sentences in a particular sequence may be identical except for a change in the *who* part of a sentence in order to focus student attention on that part. Lastly, the computer program keeps track of particular error types, and the flipbook procedure selects and constructs sentences having the characteristics needed to follow up the errors.

In addition to the formats which require underlining of 'the part that tells,' there are formats teaching and practicing recognition of whether a particular part of a sentence tells something (the opposite of the first format) and recognition of whether a sentence tells something, in addition to identifying what part (if any) does so.

Each format includes the flipbook procedure for generating similar but different (when needed) sentences. Each format includes graphic correction procedures which underline, draw boxes around, point to, and connect with arrows to show the correct answer and discriminate correct versus incorrect responses. Each format makes the subskill clear and overt by interactive graphic displays. Most formats include cued and uncued versions. And lastly, each format includes a criterion formula to determine when a student may progress to other formats.

The paraphrase track teaches the student to deal with alternate wordings. That is, the student must recognize if two different sentences mean the same thing, and answer questions that do not use words identical to those in a text.

The paraphrase formats make use of the same general methods as those for finding information. The skills are developed through clear text and graphic demonstrations. The skills are made overt by the proper selection of student response methods. The formats include both cued and uncued forms. The flipbook method is used to generate similar yet different tasks. And tasks are selected for and eliminated from a student's instructional sequence based on performance.

The first paraphrase format requires the student to replace part of a sentence with a synonomous phrase to make it a partial paraphrase and demonstrate the semantic equality of the sentences. Later formats replace all the phrases of sentences (one at a time) to make complete paraphrases. The last (and most difficult) formats require the student to: 1) locate a sentence in a full page passage that paraphrases a separate target sentence; and 2) identify in multiple choice format which of a number of sentences has the same meaning as a target sentence. The earlier 'paraphrase building' formats develop the simpler skills needed for these more difficult paraphrase formats.

DESIGN OF THE EVALUATION STUDY

We wished to evaluate primarily whether the lessons succeeded in teaching the subskills which they explicitly sought to teach, finding information and paraphrasing. Secondarily we wished to see whether the students retained the subskills over a period of time, and whether the instruction produced any general improvement in reading comprehension.

It was necessary to control for the general effect of studying on a computer. Our design therefore included two groups of students, one studying the reading lessons on PLATO and the other studying other (non-reading) lessons on PLATO. Since the non-reading students in fact spent almost all their time in math lessons, we will refer to the two groups as reading students and math students.

All students were pretested with the same reading achievement test. All students received an immediate posttest after reading students completed their lessons. Reading students were additionally given a delayed posttest (to measure skill retention) about a month after the immediate posttest.

The basic design was thus a two-way mixed analysis of variance. The first factor was instruction (reading or no reading, between subjects) and the second factor was test (pre or post, within subjects). Secondarily we had an 'overlapping' one-way analysis of variance for reading students with three repeated measures: pretest, immediate posttest, and delayed posttest.

MATERIALS

The materials included: 1) the reading lessons which have already been described; 2) the non-reading lessons (mostly math) studied by control students, which will not be described since their content is not relevant to the evaluation; and 3) the reading achievement pretests and posttests which will be described now.

As already mentioned, our primary purpose was to see if the lessons succeeded in teaching the reading subskills we call finding information and paraphrasing, while a secondary purpose was to see if general reading comprehension skills increased. To assess this second question, test questions asking about the main idea of texts were included.

We developed finding information and paraphrasing questions for sentences, paragraphs, and multi-paragraph passages. Main idea questions were only developed for passages.

Three reading achievement test forms were developed with completely different texts and questions, but which all followed the following pattern:

6 sentences, the 1st, 3rd, and 5th each followed by a question requiring the student to find information in the sentence and the 2nd, 4th, and 6th followed by one paraphrase question each.

3 paragraphs with each paragraph followed by 2 questions, the first requiring the student to find information and the second a paraphrase question.

5 passages with each passage followed by three questions, the first requiring the student to find information, the second a paraphrase question, and the third always asking "Which sentence BEST describes the whole passage?" and followed by four possible main ideas for the student to select from.

Following are examples of each question type.

Finding information in a sentence:

Although most of the world's diamonds come from Africa, some can be found in the United States near Murfreesboro, Arkansas.

Most diamonds come from
a) the United States.
b) Africa.
c) Arkansas.
d) Murfreesboro.

Paraphrase question about a sentence:

Annie was only nine when she took her father's big Kentucky rifle into the woods to hunt food for her hungry family.

Annie used a gun to
a) protect herself.
b) frighten a robber.
c) get something to eat.
d) practice target shooring.

Finding information question (first) and paraphrase question (second) about a paragraph:

Since the moment you took your first breath, your lungs have never been without air. The ribs and their muscles, by expanding and contracting, make the lungs act as a bellows. When a bellows expands, a partial vacuum is created inside, and air rushes in to fill the vacuum. When the bellows is pressed together, the air is forced out. Similarly, when the lungs are

expanded, they are full of air. The blood takes the oxygen from the air and leaves the waste gas, carbon dioxide, in the lungs. The carbon dioxide is squeezed out of the lungs when they contract.

Expansion of the lungs
a) pulls in air.
b) pushes out air.
c) pushes out waste gases.
d) pulls in carbon dioxide.

The contracting and expanding of the lungs are caused by the
a) blood stream.
b) air going in and out of them.
c) muscles inside the lungs.
d) bones and muscles of the chest.

Finding information question (first), paraphrase question (second), and main idea question (third) about a multi-paragraph passage:

Scientists in the United States are trying to breed a tree that will be ideal for city life. They say that in many cities the air is so filled with poisons that the plain, old-fashioned tree is doomed to disappear.

The ideal city tree must have shallow roots. Roots that go down too far get in the way of underground pipes. It must have fairly short branches. Branches that reach too high interfere with overhead lines. The custom made tree must not drop messy fruit on people's heads, either.

But most important, the ideal tree for the city must be tough. It must be so tough that insects can't harm it. It must resist disease. And it must be able to withstand the effects of pollution in the air.

The ideal tree would produce
a) messy fruit.
b) prickly fruit.
c) a lot of fruit.
d) no fruit.

City trees must be
a) protected from polluted air.
b) resistant to many dangers.
c) tall and pretty.
d) inexpensive and fast growing.

Which sentence BEST describes the whole passage?
a) The tree of the future must resist disease.
b) Scientists can develop trees for different purposes.
c) Scientists are trying to create a new type of tree suitable for the city.
d) The air in cities is bad for trees.

All of the questions were multiple choice as in the examples above. They were based on short reading passages extracted from the SRA Reading Laboratory 4a [4]. Sentences, paragraphs, and passages were usually changed slightly. The questions were always slightly changed or completely replaced. All items were developed without regard to test form. They were then randomly assigned to test form one, two, or three.

The three test forms were pilot tested on 37 students to determine if the test difficulty was appropriate and to check for bad texts or questions. The tests were found to be satisfactory with one or two minor wording changes. Due to the limited number of appropriate students available in the prison schools, insufficient students were available to do an analysis of test reliability.

PROCEDURE

The program was evaluated with pre-GED students at the Sheridan Correctional Center in Sheridan, Illinois. Pre-GED students are those who have completed basic education courses but are not yet ready for a high school equivalency (General Educational Development or GED) program. Pre-GED programs correspond roughly (and varying from school to school) to grades four through seven.

The education program at Sheridan, like most correctional centers, is partially open entry and exit. Thus, students appropriate for the evaluation were started in it as they entered the program.

Students were considered appropriate for the study if their score on the *Test of Adult Basic Education* (TABE) was between grades four and six. The TABE is the standard program placement test for all Illinois correctional centers.

Whenever two appropriate students were ready for the study, they were first given Form 1 of the reading achievement test. The testing was done off-line by the instructor. He instructed students to indicate the best answer for each question, and try to answer all questions. He noted the time required to complete the test as each student finished. The two students were then randomly assigned, one to study the reading lessons and the other to study math lessons. After about two weeks thirty-six students had taken the pretests and begun studying on the computer.

Students studying the reading lessons required an average of two months to complete all forty sessions. They studied the lessons on a fairly regular basis (about four days a week) spending about twenty to forty minutes per session and doing one or two sessions each day they worked.

When a reading student completed the reading lessons, both he and the math student who began at the same time were given Form 2 of the reading achievement test. In this way, time and dates working on PLATO were roughly controlled for the two groups of students. Administration procedures for Form 2 of the reading achievement test were the same as for Form 1.

The reading student then began studying math lessons on PLATO while the math student began studying the reading lessons. There were a number of reasons for this switch. First, we wanted to assess the reading students' retention of newly acquired reading comprehension subskills without continued instruction or structured practice in reading. Second, the needs of the students in the program demanded that students learn appropriate math and reading skills in preparation for GED classes. Third, we had originally hoped that the math students would also finish the reading lessons (after test Form 2) and we could then assess their reading comprehension improvement following the lessons as well.

Unfortunately, almost all students would have changed to different classes before the math students would have completed the reading lessons. As a result, we administered Form 3 of the reading achievement test only to reading students after they switched to and were studying math for a little over one month. Some students were given Form 3 slightly earlier because they were leaving the institution and would have been lost from the study if not tested before they left. Administration of the third test to reading students was identical to the other test administrations.

Five students were lost from the reading group while none were lost from the math group before taking the first posttest. The reason reported was always transfer out of the school program or the institution. It is nevertheless possible that some sampling bias may have been introduced, since transfers out of the school may tend to be more advanced students. This possibility is not supported, however, by the students' pretest scores. Furthermore, it would have biased the results against the experimental group rather than in favor of it.

RESULTS

Students studying the reading comprehension lessons required a mean of sixty-seven days (S.D. = 18) and twenty on-line hours (S.D. = 5) to complete the lessons. The mean time between the immediate posttest and the delayed posttest was forty-nine days (S.D. = 16). While studying the lessons, students' response performance for all tasks was recorded. The students responded correctly to 76 per cent (S.D. = 5.3) of all questions in the lessons.

The experimental analysis of groups and tests was done separately for each type of question: finding information, paraphrase, and main ideas.

Table 1 shows the cell means and ANOVA summary tables for the eleven finding information items. The first ANOVA represents the analysis of pre and post tests for reading and math students. The second ANOVA represents the pre, post, and delayed posttest for the reading students who stayed in the program long enough to complete all three tests.

The increase in scores for reading students was 25 per cent while it was only 3 per cent for math students. All significance tests are based on alpha = .05. The

Table 1. Finding Information Questions

	Pretest (All S's)	*Immediate Posttest (All S's)*	*Pretest (S's with Delay Scores)*	*Immediate Posttest (S's with Delay Scores)*	*Delayed Posttest (S's with Delay Scores)*
Reading Students	$\bar{x}$ = 5.8 = 53% SD = 1.8 n = 13	$\bar{x}$ = 8.5 = 78% SD = 1.3 n = 13	$\bar{x}$ = 6.2 = 56% SD = 1.8 n = 10	$\bar{x}$ = 8.6 = 78% SD = 1.4 n = 10	$\bar{x}$ = 8.7 = 79% SD = 1.9 n = 10
Math Students	$\bar{x}$ = 7.9 = 72% SD = 1.8 n = 18	$\bar{x}$ = 8.2 = 75% SD = 2.0 n = 18			

Analysis of Variance Summary Table for Group by Test Analysis

Source	*df*	*SS*	*MS*	*F*	*Probability*
Group	1	11.98	11.98	2.34	.14
Test	1	25.81	25.81	23.25	.00004
Subject	29	148.79	5.13		
Group by Test	1	22.00	22.00	19.82	.00012
Test by Subject	29	32.19	1.11		

Analysis of Variance Summary Table for Delayed Posttest Analysis

Source	*df*	*SS*	*MS*	*F*	*Probability*
Test	2	40.07	20.03	12.91	.00033
Subject	9	51.17	5.80		
Test by Subject	18	27.93	1.55		

interaction of group and test ($F_{1,29} = 19.8, p < .0015$) is very significant. A comparison of simple effects shows that the increase for students following reading instruction is significant ($F_{1,29} = 42.7, p < .01$) while the increase for math students is not. The 1 per cent increase from immediate posttest to delay posttest is insignificant. The important thing there is that the test scores did not drop.

This increase in test performance is not only statistically significant but educationally significant as well. That is, we believe a great enough increase was obtained to be worth the students' time and effort to achieve it. The reading students not only improved their test performance but retained that improvement after more than a month, while those using PLATO but not studying the reading lessons showed no appreciable change. Of course we must

be wary of a few aspects of the data. Although we made an effort to construct equivalent test forms, their equivalence and reliability is undetermined. One might argue that if test reliability is in fact low, that the results could be explained on the basis that the pretest for reading students was depressed due to the low reliability. It must be noted in this regard that math students did considerably better on the same pretest. Secondly, one cannot claim that a hypothesis has been supported based on a non-significant statistical result. Our claim that students retained what they learned, based on the one-way ANOVA, must be considered a logical rather than a statistical argument.

Table 2 shows the means and ANOVA summary tables for the same analyses as Table 1, but for the eleven paraphrase items.

The increase in scores for reading students was 11 per cent while there was a *decrease* of 10 per cent for math students. The interaction is significant

Table 2. Paraphrase Questions

	Pretest (All S's)	*Immediate Posttest (All S's)*	*Pretest (S's with Delay Scores)*	*Immediate Posttest (S's with Delay Scores)*	*Delayed Posttest (S's with Delay Scores)*
Reading Students	$\bar{x}$ = 6.3 = 57% SD = 2.2 n = 13	$\bar{x}$ = 7.5 = 68% SD = 1.2 n = 13	$\bar{x}$ = 6.2 = 56% SD = 2.2 n = 10	$\bar{x}$ = 7.7 = 70% SD = 1.2 n = 10	$\bar{x}$ = 7.7 = 70% SD = 2.0 n = 10
Math Students	$\bar{x}$ = 8.0 = 73% SD = 1.9 n = 18	$\bar{x}$ = 6.9 = 63% SD = 2.3 n = 18			

Analysis of Variance Summary Table for Group by Test Analysis

Source	*df*	*SS*	*MS*	*F*	*Probability*
Group	1	5.72	5.72	1.02	.32
Test	1	.40	.40	.19	.66
Subject	29	163.15	5.62		
Group by Test	1	19.36	19.36	9.24	.005
Test by Subject	29	60.74	2.09		

Analysis of Variance Summary Table for Delayed Posttest Analysis

Source	*df*	*SS*	*MS*	*F*	*Probability*
Test	2	15.00	7.50	4.55	.025
Subject	9	62.13	6.90		
Test by Subject	18	29.67	1.65		

($F_{1, 29} = 9.2, p < .006$). The comparison of simple effects shows that the reading students' improvement was just barely significant ($F_{1, 29} = 4.47, p < .05$) as was the decrease in scores for math students ($F_{1, 29} = 5.2, p < .05$). Again, there was no score change for reading students from immediate posttest to delayed posttest. What the students learned they appear to have retained.

The above results for paraphrase questions are not as straightforward to interpret as those for finding information questions. In addition to the cautionary considerations already noted, the increase for reading students on paraphrase questions was significant but not nearly as great as it was for finding information questions. We might question whether this is an educationally significant result, even if statistically significant. It is also strange that the scores on these questions for students who did not study reading lessons dropped significantly. A possible explanation is that there was a test difficulty effect for paraphrase questions between the test forms. More will be said on that later.

Table 3 shows the means and ANOVA summary tables for the same analyses

Table 3. Main Idea Questions

	Pretest (All S's)	*Immediate Posttest (All S's)*	*Pretest (S's with Delay Scores)*	*Immediate Posttest (S's with Delay Scores)*	*Delayed Posttest (S's with Delay Scores)*
Reading Students	$\bar{x}$ = 2.1 = 42% SD = 1.1 n = 13	$\bar{x}$ = 2.2 = 43% SD = 0.9 n = 13	$\bar{x}$ = 1.9 = 38% SD = 1.0 n = 10	$\bar{x}$ = 2.2 = 44% SD = 0.9 n = 10	$\bar{x}$ = 2.3 = 46% SD = 1.2 n = 10
Math Students	$\bar{x}$ = 2.3 = 46% SD = 1.4 n = 18	$\bar{x}$ = 2.7 = 53% SD = 1.2 n = 18			

Analysis of Variance Summary Table for Group by Test Analysis

Source	*df*	*SS*	*MS*	*F*	*Probability*
Group	1	1.92	1.92	1.08	.31
Test	1	1.03	1.03	.92	.34
Subject	29	51.63	1.78		
Group by Test	1	.37	.37	.33	.57
Test by Subject	29	32.60	1.12		

Analysis of Variance Summary Table for Delayed Posttest Analysis

Source	*df*	*SS*	*MS*	*F*	*Probability*
Test	2	.87	.43	.66	.53
Subject	9	18.80	2.09		
Test by Subject	18	11.80	.66		

as Table 1, but for the five main idea items. There are no significant differences between means for any of the tests or groups.

Simple correlation coefficients were computed between performance and time in the reading lessons and the pretest/posttest scores. Nothing surprising was found. Performance (calculated as the percentage of correct responses across all lessons) correlated positively with almost all scores for finding information and paraphrase questions, both on pretests and posttests. It is very likely that this is because performance in a lesson is a measure of ability and/or motivation, which would in turn predict test scores fairly well. Similarly, time to complete the lessons correlated negatively with most test scores for finding information and paraphrase questions. Better students generally take less time in a program which requires mastery of the material.

DISCUSSION AND CONCLUSIONS

There is evidence that the formats which teach how to find information in text are effective. Based on pre and posttest data, students who studied the reading lessons improved markedly on that subskill and retained the subskill improvement after more than a month. Control students showed no change.

The formats teaching the student how to answer paraphrase type questions showed statistically significant effectiveness, but not as much as the finding information formats. How do we interpret these results?

First, we have already pointed out the uncertain reliability of our specially constructed achievement tests. Second, we must admit the possibility that the paraphrase formats are not as effective as the finding information formats. While this explanation accounts for less improvement (compared to finding information) for reading students, it does not account for the math students' scores decreasing.

Another explanation is that there is more to answering paraphrase questions than we supposed in the lessons. Specifically, general vocabulary and knowledge of the world probably have a strong effect on dealing with paraphrased text and questions. We might suppose that our instruction on answering paraphrase questions might be improved by, at least, concomitant instruction in a vocabulary program. But an evaluation such as we conducted would still be dependent on the student's vocabulary knowledge including the specific vocabulary which was essential to the paraphrase questions in our pretests and posttests. These considerations, however, still do not jointly explain the reading students' score increase and the math students' score decrease.

The decrease in score for math students might well be due to test difficulty. That is, the paraphrase questions on Form 2 (and perhaps Form 3) might have been significantly more difficult than those on test Form 1. Although, we cannot be sure without further use of the tests, this would explain how students with no treatment would drop significantly from pretest to posttest. If this is so, the reading students' significant increase, despite a harder posttest, is even more

significant than our analysis implies. This is just hypothetical, of course, and even assuming it is so, we cannot accurately assess how much more improvement reading students would have shown given equally difficult pretests and posttests.

One other explanation should be considered for the math students' decrease. Their pretest scores may have been spuriously high. Perhaps those students felt particularly motivated when they took that test. Perhaps they were just lucky. This explanation is supported by observing that the math students' mean pretest score was 73 per cent while the reading students' mean pretest score was 57 per cent. This is quite a large difference for students presumably (at the beginning of the study) identical. Following this explanation, the math students' score decrease might have been due to chance, even though statistically we are told this is probably not so. If that be the case, we are again left with the conclusion that the reading students increased significantly but not as greatly as we could have hoped. We then return to the explanation that the paraphrase formats need to be improved or supplemented with vocabulary or other instruction.

We tested for possible improvement of a student's ability to answer questions about the main idea of passages even though our program does not yet contain formats teaching that subskill. As we expected there was no improvement of that subskill for any students. This reinforces our position that a student learns what he or she is taught. We taught students to find information and answer paraphrase questions and that is what our students learned.

Lastly, we considered the correlations found between instructional time and performance and pretest/posttest performance. We might have hoped for no significant correlations. A mastery based program aims to bring students up to the same level of achievement regardless of ability. But most educators have come to recognize this as a little bit too idealistic. We can still pursue the ideal that all of our students can learn well if taught well, while at the same time admitting that some will always learn a little bit better.

In conclusion, there is evidence that the reading comprehension lessons are teaching what they were designed to teach. Students' finding information skills improved considerably and did not diminish after more than a month. Students' paraphrase question answering skills improved significantly but not as much. The increase might have been masked by a test difficulty effect. On the other hand we might really need to improve or supplement the paraphrase formats.

On the basis of these results we will continue with development of reading comprehension lessons following the instructional design principles embodied in the current lessons. We will further investigate improvements to the existing formats that might require them. Lastly, we will continue to do formative evaluations on the instructional programs we have been or will be developing.

On this last point we wish to urge others developing instructional materials for computer-based delivery to do the same. We recognize that our own evaluation was rather limited. We investigated only the achievement of specific

learning goals and did not consider student or teacher attitudes, the effect of the program on the institution, and a host of other possible effects. We certainly agree that other effects besides achievement of these particular instructional goals are important. We also recognize that due to the constraints of the population (number of appropriate students and the length of participation needed) questions of test reliability and sampling error arise.

What we wish to encourage most is to do at least the minimum evaluation of whether an instructional program succeeds in the specific goals it sets out to accomplish. In situations where more students are available for longer periods of time (such as traditional elementary and secondary schools) even more controlled evaluations should be conducted.

With the advent of low cost microcomputers for homes and schools, individuals and publishing houses alike are beginning to get into the business of developing instruction for delivery on computers. There will be a lot of instructional material produced as a result. For the good of teachers, parents, and learners, we urge that those who develop instruction also strive to test and improve their programs until they succeed in teaching what they are intending to teach.

REFERENCES

1. M. A. Siegel, Computer-Based Education in Prison Schools, *Journal of Educational Technology Systems*, 7:3, 1978-1979.
2. _____, "PLATO Computer-Based Instruction in Reading Comprehension," paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980.
3. M. A. Siegel, L. V. Dibello, E. C. Bruner, and J. M. Felty, PCP Reading Comprehension Lessions I-IV, Computer-Based Education Research Laboratory, Urbana, Illinois, 1978, (PLATO lesson).
4. D. H. Parker, SRA Reading Laboratory 4a, Science Research Associates, Inc., Chicago, 1973.

ACKNOWLEDGEMENT

We wish to thank Mr. Robert Kaplan, Educational Administrator of Sheridan Correctional Center, for his support of this research.

Direct reprint requests to:

Stephen M. Alessi
University of Iowa
College of Education
Division of Psychological and Quantitative Foundations
Rom 361 Lindquist Center
Iowa City, Iowa 52242