DOCUMENT RESUME

ED 324 833 EC 231 835

AUTHOR	Carnine, Douglas W.
TITLE	Computer Assisted Instruction in "'-her Order Skills for Mildly Handicapped Students: rrogrammatic Research on Design Principles. Final Report.
INSTITUTION	Oregon Univ., Eugene. Coll. of Education.
SPONS AGENCY	Special Education Programs (ED/OSERS), Washington, DC.
PUB DATE	1 Oct 87
GRANT	G008400660-02
NOTE	142p.; Numerous poorly photocopied pages.
PUB TYPE	Reports - Research/Technical (143) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Computer Assisted Instruction; Drills (Practice); Educational Technology; *Instructional Design; Instructional Development; *Instructional Effectiveness; Learning Disabilities; *Mild Disabilities; Secondary Education; Simulation; Teacher Role; Tutorial Programs; Videodisks

ABSTRACT

This report summarizes a series of eight research studies related to the use of computer-assisted instruction (CAI) with mildly handicapped students at the junior high or high school level. Through videodisc and CAI the studies isolated the effects of the following design variables: (1) review cycles; (2) size of teaching sets; (3) explicit strategies; and (4) correction procedures. Studies involved three different kinds of CAI--drill and practice, tutorials, and simulations. Results indicated that properly designed CAI can be effective as an instructional medium if attention is paid to the academic task, the stage of instruction, and the role of the teacher. Detailed reports are presented in the form of preprints or reprints of journal articles with the following titles: "Applying Instructional Design Principles to CAI for Mildly Handicapped Students: Four Recently Conducted Studies" (John Woodward et al.); "Effects of Instructional Design Variables on Vocabulary Acquisition of LD Students: A Study of Computer-Assisted Instruction" (Gary Johnson et al.); "Elaborated Corrective Feedback and the Acquisition of Reasoning Skills: A Study of Computer-Assisted Instruction" (Maria Collins et al.); "Teaching Problem Solving through Computer Simulations" (John Woodward et al.); "The Effectiveness of Videodisc Instruction in Teaching Fractions to Learning-Disabled and Remedial High School Students" (Bernadette Kelly et al.); and "Closing the Performance Gap in Secondary Education" (John Woodward et al.). References accompany each paper. (DB)

FINAL REPORT

Computer Assisted Instruction in Higher Order Skills for Mildly Handicapped Students: Programmatic Research on Design Principles

G008400660-02

Principal Investigator: Douglas W. Carnine, University of Oregon

October 1, 1987

FINAL REPORT CERTIFICATION FORM

GRANT NUMBER: G008400660

INSTITUTION: University of Oregon, College of Education
Eugene, Oregon 97403

PRINCIPAL INVESTIGATOR: Douglas W. Carnine

COMPETITION: Research In Education of the Handicapped (84.023)

PROJECT OFFICER: Doris L. Cargile

TITLE: Computer Assisted Instruction in Higher Order Skills for Mildly Handicapped Students: Programmatic Research on Design Principles

SUMMARY OF FINAL REPORT

This project included a series of eight research studies related to the use of computer assisted instruction (CAI) with mildly handicapped students, either at the junior high or high school level. These studies were designed to examine a range of instructional design principles that have been previously demonstrated to be effective techniques in non-computer studies and the applicability of these for improving the effectiveness of different types of CAI programs and the videodisc. Through videodisc and computer assisted instruction, project staff were able to isolate the effects of the following instructional design variables: (1) review cycles, (2) size of teaching sets, (3) explicit strategies, and (4) correction procedures.

Several of the studies in the project involved three different kinds of CAI—drill and practice, tutorials, and simulations. Thus, a secondary interest of these studies was to examine the impact of different design variables across a range of CAI program types. The results of these studies indicate that properly designed CAI can be effective as an instructional medium. For example, the Vocabulary Instruction study demonstrated that a skill requiring considerable practice can be adequately taught on a computer, and the Reasoning Skills program was successful at teaching a more complicated academic task—logical inferences. The Math Word Problems study, however, appeared to indicate that the best way to teach skills in this area may be through teacher directed instruction first, with the computer used for guided practice. The specific outcomes of the studies in this project lead to a better understanding of the application of technology in special education as it is linked to better instructional design principles. Another implicit outcome of this project is the need for careful consideration of the academic task, the stage of instruction, and the role of the teacher in order to make optimal use of computer assisted instruction.

Abstract

Applications of computer technology in special education have all too often come about because of enthusiasm over the hardware and selected software programs. While most advocates in the field are adept at detailing technical capabilities of this medium, little has been done in the way of systematic empirical research into the use of computer assisted instruction (CAI) for the mildly handicapped. This report summarizes eight studies conducted in this area over the last three years.

Each study is grounded in the instructional design principles articulated by the senior project director on this grant (Engelmann & Carnine, 1982) and others in the field. These principles have been empirically demonstrated as effective techniques in non-computer studies and in fifteen years of research on effective curriculum design variables as part of Project Follow Through. The studies described in this involved the three different kinds of CAI (i.e., drill and practice, tutorials, and simulations) as well as videodisc instruction. Three of the eight studies are refinements of the CAI studies conducted during the first year of this grant. In each case, we attempted to build upon the earlier study, either by investigating more specific hypotheses or by testing improved versions of the software program.

As a program of research, these studies support our initial hyposthesis, namely that empirically-based instructional design principles are applicable for improving the effectivness different types of CAI programs and the videodisc media. Furthermore, our research enabled us to begin to make some recommendations about teacher practices in relation to the use of educational technology.

Research Problem

The instructional design principles investigated under this grant have been articulated by Engelmann and Carnine's (1982) Theory of Instruction. These principles have been empirically demonstrated as effective techniques in non-computer studies (e.g., Carnine, 1980; Carnine, Kameenui, & Woolfson, 1982; Darch, Carnine, & Gersten, 1984) and in our fifteen years of experience with Project Follow Through (cf. Stebbins et al., 1977).

Through videodisc and computer assisted instruction, we have been able to isolate the effects of the following instructional design variables: a) review cycles, b) size of teaching sets, c) explicit strategies, and d) correction procedures. We did this in a variety of ways. Two of the studies compared software that we developed to popular commercial CAI programs. In two other studies, we compared variants the same software program, either by using different approaches or by using different versions of the same program. Two additional studies compared the use of technology (i.e., CAI and videodisc) to traditional teaching practices. Figure 1 shows the relationship of the different instructional design variables to the eight studies.

Figure 1

Instructional Design Variables

Studies	Example Set Size	Example Selection	Review Cycles	Correction Procedures	Explicit Strategies
Vocabulary Instruction	X		X		
Math Word Problems				X	X
Reasoning Skills I				X	
Reasoning Skills II				X	X
Health Problem Solving I					X
Health Problem Solving II					X
Videodisc Instruction		X		X	X

Research Methods

All of the studies described below were conducted with mildly handicapped secondary students, either at the junior high or secondary level. All were selected from special education resource programs and screened for skill deficits in the targeted academic area. For example, all students in the math word problems study were competent in basic arithmetic operations (through division) and knew how to solve addition and subtraction word problems. Thus, it was appropriate to teach these studetns word problems involving division and multiplication on the fucus of the study. Students whose skills were above or below this level were not used in the study; those who remained were randomly assigned to treatment conditions. Each study involved a group design, with random assignment of subjects to conditions. Finally, in order to precisely measure academic development, tests were created for the particular academic skills taught in each study. When appropriate, measures of knowledge transfer were also included.

Findings

The research findings have been or will soon be published in a variety of special education and technology journals. As each article fully documents the procedures used as well as the specific findings, we have appended a copy of each to this report. Except for the last two studies, which were completed in June of 1987 and thus remain unpublished at the time of this report, all six studies are described in the following six articles. We have also included abstracts of those two unpublished studies. For ease of reference, the table below indicates which studies are associated with the following articles.

Article or Abstract	Study (or Studies)
Woodward, J., Carnine, D., Gersten, R., Gleason, M., Johnson, G., & Collins, M. (1986). Instructional design principles for CAI: A summary of four studies. Journal of Special Education Technology, 7 (1), 107-118.	Vocabulary Instruction Reasoning Skills I Math Problems Health Problem Solving I
Johnson, G., Gersten, R., & Carnine, D. (1987). Effects of instructional design variables on vocabulary acquisition of LD students: A study of computer assisted instruction. Journal of Learning Disabilities, 20(4), 206-213.	Vocabulary Instruction
Collins, M., Carnine, D. W., & Gersten, R. (in press). Elaborated corrected feedback and the acquisition of reasoning skills. A study of computer-assisted instruction. Exceptional Children.	Reasoning Skills I
Woodward, J., Carnine, D., & Gersten, R., (in press). Teaching problem-solving through computer simulations. American Educational Research Journal.	Health Problem Solving I
Kelly, B., Carnine, D. W., Gersten, R., & Grossen, B., (1986). The effectiveness of videodisc instruction in teaching fractions to learning handicapped & remedial high school students. Journal of Special Education Technology, 8(2), 5-17.	Videodisc Instruction
Woodward, J., Carnine, D., & Collins, M. Closing the performance gap in secondary education. Article submitted to the Journal of Computer Research in Education	Vocabulary Instruction Reasoning Skills II Health Problem Solving I
The effect of an explicit strategy on student comprehension of problem solving skills taught in a computer health simulation. University of Oregon: abstract	Health Problem Solving II

[Abstract indicated here was not included in document received by ERIC.]

Abstract

Health Problem Solving II

An experimental study was conducted to determine the effectiveness of teaching an explicit strategy to a group of special education students for solving a series of health problems using computer simulations. Students were matched according to reading ability and randomly assigned to either an experimental or control group. They were given a pretest to determine their knowledge of basic health concepts.

During an intervention period of fourteen days, students in both groups studied health using a traditional form of health curricula and also a compute health simulation. The only difference between groups was the experimental students were taught a specific strategy to apply to the computer simulation. The control group received no such strategy.

Detailed study was completed on student performance across six test simulations. The defined objective on the simulation was to make decisions that would help surpass the profile's given life expectancy.

After the treatment period, four separate measures were given. The first test measured the students' knowledge of basic concepts taught. The other three measures examined students' ability to generalize knowledge gained in the simulations to new contexts. These included a series of videotaped health profiles and written profiles.

Both groups scored significantly higher on the health concepts posttest than on the pretest. While playing the test simulations, the treatment group scored significantly

higher than the control in achieving the defined objective. This difference appeared to be due to the strategy that the experimental students were taught.

No significant differences were found on the three generalizations measures, with one exception. A significant difference was found on one of the videotape measures, indicating experimental S's were more likely to correctly identify health changes required.

On the written generalization measure, students were expected to identify and prioritize health changes. Although able to identify the required and corresponding behavioral changes, they lacked skills needed to prioritize these changes.

The results of an attitudinal survey indicate students generally felt they had developed a strategy. They also strongly indicated they enjoyed studying health using the computer and the simulation; few wanted to continue studying health using traditional texts.

Researchers: Hollingsworth, M., Woodward, J., & Gersten, R.

Applying Instructional Design Principles to CAI for Mildly Handicapped Students: Four Recently Conducted Studies

John Woodward
Doug Carnine
Russell Gersten
Mary Gleason
Gary Johnson
Maria Collins

Journal of Special Education Technology, 7(1), 1986

Research for these studies was sponsored under the Department of Education grant number G008400600. We would also like to thank IBM for loaning us the computer hardware that enabled us to conduct some of these studies.

Special education has passed through a phase where computers have been widely embraced and uncritically adopted. The enthusiasm over computers and their potential impact on special education can be documented with little difficulty (e.g., Budoff, Thormann, & Gras, 1984; Blaschke, 1985). While most advocates are adept at detailing the technical capabilities of this medium (e.g., immediate feedback, automatic scoring, individualized instruction), little has been done in the way of systematic research into the use of computers - in particular, computer assisted instruction (CAI) - for the mildly handicapped. This report summarizes four studies recently conducted in this area. They are the beginnings of what we consider to be systematic research into CAI for the mildly handicapped.

The limited research on the instructional effectiveness of CAI for handicapped and non-handicapped populations is complicated and often contradictory. After a comprehensive search of the literature, Forman (1982) concluded that achievement was rarely enhanced by CAI, even though students exhibited positive attitudes toward such instruction. When studies do show effects on learning (Edwards, Morth, Taylor, Weis, & Dusseldorp, 1975; Burns & Boseman, 1981), they are modest and isolated, far from the more generalized impact on thinking skills that educators and enthusiasts have long claimed would result from CAI (Bangert-Drowns, Kulik, & Kulik, 1985). We are not surprised by these mixed findings, as little of the available software used in special education settings makes use of even the most rudimentary principles of sound instructional design and effective teaching (cf. Stevens & Rosenshine, 1981; Engelmann & Carnine, 1982; Brophy & Good, 1984).

In 1984, we began a series of CAI studies that examined different instructional design principles that have been articulated by Engelmann and

Carnine (1982) and others. These principles have been empirically demonstrated as effective techniques in non-computer studies (e.g., Carnine, 1980; Carnine Kameenui, & Woolfson, 1982; Darch, Carnine, & Gersten, 1984) and in our fifteen years of experience with Project Follow Through (cf. Stebbins et al., 1977). The studies described below involved the three different kinds of traditional CAI: drill and practice, tutorials, and simulations. Thus, a secondary interest of this research was to examine the impact of different instructional design variables across a range of CAI program types.

Through computer assisted instruction, we have been able to isolate the effects review cycles, size of teaching sets, explicit strategies, and correction procedures. We were able to do this in a variety of ways. Two of our studies compared software that we developed to popular commercial programs. In another, we examined the effect of one variable (a correction procedure) by modifying our version of the software. In the last study described in this report, we used our software as an adjunct to a written curriculum to teach specific problem solving skills. Figure 1 shows the relationship of the different instructional design variables to the four studies.

[Insert Figure 1 about here]

All of the studies described below were conducted with mildly handicapped secondary students. The two tutorial studies involved students from different junior high schools in a medium sized district. In each of the two remaining studies, the participating students were all from different class periods at the same high school. Students were mostly while and from middle and lower middle class families. All were selected from special education resource programs and screened for appropriate skill deficits before it was determined if practice programs for teaching vocabulary to mildly handicapped adolescents (Johnson, Gersten, & Carnine, 1986). The study examined the effect of size of the daily teaching sets and provisions for daily and cumulative review on the acquisition and maintenance of word meaning. Two CAI vocabulary programs were used to present the same 50 words and definitions.

Two designs were used in this study: 1) a time to mastery (Will there be a significant difference between times required to meet mastery criterion on the 50 words by students taught with the two different CAI programs?), and 2) fixed design, in which all subjects were tested after the seventh session. We also looked at differences between pretest and posttest scores as well as maintenance of effects two weeks after students achieved mastery.

<u>Method</u>

Twenty-four mildly handicapped high school students from an initial pool of 38 students were matched by scores on a 50 item vocabulary pretest and randomly assigned to one of the two CAI programs. The students were all referred to the school's special education program for remedial reading or language arts instruction and were identified by the resource teacher as needing vocabulary instruction. Students worked individually on th · ·ssigned program 20 minutes a day for 11 days. All of the words, which were the same for both programs, were considered important by two or more special education teachers. A final list composed of 25 verbs and 25 adjectives was used.

<u>The CAI Programs.</u> One program used in the study, the Small Teaching Set program, tests students on words and then creates lessons with the words they cannot identify (Carnine, Rankin, & Granzin, 1984). After testing the students on new words, the program provides instruction on a "teaching set" of no more than three words which the student miss·d on the test. Each lesson

also includes a "practice set" with a maximum of seven words. The student must meet a specific mastery criterion on each word before it is removed from the practice set. The program tests the student on new words and adds words the student does not know to the practice set. Once the student has mastered ten words, the program presents a cumulative review lesson on those words.

The other program, the Large Teaching Set program, teaches words in sets of 25 words (Davidson & Eckert, 1983). The student may choose to see the words in any of four types of formats: (a) a teaching display which shows the word, its definition, and one example sentence; (b) a multiple choice quiz format; (c) an exercise in which a definition is displayed and the student must spell the correct missing word to complete a sentence; and (d) an arcade-type game in which the student matches words to their definitions.

Measures. A 50 item, multiple choice test was developed for the study (.79 coefficient alpha). This test was administered to all subjects as a pretest, as a criterion reference test at the end of seven sessions, immediately after mastery (or at the end of the eleventh session), and two weeks after mastery. There were also two transfer measures. One was a 10 item objective test in which students defined words orally. The other test required students to answer comprehension questions that require knowing the meaning of words in several short passages.

Results

Eight of the twelve subjects (67%) in the Large Teaching Set program and ten of the twelve subjects (83%) in the Small Teaching Set program met mastery criterion by the end of 11 sessions. The study was terminated after eleven sessions because the experimenter felt that the subjects who were still struggling to reach mastery were no longer benefiting from instruction. The mean number of sessions to mastery (for those who reached mastery) was 7.6 for those in the Small Teaching Set and 9.1 in the Large Teaching Set program. Table 1 presents the descriptive statistics for both groups. Results of a t - test indicate this difference is significant ($p < .05$). Hence, subjects in the Small Teaching Set program met mastery in significantly less time.

[Insert Table 1 about here]

A 2 x 2 analysis of variance (ANOVA) was performed on posttest and maintenance test scores, indicating no significant main effect for type of instruction. Results on the multiple choice test in the fixed time design (i.e., the test administered to all students after seven sessions) indicates a slight, but nonsignificant difference in means favoring subjects in the Large Teaching Set program. Differences between scores on two transfer measures were also statistically nonsignificant.

Discussion

The unequivocal finding of the study was that the subjects taught with the Small Teaching Set program reached mastery criterion on the set of 50 words faster than subjects with the Large Teaching Set program. In addition, more students in the Small Teaching Set program reached mastery within eleven lessons. Given that the groups achieved equivalent levels of performance on the multiple-choice tests, their difference in acquisition rates becomes even more meaningful. Subjects taught with the Small Teaching Set program required less time to meet mastery criterion on the words, yet their posttest performance was equal to that of subjects in the other treatment who took longer reaching mastery. In addition, the shorter instructional time which the Smaller Teaching Set program subjects required did not negatively affect their retention of word meanings.

Tutorials

Reasoning Skills: Correction Procedures

Much of the recent literature on improving special education teaching practices has stressed the importance of providing academic feedback to students when they made errors (Carnine, 1980; Rieth, Polsgrove & Semmel, 1981; Stevens & Rosenshine, 1981). Furthermore, a recent meta-analysis of the limited research on corrective feedback by Lysakowski and Walberg (1981) suggests that detailed corrective feedback is superior to merely telling students whether their answers are right or wrong. Just telling students they are wrong (called a "basic correction") does not help them solve the problem correctly. These authors suggest that students need to see an overt model of all the steps necessary for an appropriate response. By observing a model of all the steps necessary in obtaining a correct response, students receive detailed information on how to solve the problem. This procedural knowledge should be of use when they encounter similar types of problems. This type of correction is referred to as an "elaborated correction."

This was the first of two computer tutorial studies, and the primary intent here was to examine whether remedial and mildly handicapped students who receive elaborated correction procedures would perform significantly better than students provided with basic corrections (Collins, Carnine, & Gersten, in press). We also examined any differences regarding acquisition time between students. Reasoning skills were chosen as a subject because they tend not to be routinely taught to special education students (Zetlin & Bilsky, 1980). Furthermore, the strategy used to teach these skills (i.e., basic analysis of three statement arguments) was highly rule-governed, thus it is an appropriate for an elaborated correction procedure.

Method

Twenty-eight mildly handicapped and remedial junior high school students from a pool of 34 students were selected and randomly assigned to the Basic Correction or Elaborated Correction group. Participating subjects had at least a fifth grade reading level (but also had a reading comprehension deficiency of no more than three years) and passed a screening test that measured their understanding of large and small classes. The Elaborated Correction group used an unaltered copy of the CAI program used in the study. The Basic Correction group used a modified version of the program. If a student in this group made an error, they were only given the correct answer. This was the only difference between the two conditions. In both conditions, students worked individually on a microcomputer. Students worked on their respective version of the program until they completed five lessons.

The CAI Program. The Reasoning Skills program (Engelmann, Carnine, & Collins, 1983) was designed to teach students two major objectives: a) to draw conclusions from two statements of evidence and b) to determine whether a three-statement argument was logical or illogical. The program taught students about overlapping classes and non-overlapping classes. They learned that there are three possible key words (some, all, no); the same rule holds for all three. It also taught students relevant rules for constructing and analyzing arguments. The other major objective of the program was to teach students to identify unsound arguments. For logically unsound arguments, students were taught to specify one of three reasons why an argument was unsound.

Measures. The Test of Formal Logic (Collins, 1984) was the primary dependent measure in the study. The purpose of this test was to measure a student's ability to construct and analyze syllogistic arguments. Two alternative forms of the test were designed; Form A was used as the pretest and maintenance measure (given two weeks after treatment terminated) and Form B was used as the the posttest measure (given immediately after the treatment). The internal consistency reliability (coefficient alpha) for Form A was .90 and .91 for Form B. Parallel form reliability between Forms A and B was .84.

There was also a 15 item transfer test that evaluated subjects' abilities to generalize what they had learned on the computer to similar analytic tasks, but in prose paragraph form. The transfer test was devoted to the more difficult objective on the program - deciding whether arguments were sound, and, if not sound, giving a reason. This test was given to subjects on the day after they completed training on the CAI program.

Results

A 2 x 3 analysis of variance (ANOVA) with one between subjects factor (Type of Correction) and one within subjects factor (Time of Test) was performed on the data. This analysis involved a planned comparison that looked at the post and maintenance tests only. Table 2 presents the descriptive statistics for the pretest, posttest, and maintenance tests. The ANOVA indicated a significant difference favoring the Elaborated Corrections group ($p < .001$) There was also a significant difference between the two groups on the transfer test, again favoring the Elaborated Correction group ($p < .05$).

[Insert Table 2 about here]

Data were collected on the time students took to complete each of the five lessons. The purpose of this analysis was to see whether students in the Elaborated Corrections group took more time to complete the lessons. A 2 x 5 analysis of variance (ANOVA) with repeated measures was performed on the time-per-lesson data and non-significant difference between groups was found.

Discussion

This study was the first to explore experimentally the effectiveness of elaborated corrective feedback in teaching a complex cognitive skill to handicapped learners. The results indicate this is an effective instructional procedure.

The roughly equivalent time for the two groups to complete the five lessons seems anomalous at first. With more text to read in elaborated corrections, that treatment would seemingly take longer to complete the lessons. Completion times were not significantly greater for the elaborated corrections group, however. The extra time required to read the elaborated corrections may have been compensated for by faster acquisition of the material. In both versions of the program, the computer would return a student to items that were missed earlier in a lesson. If elaborated corrections resulted in fewer mistakes, students would spend less time returning to missed items. This interpretation suggests that taking more time early in a complex instructional sequence to offer elaborated corrections may, in fact, lead to savings in instructional time later in a program.

Both the basic and elaborated correction groups improved their reasoning skills as measured by the dependent variable. The groups demonstrated a mean score of 68 - 70% on the posttest (a dramatic gain from the mean scores of 26 to 34 percent on the pretest). The systematic design of instruction - particularly through a series of carefully controlled rules - may have contributed to this gain. Reasoning skills were acquired without any instruction from the teacher. Typically, CAI programs merely provide drill and practice exercises to supplement teacher instruction. Here the program was a true tutorial and did all the instructing.

Math Word Problems: Explicit Strategies

In our second study of tutorials, we examined the effectiveness of this kind of CAI program in teaching math word problems. Unlike the basic analysis of a three statement argument, solving word problems is a much more complex skill. Students are required to make many discriminations and success depends, to a large degree, on linguistic analysis (Jerman & Mirman, 1974). Although a large number of studies have been conducted in math problem solving, few have yielded any adequate information for building effective interventions because of flaws in research design (Kilpatrick, 1978; Silbert, Carnine, & Stein, 1981), and varying definitions of problem solving and the tasks to measure problem-solving ability (Silver & Thompson, 1984). Furthermore, the success of future problem solving research depends on less on a continued analysis of the learner and his or her deficiencies and more on 1) an analysis of the limits of instruction the students are currently receiving and 2) development of strategies that will work with low achieving students.

The specific purpose of the study was to determine whether handicapped students could learn to solve multiplication and division math story problems if taught a strategy that focused on how to choose the correct operation (Gleason, 1985). It was hypothesized that students who received explicit instruction on choosing the operation would solve more problems correctly than students who received instruction that did not specifically focus on the choice of operation and concentrated, instead, on a more general strategy of manipulating units. A wider interest of this study, one directly related to our systematic research agenda, was in whether or not a CAI tutorial alone would be effective in teaching a complex cognitive skill.

Method

A pretest-posttest design with random assignment of subjects to treatment groups was used to examine the effectiveness of two procedures for teaching mildly handicapped students to solve math word problems. Twenty-six junior high school students from two different schools were randomly assigned to either a "direct instruction" math story problems program or a highly regarded commercial program. Each student worked at a computer for 15 to 30 minutes a day for 11 days.

The CAI Programs. The direct instruction program Analyzing Word Problems (Carnine, Hall, & Hall, 1983) is based on principles of a theory of instruction described by Engelmann and Carnine (1982). The approach requires direct teaching of a clearly-specified, step-by-step strategy. When instructing the students, the teacher models each step in the process, heavily prompting the students as they continue to use the process. The prompts are systematically faded until students reach independence. When students make errors, the teacher again models or provides a prompt based on a previously taught rule. This kind of strategy instruction is incorporated into the Analyzing Word Problems program. The program teaches students how to solve multiplication and division word problems in a step-by-step fashion. When students err, they are given a rule-based correction (e.g., "Does the statement contain the word Each or Every? If so, what kind of problem is it?").

The Semantic Calculator (Sunburst Communications, 1983) was used as a contrast to the DI program. This program is based on the premise that the major difficulty in solving word problems comes from inappropriate manipulation of units (e.g., weeks, apples, dollars, etc.). If a student could be taught to extract from the problem the quantities needed to solve the problem and the correct units for the answer, they would be able to solve the problem. In this program, a student is guided through story problems by answering "How many?" and "What?" questions about word problems that are written on worksheets. Next, the student uses the letters A and B to type in the operation that should be performed on the numbers (i.e., "A/B" to divide and "A x B" to multiply) and then predict the units in the answer. The computer shows the student what units were used to express the answer to the problem. If the student answer did not match that of the computer's, the student knew that he or she should go back and try again.

Measures. Both the pretest and posttest were 28-item tests comprised of 11 multiplication, 10 division, 2 addition, and 5 subtraction problems. All items were selected from three major arithmetic intermediate level textbooks and from the California Achievement Test. Sixty-eight percent of the problems on the test were like ones included in the instructional lessons; the remaining 32% were transfer problems.

Results

Results indicated no significant differences between performance of the Direct Instruction group and the Semantic Calculator group on the posttest and in the amount of time used to take the posttest. Interviews with students as they performed problems (i.e., choosing the correct operation and telling a reason for the choice) did yield a statistically significant difference between the groups favoring the Direct Instruction group, but the mean performance levels for both were not considered educationally significant. Table 3 shows the descriptive statistics for the two groups on the pre- and posttests.

[Insert Table 3 about here]

Discussion

There are many possible reasons for why there were no significant differences between groups. Eleven days at 25 minutes a day may have been too short of an intervention. With a longer treatment period, it would have been more certain that an unacceptable level of performance was attributable to other factors. Further, observations of student performance during the study indicated that many students typically ignored prompts on the screen that told them what to do next. Hence, through a failure to attend the students may well have missed opportunities to learn from their errors.

It is also conceivable that mildly handicapped students may need more teacher-directed instruction before using a computer for additional practice opportunities. The presentation of the problem-solving strategy on the computer lacked the subtlety and flexibility that a teacher adds to instruction. Good teachers gather a considerable amount of information about how students are learning a new skill - particularly one as difficult as problem solving - and modify their teaching accordingly. Observations of teacher directed instruction that uses the same strategy that is presented in the Analyzing Word Problems program support this hypothesis (Gleason, 1985). Therefore, the most appropriate use of a computer for students such as these may be for guided practice (i.e., as a medium for reviewing material that is already familiar to the students).

Simulations

Health Ways: Problem Solving Skills

Secondary students spend a considerable amount of their time completing seatwork activities (Doyle, 1983). These academic tasks often involve higher order cognitive skills, and students are asked to make a variety of inferences about a subject area by prudently using facts, concepts, and strategies or problem solving skills. Some writers (Doob, 1972; Greenblat & Duke, 1975; Budoff, Thormann, & Gras, 1984) have suggested that one way to enhance the higher order skills of students is through educational simulations.

While much of the research has concluded that computer and non-computer simulations are no more effective than conventional instruction, many of these studies have been plagued by fundamental weaknesses in research design (Fletcher, 1971; Pierfy, 1977). In the study below, we addressed many of these problems and created an instrument that reflected the problem solving skills actually taught in the computer simulation. Finally, we have addressed a curious feature of previous simulation research: the general reluctance to combine simulation instruction with conventional instruction.

Only on a few occasions have simulations and conventional instruction been compared to conventional instruction alone. Nor is it clear in most of these studies what constitutes conventional instruction. O[illegible] of our interests in studying simulations was to investigate how effective instructional practices could be used to enhance - rather than replace - secondary level instruction, not only in terms of their effect on basic fact and concept retention, but as they relate to higher order skills (Woodward, Carnine, & Gersten, 1986).

Method

Thirty mildly handicapped high school students from a pool of 38 potential students were randomly assigned to either the conventional or simulation condition. Students were selected for the study on the basis of Metropolitan Achievement Test (MAT) reading comprehension scores. Only those who had, at a minimum, a sixth grade reading level or were at least two years below grade level in comprehension were selected for the study.

All students were instructed for 40 minutes per day for twelve days. The lesson consisted of two parts. The first part, called <u>structured teaching</u>, was identical for subjects in both student conditions. Instruction was conducted in a large group of 12 to 15 students for this part of each lesson.

At the end of the initial instruction, students separated into two groups - one which worked on application activities (the conventional group) and the other with the computer simulation (the simulation group). The conventional group worked in the resource room under the supervision of the resource room teacher, who presented these students with a variety of application or review activities.

Simulation students, on the other hand, were taught in a computer lab, each student working individually at a microcomputer. The twelve day course of instruction for these students was broken into three phases: initial modeling (three days), guided practice on three simulation games (two days), and independent practice with individual feedback from the instructor (seven days).

<u>The CAI Program.</u> <u>Health Ways</u> is a commercial software program developed by Carnine, Lang, and Wong for the Apple II and IBM PCjr computers. The simulation provides extensive practice on analyzing health profiles. An initial screen showing profile characteristics can be seen in Figure 2. The <u>Health Ways Supplementary Curriculum</u>, an accompanying written curriculum developed by Woodward and Gurney, extended information presented in <u>Health Ways</u> and the original <u>Health Ways</u> teachers guide. Material was taken directly from two widely used junior high school health textbooks. All of the information was rewritten to control for vocabulary and amount of new information. Clippings from newspapers, news magazines, journal articles, and health pamphlets were also used in this supplementary

curriculum. The reading 'evel of the curriculum is approximately sixth grade.

[Insert Figure 2 about here]

Measures. Students were assessed one day, two days, and two weeks following the instruction. On the first day, student's acquisition of basic facts and concepts about health taught in the curriculum was measured by the Health Ways Nutrition and Disease Test. The first 20 questions of this test were solely from the written curriculum. The remaining 10 were questions over material that appears in both the written curriculum and the Health Ways simulation. This latter section will be referred to as the reinforced section of the test because information pertaining to the items was reviewed in the computer simulation. Internal consistency reliability (coefficient alpha) of this measure is .84. On the second day, the students were given the Health Ways Diagnosis Test, an individually administered test measuring prioritizing skills. This test was a set of three written profiles and measured health related problem solving skills (i.e., the student's ability to detect important health problems facing an individual, identify and change related health habits, and control stress as it increased due to the health changes). The Health Ways Diagnosis Test has a test - retest reliability of .81. Two weeks after the instruction the students were again given the Health Ways Nutrition and Disease Test. This served as a retention measure.

Results

Health Ways Nutrition and Disease Test. This test was used in both the posttest and maintenance (retention) phases, a 2 x 2 analysis of variance (ANOVA) was used. The 30 items in the test were broken into two subscales: a) items reinforced by the Health Ways simulation and b) items taught in the curriculum and not reinforced by the simulation. Separate 2 x 2 ANOVAs with repeated measures were performed on each subscale. The effect on items reviewed or reinforced by Health Ways was significant ($p < .01$) and nonsignificant for those items not reinforced ($p < .06$). Table 4 presents the descriptive statistics for reinforced section of the test. This indicates that the simulation was an effective vehicle for reviewing material that had already been taught in the written curriculum.

[Insert Table 4 about here]

Health Ways Diagnosis Test. t-tests performed on this test demonstrate a significant difference between the two groups ($p < .001$) in problem solving skills. Simulation students were better able to diagnose health problems, prioritize them as to their effects on an individual's longevity, and prescribe appropriate remedies. Although most simulation students did not receive perfect scores on this measure (i.e., they followed the strategy presented in the treatment exactly), performance was almost always above the criteria set by the experimenters for appropriate diagnosis and the suggested sequence of remedies.

Discussion

The results of this study support the use of computer simulations in teaching material not easily taught by traditional means. Further, a structured approach in simulations, one where outcomes are specified and controlled, does produce significant educational results.

We infer from the results that the explicit strategy instruction used to teach the simulation students about Health Ways was a successful bridge to "less" direct instruction activities. It is not clear that computer simulations, by themselves, are adequate in teaching higher order skills. For this reason, one might speculate that added instruction is necessary; at least to the point that it focuses the student on the instructional goals of the simulation.

Conclusion

The results of these four studies indicate that properly designed CAI can be effective as an instructional medium. These findings are consistent with our non-computer research that we have conducted over the last ten years. Sophisticated design principles can make a considerable difference in the effectiveness of any instructional program. Another outcome implicit in these studies is that they begin to identify - with much greater clarity - the role of the teacher and his/her instruction away from the computer. This perspective deviates from original questions about computer assisted instruction (e.g., Is CAI more effective or efficient than conventional instruction?). It forces us to look closely at the intersection of the teacher, the academic task, and the stage of instruction (e.g., Is the skill just being introduced, practiced, or reviewed?).

The Vocabulary Instruction study, for example, demonstrates that a skill requiring considerable practice can be adequately taught on a computer. Such a task is time consuming for a teacher and can be handled effectively by CAI. Furthermore, there is very little variation as instruction moves from one stage of instruction to the next (i.e., from introduction and modeling to guided practice to independent practice). Note, however, that the task, as it was defined in the study, was one of memorizing vocabulary words. We did not teach nor assume that students would necessarily learn how to use the words expressively or detect their meaning from context. This would have required a different analysis.

The Reasoning Skills program, a teacher independent tutorial, was successful at teaching a more complicated academic task: logical inferences. It might be argued that the particular academic skills taught in this program were more discrete than, say, math word problems or the subtle problem solving skills addressed in the Health Ways study. If this observation is correct, then computer assisted instruction - carefully developed with instructional design principles and field tested - can be an effective, "stand alone" form of instruction. As with the vocabulary program, such instruction is efficient insofar as it allows the teacher to attend to other instructional activitie. while students work on the computer.

Math word problems, on the other hand, do no share the same task complexity as the syllogisms. In this study, we speculate the best way to teach skills such as math word problems, notoriously difficult for mildly handicapped students, may be through teacher directed instruction first, with the computer used for guided practice. This conclusion is admittedly tentative, and we will conduct another study after we revise our strategy.

Finally, the Health Ways study strongly suggests that facts and concepts, which were preskills to the problem solving activities, can be efficiently taught in group instruction without computers. The computer can be an effective tool after the preskills have been introduced and explicit strategies for using the simulation have been taught by the teacher. In this sense, a complex task like problem solving can be effectively taught in the guided and independent practice phases of instruction.

As a program of research, these studies demonstrate that empirically-based instructional design principles are applicable across different types of CAI programs. Although the research focuses on these principles, we can extract some recommendations about teacher practices. With effective software, the teacher can use CAI programs for the purpose of efficiency (i.e., letting the computer teach low level or time consuming tasks) or to present material that would be difficult to do through traditional media. The latter was certainly the case with the Health Ways simulation. Looking beyond the specific outcomes of this research, an optimal use of computer assisted instruction requires more than a set of design principles. It entails a careful consideration of the academic task, the stage of instruction, and the role of the teacher.

References

Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), Comprehension and teaching: Research reviews. Newark, DE: International Reading Association.

Bangert-Drowns, Kulik, J. A., & Kulik, C. L. C. (1983). Effects of coaching on achievement test performance. Review of Educational Research, 53, 571-85.

Beck, I.L., Perfetti, C.A., & McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. Journal of Educational Psychology, 74, 506-521.

Blaschke, C. (1985). Technology trends in special education. T.H.E. Journal, 12 (6), 73-77.

Brophy, J., & Good, T. (1984). Teacher behavior and student achievement. In M. Wittrock (Ed.) Third handbook of research on teaching (in press).

Budoff, M., Thormann, J., & Gras, A. (1984). Microcomputers in special education. Cambridge, MA: Brookline Books.

Burns, P. K., & Boseman, W. C. (1981). Computer-assisted instruction and mathematics achievement: Is there a relationship? Educational Technology, 21(10), 32-39.

Carnine, D. (1980). Phonics versus whole-word correction procedures following phonic instruction. Education and Treatment of Children, 3 (4), 323-330.

Carnine, D., Hall, T., & Hall, R. (1983). Analyzing word problems. Unpublished computer program, Eugene, Oregon.

Carnine, D., Kameenui, E., & Woolfson, N. (1982). Training of textual dimensions related to text-based inferences. Journal of Reading Behavior, 14 (3), 335-340.

Carnine, D., Lang, D., & Wong, L. (1984). Health ways. Unpublished computer program, Eugene, Oregon.

Carnine, D., Rankin, G., & Granzin, A. (1984). Learning vocabulary. Unpublished computer program, Eugene, Oregon.

Collins, M. (1984). Test of formal logic. Unpublished doctoral dissertation, University of Oregon.

Collins, M., Carnine, D., & Gersten, R. (In press). Elaborated corrected feedback and the acquisition of reasoning skills. A study of computer assisted instruction. Exceptional Children.

Darch, C., Carnine, D., & Gersten, R. (1984). Explicit instruction in mathematics problem solving. Journal of Educational Research, 77, 351-359.

Davidson, J.G., & Eckert, R.K. (1983). Word attack! [computer program]. Ranchos Palos Verdes, CA: Davidson & Associates.

Doob, P. (1972). Prospects for simulation gaming in health planning and consumer health education. ERIC ED 070 596.

Doyle, W. (1983). Academic tasks. Review of Educational Research, 53, 159-199.

Durkin, D. (1979). What classroom observations reveal about reading comprehension instruction. Reading Reserach Quarterly, 14, 481-533.

Edwards, J., Norton, S., Taylor, S., Weiss, M., & Dusseldorp, R. (1975). How effective is CAI? A review of the research. Educational Leadership, 33, 147-153.

Engelmann, S., & Carnine, D. (1982). Theory of instruction. New York: Irvington.

Engelmann, S., Carnine, D., & Collins, M. (1983). DIAL reasoning skills program [computer program]. Eugene, Oregon: Engelmann - Becker Corp.

Fletcher, J. (1971). The effectiveness of simulation games as learning environments. Simulation & Games, 2, 425-554.

Forman, D. (1982). Review of the literature. The Computing Teacher, 9 (5), 37-50.

Gleason, M. (1985). A comparison of two CAI approaches in teaching math word problems to handicapped learners. Unpublished doctoral dissertation, University of Oregon.

Greenblat, C., & Duke, R. (1975). Gaming simulation: Rationale, design, and application. New York: Halsted.

Jerman, M.E., & Mirman, S. (1974). Linguistic and computational variables in problem solving in elementary mathematics. Educational Studies in Mathematics, 4, 306-323.

Johnson, G., Gersten, R., & Carnine, D. (1986). Effects of instructional design variables on vocabulary acquisition of LD students: A study of computer assisted instruction. Manuscript submitted for publication.

Kilpatrick, J. (1978). Research on problem solving in mathematics. School Science and Mathematics, 78, 189-192.

Lysakowski, R., & Walberg, H. (1982). Instructional effects of causes, participation and corrective feedback: A quantitative synthesis. American Educational Research Journal, 19 (4), 559-578.

Pearson, P.D., & Gallagher, M. (1983). The instruction of reading comprehension. Contemporary Educational Psychology, 8, 317-344.

Pierfy, D. (1977). Comparative simulation game research: Stumbling blocks and stepping stones. Simulation & Games, 8, 255-268.

Reith, H., Polsgrove, L., & Semmel, M. (1981). Instrucitonal variables that make a difference: Attention to task and beyond. Exceptional Education Quarterly, 2(3), 61-71.

Roser, N., & Juel, C. (1981, December). Effects of vocabulary instruction on reading comprehension. Paper presented at the annual meeting of the National Reading Conference, Dallas.

Semantic Calculator [computer software] (1983). Pleasantville, NY: Sunburst Communications.

Silbert, J., Carnine, D., & Stein, M. (1981). Direct instruction mathematics. Columbus, OH: Merrill.

Silver, E., & Thompson, A. (1984). Research perspectives on problem solving in elementary school mathematics. The Elementary School Journal, 84, 529-545.

Stebbins, L., St. Pierre, R., Proper, E., Anderston, R., & Cerva, T. (1977). Education as experimentation: A planned variation model, Vols. IV A-D. Cambridge, MA: Abt Associates, 1977.

Stevens, R., & Rosenshine, B. (1981). Advances in research on teaching. Exceptional Education Quarterly, 2, 1-9.

Tierney, R.J., & Cunningham, J.W. (1984). Research on teaching reading comprehension. In P.D. Pearson (Ed.), Handbook of reading research (pp. 609-656). New York: Longman.

Woodward, J., Carnine, D., & Gersten, R. (in press). Teaching problem solving through computer simulations. American Educational Research Journal.

Zeitlin, A. & Bilsky, L. (1980). Reasoning by trainable mentally retarded and young non-retarded individuals. Journal of Mental Deficiency Research, 24 (1), 65-71.

Table 1

Mean Score, Standard Deviation, and Mean Percent Correct on 50-item Pretest, Posttest, and Maintenance Test for Small Teaching Set and Large Teaching Set Samples: Time-to-Mastery Design

		Pretest			Posttest			Maintenance Test		
Group	N	M	SD	Mean Percent Correct	M	SD	Mean Percent Correct	M	SD	Mean Percent Correct
Small Teaching Set	12	24.7	8.1	49.4	42.0	4.0	84.0	40.5	7.1	81.0
Large Teaching Set	13[a]	24.9	7.8	49.8	43.7	7.7	87.4	42.0	8.7	84.0

[a]One subject completed only seven sessions. This subject's posttest scores appear only in the fixed-time design.

Table 2

Mean Score, Standard Deviation, and Mean Percent Correct on Test of Formal Logic: Pretest, Posttest, and Maintenance Test for the Elaborated and Basic Correction Groups

		Pretest			Posttest			Maintenance Test		
Group	N	M	SD	Mean Percent Correct	M	SD	Mean Percent Correct	M	SD	Mean Percent Correct
Elaborated Corrections	17	10.8	4.3	30.7	24.6	5.0	70.4	25.1	3.7	71.6
Basic Corrections	17	11.5	4.2	32.9	21.5	6.6	61.1	20.6	7.5	59.0

Table 3

Mean Score, Standard Deviation, and Mean Percent Correct on the 28-item Story Problem Test: Pretest and Posttest for the Analyzing Word Problems and Semantic Calculator Groups

		Pretest			Posttest		
Group	N	M	SD	Mean Percent Correct	M	SD	Mean Percent Correct
Analyzing Word Problems	13	13.5	3.7	48.0	14.8	5.5	53.0
Semantic Calculator	13	13.9	3.5	50.0	11.6	6.3	41.0

Table 4

Mean Score, Standard Deviation, and Mean Percent Correct on the Reinforced Section of the Health Ways Nutrition and Disease Test: Posttest and Maintenance Test for the Conventional and Simulation Groups

		Posttest			Maintenance Test		
Group	N	M	SD	Mean Percent Correct	M	SD	Mean Percent Correct
Simulation	15	7.3	1.4	73.3	6.5	2.0	65.3
Conventional	15	5.6	2.2	56.0	5.0	2.2	50.0

Figure 1

	Instructional Design Variables			
	Example Set Size	Review Cycles	Correction Procedures	Explicit Strategies
Studies				
Vocabulary Instruction	X	X		
Math Word Problems			X	X
Reasoning Skills			X	
Health Problem Solving				X

Effects of Instructional Design Variables on Vocabulary Acquisition of LD Stude..ts: A Study of Computer-Assisted Instruction

Gary Johnson, Russell Gersten, and Douglas Carnine

Two computer-assisted instructional vocabulary programs were used to teach definitions of 50 words to 25 learning disabled high school students. These students were matched on pretest scores and then randomly assigned to one of two computer-assisted instructional programs. The major differences between the programs were (a) the size of the teaching sets and (b) the procedures for cumulative review. One program provided teaching and practice exercises on small sets of words and cumulative review exercises on words the students learned in the program. The other program provided exercises on a large set of 25 words at a time and no cumulative review. Students received computer-assisted vocabulary instruction for a maximum of eleven 20-minute sessions. The major finding was that significantly more students who were taught with the small teaching set program reached mastery within 11 sessions than students in the comparison group. Students in both programs learned as much (as measured by the criterion-referenced test) and retained as much, as measured by the maintenance test. Yet one group learned the material more efficiently. No significant differences were demonstrated on two transfer measures, an oral test of word meanings and a passage comprehension test.

Based on the premise that word knowledge correlates highly with reading comprehension (Anderson & Freebody, 1981; Mezynski, 1983; Pearson & Gallagher, 1983; Stahl, 1983; Tierney & Cunningham, 1984), several investigators have attempted to improve students' reading comprehension skills by teaching vocabulary. In only a few of these studies were subjects identified as either low performing or learning disabled.

While all studies produced evidence of improved vocabulary knowledge, the effects on reading comprehension have been varied and inconclusive. However, the methods which were most successful in teaching new vocabulary consumed the most instructional time, and even then the gains, in terms of numbers of words learned, were modest. An area in nee[illegible] research is the identification of [illegible]ods for increasing the direct instruction of vocabulary for low performing students. An important concern, particularly with mildly handicapped students, is increasing instructional time without increasing the demands on teachers' already limited time. One potential vehicle for delivering this instruction is computer-assisted vocabulary instruction.

INSTRUCTIONAL DESIGN FOR COMPUTER-ASSISTED INSTRUCTION

Features of computer-assisted instruction (CAI) which are seen as advantageous for instruction with special education students include individualization and self-pacing, immediate feedback about performance, consistent correction procedures, patient repetition, carefully sequenced instruction, frequent student responding, and motivation (Budoff, Thormann, & Gras, 1984). Yet, much currently existing software fails to provide these features in programs for special education students (Thormann, Gersten, Moore, & Morvant, in press). Two instructional design considerations that were explored in this study were (a) optimal set si[illegible] for introducing new words and (b) schedules of re[illegible] Relevant research on each varia[illegible] discussed below.

"Set Size" for Daily Lesson[illegible]

Drill and practice programs a[illegible] most widely used CAI software [illegible]doff et al., 1984), but little attentio[illegible] been directed to increasing th[illegible] ficiency and efficacy of CAI dril[illegible] prac[illegible]e strategies (Merrill & [illegible]bury, 1984; Siegel & Misselt, 1984[illegible] area of needed research is the [illegible]mination of the appropriate size [illegible] learning set for lessons in a dril[illegible] practice program. Concerns for [illegible]riculum designers are how many [illegible] can reasonably be presented d[illegible] each lesson and how the set size a[illegible] the learner's ability to master [illegible] remember material. This stud[illegible] dresses that issue by evaluatin[illegible] software programs with a ver[illegible] ferent approach towards introdu[illegible] set size.

The early behavioral studies ([illegible] & Hulse, 1967; Furukawa, 1970; [illegible] 1971; McGeoch & Irion, 1952) on [illegible]grammed instruction lend supp[illegible] Miller's (1956) "magical numbe[illegible] seven as the optimal set size (i.e[illegible] average number of unrelated c[illegible] letters, or words that an adult [illegible] recall after one exposure). Be[illegible] most of the research in this are[illegible] been conducted with college level [illegible]jects (cf. Deese & Hulse, 1967; [illegible] 1971; McGeoch & Irion, 1952), an[illegible] results of studies with school-age [illegible]dren were equivocal, further res[illegible] with school-age children wit[illegible] lea[illegible] disabilities is needed. One purp[illegible] the present study was to compar[illegible] effects of different set sizes on th[illegible] quisition of word meanings by m[illegible] handicapped students.

Review Schedule

The element of review was a[illegible] portant variable in the Beck, Pe[illegible] and McKeown (1982) study of v[illegible]ulary acquisition. Words were pre[illegible]ed in two different levels of frequ[illegible] In one treatment, each set of w[illegible] was taught in a 5-day cycle. A[illegible] words in the set were introduce[illegible] the first day of the cycle and reviewed when necessary, but only during that 5-day cycle. There was no subsequent review. In a second treatment words were introduced in the same manner, but additional practice was provided after the fifth day in special review cycles. Words in this treatment appeared *16 to 22 additional times each.*

The effects of the extra reviews were clear. Students remembered more of the words which received special reviews, and on a lexical decision test, students were able to access the definition of these words more quickly. A replication of the original study by McKeown, Beck, Omanson, and Perfetti (1983) again demonstrated that extra review enhanced students' ability to answer multiple choice questions about story passages. In the replication, the extra reviews positively affected both students' retention of word meanings and their comprehension of passages containing instructed words.

Research in computer-assisted instruction has demonstrated that several short-spaced reviews are more effective in increasing retention than are a few massed reviews (Gay, 1983; Siegel & Misselt, 1984).

Merrill and Salisbury (1984) propose a strategy that would provide spaced reviews during a CAI drill and practice program. New items are presented to students, and only items they do not know become part of a "working pool." The number of items in the working pool would be determined empirically (which is one purpose of the present study). Once the student meets a specified criterion on an item in the working pool, that item is removed and placed in a "review pool." Each item in the review pool is reviewed on specified dates for a specified number of times.

Description of the Software Programs Used in the Study

Two programs were used in the study. The comparison program, called the Large Teaching Set, was a commercial program that used larger teaching sets and did not provide cumulative review. The experimental program, the Small Teaching Set, was intended to exemplify the two principles of instructional design discussed above—optimal set size and cumulative review.

The distinctive instructional design features of the Small Teaching Set program include (a) individualized lessons which provide teaching and practice only on words the student does not know, (b) a practice set which consists of no more than seven words at any time, (c) a specified mastery criterion which must be met two consecutive lessons before a word is considered learned, and (d) cumulative reviews on learned words to ensure retention.

The review procedures of the two programs are different. The Small Teaching Set program provides daily review on words in the student's practice set and periodic cumulative review of words which the student has learned in the program. Once the student has mastered 10 words, the program presents a cumulative review lesson on those words. In contrast, the Large Teaching Set keeps no cumulative record of student errors, so no cumulative review is provided.

The Large Teaching Set was adapted from a commercial program developed by Davidson and Eckert (1983). Words are taught in sets of 25. The student may choose to see the words in any of four types of formats: (a) a teaching display which shows the word, its definition, and one example sentence; (b) a multiple choice quiz format; (c) an exercise in which a definition is displayed and the student must spell in the correct missing word to complete a sentence; and (d) an arcade-type game in which the student matches words to their definitions.

In summary, then, the purpose of this study was to compare two methods of computer-assisted instruction for teaching vocabulary to mildly handicapped adolescents. The study examined the effect of various cycles of practice and review on the acquisition and maintenance of word meanings. Two CAI vocabulary programs were used to present the same 50 words and definitions. The size of the daily teaching sets and provisions for daily and cumulative review varied between the two programs. Acquisition was assessed by multiple choice measures. In addition, students were asked to define 10 of the words and answer written comprehension questions on passages containing 10 of the most frequently missed words. A maintenance test was administered 2 weeks after instruction ended.

METHOD

Subjects and Setting

Thirty-eight high school students with learning disabilities in grades 9 through 12 were considered eligible for the study. Each of these students scored at least 3 years below grade level on the Reading subtest of the Woodcock-Johnson (Woodcock & Johnson, 1977). Because the Woodcock-Johnson had been administered at different times to different students by the district, the experimenters administered the Advanced 1 Level Reading subtest of the Metropolitan Achievement Test (Prescott, Balow, Hogen, & Farr, 1978) to all students. The test was administered to the entire group 3 weeks after the conclusion of the study.

The mean performance of the entire sample corresponded to the eighth percentile. Scores ranged from the 1st to 22nd percentile. In contrast, the mean score for the district was at approximately the 72nd percentile. Thus, all students performed significantly below their peers on standardized measures of reading achievement.

All students were administered a multiple choice, 50-item vocabulary pretest. They were then matched by pretest scores and randomly assigned to one of the two treatments, the Small Teaching Set program or the Large Teaching Set program. Six students who scored over 80% correct on the pretest were excluded from the study. Two students decided not to participate. During the study, four students were dropped due to frequent absences, and one student was dropped when his performance indicated that his pretest score was inaccurate. Thus, a total of 25 subjects actually participated in the study.

The study was conducted in a large, special education resource classroom in a high school in the northwest. IBM computers and color monitors were set up in the back of the classroom, away from other instructional groups.

Materials

The Small Teaching Set program (Carnine, Rankin, & Granzin, 1984) constructs individualized CAI vocabulary lessons by first testing a student on new words and then composing teaching and practice sets of only those words which the student does not know. An example of a teaching frame appears in Table 1.

The exercises in the practice set consist of three types of multiple choice items: the new word appearing alone with the correct definit[illegible] as one of five choices, the word appearing in a sentence with the correct definition as one of five choices, and a synonym (or short definition) for the word appearing in a sentence with the word as one of five choices. Examples of practice exercises appear in Table 2. For the practice exercises, the program picks from a pool of four items and randomly selects items to present. The student must get two items per word correct before the lesson ends, unless time runs out and the student selects the "escape" option to terminate the lesson.

In order to reach mastery criterion on a word and have it removed from the practice set, the student must identify the word's meaning two consecutive times on two consecutive lessons, or, in other words, four times in a row across two lessons. The word then becomes a "learned" word and moves from the practice set to the "review set." Once 10 words have been "learned" and moved to the review set, the program provides a cumulative review test on the review set. Any words missed on this cumulative review test are put back in the practice set, and the student must again meet mastery criterion on [illegible] word.

The Large Teaching Set program (Davidson & Eckert, 1983) teaches words in sets of 25. The program comes with nine levels of 75 words each for grades 4 through 12. However, for the purposes of this study, the same 50 words used in the Small Teaching Set program appeared in the Large Teaching Set program as two sets of 25 words (see "Word Selection" below).

Each time the program is run, the student goes through the same 25 words in the same order. Unlike the small set programs, some of the words are words the student already knows, since there is no individualization. At the beginning of the lesson, the program presents a menu with a choice of four formats: "word display," "multiple choice quiz," "sentence completion," or an arcade-type game.

These activities include two word display and multiple choice quizzes similar to the Small Teaching Set program, and two that are quite different. Sentence completion involves spelling the new words, and the arcade activity involves matching exercises in a game format. (For details, see Johnson, 1985, pp. 29-34, 44-51.)

Feedback to Students. Both CAI programs provided comparable immediate feedback to subjects on the accuracy of their responses. In the Large Teaching Set program, when the subject answered an item correctly, a message such as "Nice going" or "Keep it up, (name)," appeared. When the subject answered an item correctly in the Small Teaching Set program, the message, "Yes, the answer is ______," appeared.

The arcade-type game provided a type of reinforcement not available in the Small Teaching Set program. When the subject accurately "shot" the correct answer, the answer was momentarily highlighted, and a score for that shot appeared briefly in the middle of the screen. Accompanying sound effects were turned off in order not to distract other students and teachers in the room.

Both programs also provided feedback on the number of words correct. The Large Teaching Set program did this by giving the subject a percentage correct score at the end of an activity and then displaying any words missed. The Small Teaching Set program listed words on which the subject had yet to meet mastery ("curre[illegible] reviewing") and words mastered ([illegible] ready learned").

Selection of Words for the Stu[illegible] The Large Teaching Set program p[illegible]vides words, definitions, and exerci[illegible] for 25 nouns, 25 verbs, and 25 adj[illegible]tives for each level. Prior to the stud[illegible] list of these 450 words was given to middle school and high school edu[illegible]tion teachers in the district in whi[illegible] the study was to be conducted. Th[illegible] teachers picked words from this l[illegible] which they considered important a[illegible] useful for mildly handicapped s[illegible]ondary special education students know. An initial list was constructed [illegible] 107 words which were considered i[illegible]portant by two or more special edu[illegible]tion teachers. A final list of 25 ver[illegible] and 25 adjectives was developed [illegible] use in the study. All of these wor[illegible] were from the words commonly co[illegible]ered in grades, 7, 8, and 9.

Table 1
A Teaching Frame from the Small Teaching Set Program

The word ESTABLISH means SET UP.

Susie will ESTABLISH a new procedure fc[illegible] our meetings.
Susie will SET UP a new procedure for our meetings.

The bank is going to SET UP a new branch o[illegible] the other side of town.
The bank is going to ESTABLISH a new branch on the other side of town.

Table 2
Two Practice Forms from the Small Teaching Set Program

They are working to ESTABLISH an organization to protect whales.
1. make legal
2. elect
3. fund
4. set up
5. join

The doctors are going to SET UP a new eye care clinic
1. employ
2. attend
3. operate
4. cancel
5. establish

The same 50 words, 25 adjectives, and 25 verbs were entered and used in both the Small Teaching Set and Large Teaching Set programs. The same definitions were used in both the Small Teaching Set and Large Teaching Set programs. For the purposes of the study, exercises written for use in the Small Teaching Set program were the same or very similar to items which appeared in the Large Teaching Set program. The differences between the effects of the programs, if any, were intended to be a function of instructional design features.

Procedures

Following pretesting (see "Measures," below), all subjects received computer-assisted vocabulary instruction during a 20-minute session each Monday through Thursday. Since the 45-minute periods were divided into two separate sessions, some subjects began the period with a computer session and then returned to their regular instructional group, while other subjects first attended their instructional group and then completed a session on the computer.

The experimenter for the study was a doctoral student in special education at the University of Oregon. The experimenter was present for each session to ensure that the sessions lasted exactly 20 minutes, that subjects actively worked on the computer with minimal talking, that they completed as many lesson activities as possible during the 20 minutes, and that they took the optional reviews of missed words at the end of the "multiple choice quiz" exercises in the Large Teaching Set program. The experimenter also completed checklists on each subject's daily progress.

Familiarization with the Computer and Word Reading Practice. During the first 5 minutes of sessions 1 and 2, the experimenter taught the subjects how to load the program disks and start up their programs. Most subjects had little, if any, experience operating a computer. During the next 5 minutes of the first two sessions the experimenter provided word reading practice on words that were to appear in the program. The words were printed in short columns on two practice sheets, and subjects took turns reading columns out loud. If a subject misread a word, the experimenter told the subject the word and directed the subject to read the word and reread the column from the beginning.

Eight subjects, four in each treatment, displayed difficulty in accurately decoding and pronouncing words during word reading practice. Two additional word practice sessions were provided to these students.

Length of Sessions. The time scheduled for each daily computer session was exactly 20 minutes. The number of lessons that the subjects completed varied from one to two lessons. If the subject was in the middle of a lesson when the 20-minute session was about to end, the experimenter directed the subject to use the "escape" option to terminate the lesson on time. If the "escape" option was selected, no credit toward mastery was counted for words practiced during the aborted lesson.

Mastery Criterion for Large Teaching Set. The experimenter told the subjects in the Large Teaching Set program that their goal was to get a score of 84% correct (21 correct out of 25). They were told that if they scored 84% or higher, they could play the arcade-type game, and if they scored 84% or higher on two consecutive days, they would move on to a new activity.

On each day that the subject scored 84% correct or more, the subject completed the reviews and then spent the remainder of the session playing the arcade game, usually no more than twice in the time remaining. After meeting the criterion of 2 days at 84% correct or more on the multiple choice quiz, the subject was told to select the sentence completion activity; this was done only once, without review.

After meeting criterion on the multiple choice quiz and doing the sentence completion activity one time, the subject began the second list of 25 words, the same 25 verbs which were words 26 through 50 in the Small Teaching Set program. The subject followed the same sequence of activities for this second list of words. Once the subject completed all activities on both word sets, the subject was considered to have met mastery criterion for the study.

Measures

Pretest, Posttest, and Maintenance Test. A 50-item, multiple choice test requiring the student to select the correct definition of a word was developed for use in this study. Items were similar to what appeared in the practice frames (see Table 2). This instrument had a coefficient alpha reliability of 0.79. The pretest to posttest correlation was .68.

Additional Measures. All subjects took a 10-item, open-ended oral test ("What does ______ mean?"). The test was designed to measure the student's ability to recall the definition of the words taught. The training only involved recognition, a much easier response mode. The tester wrote down as much of the subject's responses as possible and also audiotaped the test. For each word the subject was asked to give a definition and to use the word in a sentence. After a subject met mastery criterion (or after session 11 for those subjects who did not meet mastery), a test consisting of three passages and accompanying inferential comprehension questions was administered. These passages contained a total of 10 verbs which were most frequently missed on the pretest. The passages were designed to assess subjects' understanding of the words in contexts other than the sample sentences which appeared in the programs.

An example from the comprehension test appears in Table 3.

Attitude Survey. The attitude survey questioned students about working on the computer and the specific CAI programs. The items asked the students how they felt about working on the computer and how much they felt they learned.

Table 3
Sample Item from Comprehension Test

Denise enjoyed her back yard. In the fall, the yard was covered with leaves. Denise had procrastinated. Saturday was cool and crisp. Denise decided to rake the leaves. At first, her hand felt cold and stiff on the rake handle. Soon, she acclimated. She enjoyed the clear, sunny skies and the rustle of the leaves.

1. Denise raked the leaves in her yard
   a. before she was supposed to.
   b. just when she was supposed to.
   c. after she was supposed to.
2. When Denise finished raking,
   a. her hands were still stiff.
   b. her hands felt fine.
   c. her hands were hot.
3. Denise enjoyed the clear, sunny skies and
   a. the sound of the leaves.
   b. the fall colors.
   c. the smell of the leaves.

RESULTS

Time to Mastery

A time-to-mastery design was used to examine whether there was a significant difference between the times required to reach mastery by subjects in the two programs. Table 4 presents a summary of the number and percentages of subjects who met mastery within 11 sessions and the mean number of sessions to mastery for subjects in both treatments. The study was terminated after the 11th session because the experimenter felt that the subjects who were still struggling to reach mastery were no longer benefiting from instruction. For the students who met criteria, the mean number of sessions to mastery was 7.6 for subjects in the Small Teaching Set program and 9.1 for subjects in the Large Teaching Set program. Results of a t-test indicate that this difference in sessions to mastery is statistically significant, $t(16) = 1.87$, $p < .05$.

Posttest and Maintenance Test

The 50-item, criterion-referenced, multiple choice test was administered to each subject as a posttest after meeting mastery. Those subjects who did not meet mastery by the end of the 11th session were administered the multiple choice test after session 11. The same test was re-administered (with the order of items changed in a random fashion) as a maintenance test 2 weeks later. A summary of pretest, posttest, and maintenance test results is presented in Table 5.

A 2 × 2 analysis of variance was performed on the posttest and maintenance test scores (see Note 1). The between-subjects factor was type of instruction, and the within-subjects factor was time of testing. Results of the ANOVA indicated that there was no effect for type of instruction, $F_{(1,22)} = 0.33$. Results of the ANOVA demonstrated that there was a slight drop in performance between posttest and maintenance test for subjects in both groups, $F_{(1,22)} = 4.94$, $p < .05$. Mean performance was close to mastery level for both groups on both measures, 84% to 87% on the posttest and 81% to 84% on the maintenance test. Subjects in both programs learned as much, as measured by the criterion-referenced posttest, and retained as much, as measured by the maintenance test.

Additional Measures

Each subject was administered an open-ended oral test on word meanings after session seven. A maximum of 2 points was awarded to each item, 1 for a correct definition and 1 for a appropriate sentence. Partial credit (point) was given to responses whic were correct but incomplete. Results a t-test on subjects' scores on th measure indicated that differences b tween groups were nonsignifican $t(22) = .45$. The mean was 6.4 for th Small Teaching Set group and 7.2 fc the Large Teaching Set group; star dard deviations were 4.7 and 4.4, r spectively. On the written compreher sion test, means were 5.1 ($SD = 1.4$) fc the Small Set and 4.7 ($SD = 2.1$) for th Large Set. This difference was nc significant.

Attitude Survey

Results of the attitude survey ir dicated that, for the most part, subjec responded favorably toward con puter-assisted instruction and the prc grams. Twenty-three of twenty-fou subjects felt the computer helped the learn new words, and one subject ir dicated that "maybe" the compute helped.

In answer to the question "Did yo enjoy working on the computer?" sub jects answered on a 4-point scale, wit 1 being "not very much" and 4 bein "very much." The mean scores were 3. for subjects in the Small Teaching Se program and 2.8 for subjects in th Large Teaching Set program. Result

Table 4
Percentage of Subjects Reaching Mastery Within 11 Sessions and Mean Number of Sessions to Mastery for Both Experimental Groups

Group	N	Number of Subjects Reaching Mastery	Percentage of Subjects Reaching Mastery	Mean Number of Sessions to Mastery	Standard Deviation
Small Teaching Set	12	10	83	7.6	1.9
Large Teaching Set	12	8	67	9.1	1.5

Table 5
Mean Score, Standard Deviation, and Mean Percentage Correct on 50-item Pretest, Posttest, and Maintenance Tests

		Pretest			Posttest			Maintenance Test		
Group	N	M	SD	Mean %	M	SD	Mean %	M	SD	Mean %
Small Set	12	24.7	8.1	49.4	42.0	4.0	84.0	40.5	7.1	81.0
Large Set	12	24.75	7.8	49.5	43.7	7.7	87.8	42.0	8.7	84.0

of a Mann-Whitney U Test indicate that this difference was significant. $U = 43.5, p < .01$. Nineteen subjects indicated they would like to learn more on a computer, and three subjects indicated that "maybe" they would. Two subjects, both in the Large Teaching Set program, indicated they would not like to learn more on a computer.

Table 6
Comparison of Mildly Handicapped with Nonhandicapped Samples: Multiple Choice Test

Group	Test	*N*	*M*	*SD*	Mean Percentage Correct
Small Teaching Set	Posttest	12	42.0	4.0	84.0
Large Teaching Set	Posttest	12	43.7	7.7	87.4
Nonhandicapped Comparison (10th grade)	—	26	40.3	4.9	80.6

Comparison with Nonhandicapped Subjects

One purpose of this study was to see if, with special instruction, students with learning disabilities could perform at the same level as students without learning difficulties. Thus, the 50-item multiple choice test was administered to nonhandicapped 10th-grade students in a randomly selected English class in order to provide a comparison standard. As Table 6 demonstrates, the mean posttest scores of the mildly handicapped subjects were at a similar level to the nonhandicapped students' mean scores. The LD students' scores were 84% and 87.4%, the nonhandicapped students' mean was 80.6%. Despite a significant difference in reading ability, as measured by the Metropolitan Achievement Test (Prescott et al., 1978) and the Woodcock-Johnson (Woodcock & Johnson, 1977), these LD students were able to learn word meanings so that, in this instance, they could perform on a level similar to that of their nonhandicapped peers. The difference in performance between the LD groups and the nonhandicapped group was not significant. This result indicates that with intelligent use of CAI, the LD students performed at a level similar to that of their nonhandicapped peers.

DISCUSSION

The results of this comparison of two methods of computer-assisted vocabulary instruction with mildly handicapped high school students will be discussed in terms of (a) time required to reach mastery criterion, (b) growth in word knowledge, (c) transfer of learning, and (d) student attitudes toward computer-assisted instruction.

Time to Mastery. In previous studies which attempted to improve students' word knowledge through the direct teaching of word meanings, the effects of various instructional procedures were compared. Those studies which demonstrated sizable gains did so at a large expense of instructional time. This study was the first to focus on efficiency as a dependent variable.

The one unequivocal finding of the study was that subjects taught with the Small Teaching Set program reached mastery criterion on the set of 50 words significantly faster than subjects taught with the Large Teaching Set program. Also, more students in the Small Teaching Set program reached mastery within 11 sessions.

Given that the groups achieved equivalent levels of performance on the multiple choice tests, their difference in acquisition rates becomes even more meaningful. Subjects taught with the Small Teaching Set program required less time to meet mastery criterion on the words, yet their posttest performance and retention was equal to that of subjects in the other treatment. These findings have important implications for teachers of low performing or reading disabled students. An efficient, computer-assisted method of vocabulary instruction could provide an additional tool for teaching vocabulary, without placing further burdens on teachers' time.

Growth in Word Knowledge. The growth in word knowledge evidenced by both groups provides encouraging support for the use of computer-assisted instruction in vocabulary with mildly handicapped students. Each group started with a pretest mean score of about 50%; after seven 20-minute sessions, each group's mean score was around 80% (Johnson, 1985). When subjects were tested after reaching the mastery criterion determined for their program, or after 11 sessions for those 6 subjects who did not reach mastery criterion, each group's mean score was around 85%. These scores reflected a commonly accepted minimum mastery level: approximately 85% correct. Finally, on the maintenance test, administered 2 weeks later, each group's mean score was above 80%. Although the drop between posttest and maintenance test was statistically significant, 80% is still a high level, especially considering that subjects began at a 50% level.

Performance on Measures Requiring Alternative Response Formats. Subjects' low scores on the two transfer measures should not be surprising. Students were taught with a multiple choice format, requiring recognition. The open-ended oral test of word meanings is a more difficult task. The test of passage comprehension also required response modes considerably more difficult than the multiple choice response mode of the CAI programs. While subjects scored 85% correct on the multiple choice posttest, they scored approximately 35% on the open-ended oral test of word meanings and 50% on the comprehension test. Lack of specific training on the kinds of tasks tested by the transfer measures was the primary reason for subjects' low scores. Without specific training on the response characteristics of transfer tasks, mildly handicapped adolescents often fail to generalize academic skills (Alley, Deshler. Clark, Schumaker, & Warner, 1983). The implication is clear. Students need training in transfer of skills learned in CAI formats.

Student Attitudes Toward Instruction. On the attitude survey, most subjects indicated they enjoyed computer-assisted instruction and the CAI programs. When asked to indicate what they specifically liked about working on the computer, perhaps the most telling response was, "It helps keep your mind on what you were [sic] doing." Subjects' positive response to computer-assisted instruction lends credence to the claims of Budoff, Thormann, and Gras (1984) that advantages of CAI with special education students include increased attention, immediate feedback about performance, immediate reinforcement, and motivation.

On the question "Did you enjoy working on the computer?" subjects rated the Small Teaching Set program significantly higher, as results of a Mann-Whitney Test demonstrated. This finding is interesting as it relates to the design of CAI programs. While the Large Teaching Set program had an "arcade" type game, the Small Teaching Set was designed to foster rapid learning. During the study, some subjects in the Small Teaching Set program occasionally asked the experimenter why they didn't get to play a game like the one in the other program. The experimenter wondered if this difference in programs might bias the subjects against the Small Teaching Set program. The results indicated, however, that the subjects in the Small Teaching Set program, which tailored lessons to their individual learning needs, rated that program more highly.

This finding is important for those designers of CAI programs who apparently believe that for educational software to be motivating, it must approximate computer games which are popular in video arcades and in the home video market. Results of the attitude survey in this study do not support such reasoning.

Suggestions for Future Research. This study contrasted two packaged CAI programs. The major differences between the programs related to the size of the teaching and practice sets and the procedures for individualization and cumulative review. Yet other subtle differences between the software programs may have affected the outcomes. Since an effect for time to mastery was clearly demonstrated for the Small Teaching Set program, and posttest performance levels were equivalent for the two groups, future research might focus on only the Small Teaching Set program. By varying the size of teaching and practice sets, and by comparing different schedules for cumulative review exercises, more exact effects of these variables could be measured.

Although most subjects with learning disabilities learned from the CAI programs, as expected, performance levels were low on the transfer measures. If disabled readers are to benefit from computer-assisted vocabulary instruction, future studies need to investigate the effects of integrating computer-assisted and teacher directed instruction. This integration should follow some of the principles articulated by Beck et al. (1982), although, hopefully, in a more efficient fashion.

ABOUT THE AUTHORS

Gary Johnson received his doctorate in special education in 1985 from the University of Oregon. Currently he is responsible for teacher training in Washington School District in Phoenix, Arizona. His research emphasis is on supervision and administration of a non-pullout program for special education and Chapter I students in reading. He has coauthored several texts commonly used with special education students including the Science Research Associates Corrective Reading Series. ***Russell Gersten*** *is an associate professor of special education at the University of Oregon. He received his doctorate from Oregon in 1978. Since then his research has focused on longitudinal evaluations of direct instruction programs, effective staff development strategies for schools with students achieving below the norm, and instructional design. These studies have been published in major journals such as American Educational Research Journal, Exceptional Children, and Educational Leadership. His current research focuses on teaching strategies in reading comprehension, effective strategies for working with language minority students, and longitudinal research.* ***Douglas Carnine*** *is associate professor of special education at the University of Oregon and directs several grants which apply to videodisc and computer technology in special education. He also develops CAI and videodisc programs as well as print programs. He is coauthor Direct Instruction Reading, Direct Instruction Mathematics, and Theory of Instruction. Address: Johnson, Mountainview Elementary School, Mountainview Rd., Phoenix, AZ 85021.*

ACKNOWLEDGMENTS

This research was supported in part by Grant G00840060 from the Office of Special Education and Rehabilitation Services, U.S. Department of Education. We wish to acknowledge the assistance of John Woodward, research coordinator at the University of Oregon, and Annette Howarth Zwettler, Patty Bonine, Nick Chrones, and Jeu Maines of the school district. We thank Craig D Lisa Moore, and Bruce Miller for their editorial assistance. We would also like to thank IBM for providing equipment for this research.

NOTE

1. The pretest was not included in the ANOVA for the following reason. The time between posttest and the maintenance test was the same for all subjects (2 weeks), but the time between pretest and posttest was not the same for different subjects. Thus repeated measures could not be used.

REFERENCES

Alley, G.R., Deshler, D.D., Clark, F.L., Schumaker, J.B., & Warner, M.M. (1983). Learning disabilities in adolescent and adult populations: Research implications (part II). Focus on Exceptional Children, 15(9), 1-14.

Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), Comprehension and teaching: Research reviews (pp. 77-1... Newark, DE: International Reading Association.

Beck, I.L., Perfetti, C.A., & McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. Journal of Educational Psychology, 74, 506-521.

Budoff, M., Thormann, J., & Gras, A. (1984). Microcomputers in special education. Cambridge, MA: Brookline Books.

Carnine, D., Rankin, G., & Granzin, A. (198... Learning vocabulary. Unpublished computer program. Eugene, OR.

Davidson, J.G., & Eckert, R.K. (1983). Word attack [Computer program]. Ranchos Palos Verdes, C... Davidson & Associates.

Deese, J.E., & Hulse, S.H. (1967). The psychology of learning. New York: McGraw-Hill.

Furukawa, J.M. (1970) Chunking method of determining size of step in programmed instruction. Journal of Educational Psychology, 61, 24... 254.

Gay, L.R. (1983). Temporal position of reviews and its effect on the retention of mathematics rules. Journal of Educational Psychology, 64, 17... 182.

Holl, J.F. (1971). Verbal learning and retention. Philadelphia: Lippincott.

Johnson, G.L. (1985). Computer-assisted vocabulary instruction with mildly handicapped ...

school students. Unpublished doctoral dissertation. University of Oregon. Eugene.

McGeoch. J.A. & Irion. A.L. (1952). The psychology of human learning. New York: Longmans. Green.

McKeown. M.G.. Beck. I.L.. Omanson. R.C.. & Perfetti. C.A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. Journal of Reading Behavior. 15. 3-18.

Merrill. P.F.. & Salisbury. D. (1984). Research on drill and practice strategies. Journal of Computer-Based Instruction. 11(1). 19-21.

Mezynski. K. (1983). Issues concerning the acquisition of knowledge. Effects of vocabulary training on reading comprehension. Review of Educational Research. 53. 253-279.

Miller. G.A. (1956). The magical number seven. plus or minus two: Some limits on our capacity for processing information. Psychological Review. 63. 81-97.

Pearson. P.D.. & Gallagher. M. (1983). The instruction of reading comprehension. Contemporary Educational Psychology. 8. 317-344.

Prescott. G.A.. Balow. I.H.. Hogan. T.P.. & Farr. R.C. (1978). Metropolitan achievement tests: Advanced level I. New York: Harcourt Brace Jovanovich.

Siegel. M.A.. & Misselt. A.L. (1984). Adaptive feedback and review paradigm for computer-based drills. Journal of Educational Psychology. 76. 310-317.

Stahl. S.A. (1983). Differential word knowledge and reading comprehension. Journal of Reading Behavior. 15. 33-50.

Thormann. J.. Gersten. R.. Moore. L. & Morvant. M. (in press). Microcomputers in special education classrooms: Themes from research and implications for practitioners. Computers in the Schools.

Tierney. R.J.. & Cunningham. J.W. (1984). Research on teaching reading comprehension. In P.D. Pearson (Ed.). Handbook of reading research (pp. 609-656). New York: Longman.

Woodcock. R.W.. & Johnson. M.B. (1977). Woodcock-Johnson psycho-educational battery. Boston: Teaching Resources.

Elaborated Corrective Feedback and the Acquisition of Reasoning Skills

A Study of Computer-Assisted Instruction

Maria Collins
Boise State University

Douglas Carnine

Russell Gersten
University of Oregon

TE2

Author Note

This research was supported in part by U.S. Department of Education Grant Number G008400660, Office of Special Education and Rehabilitation. The authors wish to thank John Woodward and Phillip Bourbeau for their editorial assistance.

TE2

Figure Caption

Figure 1. Example of a computer display of a valid argument.

TE2

Elaborated Corrective Feedback and the Acquisition of Reasoning Skills

A Study of Computer-Assisted Instruction

Much of the recent literature on improving special education teaching practices has stressed the importance of providing specific academic feedback to students when they make errors (Carnine, 1980; Peith, Polsgrove & Semmel, 1981; Stevens & Roseshine, 1981). Correlational studies have shown that more effective teachers provide immediate corrective feedback to students when errors are made (Englert, 1984; Gersten, Carnine & Williams, 1982). Furthermore, effective teachers offer students information on strategy or process that can be used to deduce the correct answer (Englert, 1984; Gleason, et al., 1985; Good & Grouws, 1977; Stallings, 1975). However, relatively little experimental research has evaluated the impact of the type of corrective feedback provided to learners.

A meta-analysis of the limited research on corrective feedback by Lysakowski and Walberg (1982) suggests that detailed corrective feedback is superior to merely telling students whether their answers are right or wrong. These authors suggest that students need to see an overt model of all the steps necessary for an appropriate response. By observing a model of all the steps necessary in obtaining a correct response, students receive detailed information on how to solve the problem. This procedural knowledge should be of use when they encounter similar types of problems. Merely telling students they are wrong, does not help them to solve the problem correctly.

The corrective feedback provided in these various studies can be placed in one of three groups: a) minimal feedback (telling students if their answers are

TE2

correct or incorrect, b) basic feedback (telling students if their response are correct, and, if incorrect, supplying the correct answer) and c) elaborated corrective feedback (telling students if their responses are correct, and, if correct, supplying the correct answer and providing additional information about the correct answer). Elaborated corrective feedback is similar to the correction procedure Lysakowski and Walberg (1982) advocate, i.e., detailed feedback on how to correctly solve the problem. Only two studies have addressed elaborated corrective feedback directly with a handicapped population. The results of the research have been mixed.

A longitudinal study by Siegel and Crawford (1983) examined the effects of elaborated feedback on mentally retarded students selected from both elementary and secondary school settings. The researchers compared students taught with elaborated corrective feedback with students trained on the discrimination tasks without any corrective feedback. The elaborated corrections group demonstrated superior results over the no feedback group on a transfer test given shortly after training as well as on the two-year follow-up test. However, this study did not compare the effects of elaborated feedback with basic feedback in which students are told whether their response is correct, and, if correct, are told the correct response.

Only one study, by Meyer (1982), addressed this issue with handicapped learners. Meyer examined the relative effectiveness of elaborated correction procedures and basic correction procedures in teaching word attack skills to two groups of learning disabled students in the secondary grades. The only difference between the two groups was that, when students in the basic correction group made an oral reading error, the teacher told them that they were incorrect

TE2

and then sounded out the word correctly for them. In the elaborated correction group, when student made errors, the teacher reviewed the appropriate decoding rules, and then had the student sound out the word with her or him. Meyer reported no significant difference in oral reading performance between students in the two groups. Her findings are different than those of Carnine (1980), in which non-handicapped pre-schoolers who received elaborated correction procedures were able to decode more new words than students who received basic corrections.

There are two possible reasons for this anomaly. This first is that the error rate was so low in Meyer's (1982) study--less than 4 percent--that elaborated corrective feedback may not have been necessary. The few errors may have been due to temporary inattentiveness, slips of the tongue, rather than failure to know--and know how to apply the word atta rules. The second possible explanation relates to the difficulty of the skill to be taught. Grant, McAvoy, & Keenon (1982) argue that elaborated corrective feedback is only necessary for cognitive skills that are new and/or complex for the learner. The word attack skills for sounding out CVC and CVCe words were hardly new, or terribly complex to the high school students, many of whom had been exposed to similar rules for eight years. On the other hand, these same phonics applications to word reading would seem complex to preschoolers (Carnine, 1980).

The current study, like the Meyer study, compared the effects of elaborate corrective feedback and basic feedback in teaching remedial and learning disabled students at the high school level. However, in this case, a complex cognitive task, inferential logic, was selected.

Formal logic was selected as the content for studying feedback because of its complexity. Also, instruction in logic and reasoning skills has been generally lacking in the curricula of both "regular" and "special education" students in elementary and junior-high settings (Cnerkes, 1979). The difficulty in teaching reasoning skills may account for its absence from the curriculum. A major purpose of the study was then to evaluate whether remedial and learning disabled secondary level students could learn logical analysis skills from computer-assisted instruction.

However, the primary intent of the study was to examine whether low-performing students who receive elaborated correction procedures would perform significantly better than students provided basic corrections. We also examined any differences regarding acquisition time between students as well as attitudes toward the instruction.

Method

Subjects

Subjects were selected from six remedial (Chapter One) and special education classrooms in two secondary schools in western Oregon. To qualify for the study, subjects had to demonstrate: a) a reading comprehension deficiency of no more than three years on district-administered, standardized reading comprehension tests; b) at least fifth grade oral reading level, as determined by teacher judgment; and, c) an understanding of the concept of large and small classes, as evidenced by passing an experimenter-developed classification pretest. (See Measures section.) Reading comprehension scores were considered to insure that subjects did not have a severe problem with reading comprehension. If reading comprehension was a serious problem, instruction would need to focus on reading comprehension, rather than reasoning skills.

One-hundred and eighteen secondary students in remedial and special education were screened. Of the 28 who completed the study,13 were learning disabled and 15 were remedial. The Woodcock Reading Mastery Test (1978) was administered to all 34 subjects. The subjects were matched on scores from the Word Comprehension subtest of the Woodcock Reading Mastery Test and then randomly assigned to the Basic Correction or Elaborated Correction group (Cook and Campbell, 1979). Mean scores for the two groups on the screening test were virtually identical, 23.67 (SD = 1.15) for the Elaborated Corrective Feedback Group and 23.85 (SD = 1.28) for the Basic Corrective Feedback Group. Mean scores on the Woodcock Word Comprehension Subtest were identical for both groups. Both were 35.79. (Standard Deviations were 6.6 and 8.7 respectively).

Materials

The Reasoning Skills computer-assisted instruction program (Engelmann, Carnine & Collins, 1983) was designed to teach students two major objectives: a) to draw conclusions from two statements of evidence and b) to determine whether a three-statement argument was logical or illogical. To enter the program students had to have class inclusion skills, i.e., place members of a smaller class in a larger classes (e.g., car fits in the class of vehicle, fox terrier fits in the class of dog). The program then taught students about overlapping classes (e.g., tall things and buildings) and non-overlapping classes (e.g. boys and girls).

Students learned three key words (all, some, no) and their corresponding class relationships:

for all the smaller class is included in the larger class

TE2

for <u>some</u> the classes overlap

for <u>no</u> the cla'ses do not overlap.

Table 1 presents cases of <u>all</u>, <u>no</u> and <u>some</u> arguments.

Insert Table 1 about here

The following three rules formed the basis for the first phase of instruction -- constructing arguments. The first rule is:

"<u>Start the conclusion with the key word</u>. Before learning this rule, students learn what a key word is. They learn that in a syllogistic argument, one of the statements of evidence begins with the word <u>all</u>. The word at the beginning of the <u>other</u> statement is the key word. The key word becomes the <u>first word</u> in the conclusion (see figure 1).

Insert Figure 1 about here

In a conclusion, the key word is followed by two classes. The second rule tells students how to identify the two classes named in a conclusion:

"<u>The conclusion names the classes that are named only once in the evidence</u>." In any syllogistic argument, one class is always named twice in the evidence. In Figure 1, men is named twice. The once-named classes ("humans" and "those with brown hair") are underlined in Figure 1. These are the two classes that appear in the conclusion.

The third rule applies to a subgroup of conclusions, only those with <u>all</u> as the key word. An <u>all</u> statement always places a smaller class in a larger class; thus a conclusion beginning with <u>all</u> must name the smaller class first. (Class order isn't important for some and no conclusions beginning with some or no). In the following all statements, a small class is placed in a larger class:

All men are animals.

All animals are living things

Men is a smaller class than animals. Animals is a smaller class than living things. The small class must be named at the beginning of the all statement. Thus, in the conclusion ("All men are living things") small class, men, must be named first.

The following example is taken from the program. It includes the Elaborated and the Basic Corrections provided when an error was made.

All athletes are humans.

All football players are athletes

Question: Enter the number of the smallest class in the evidence.

The correct answer is football player. If student responses were a) athletes or b) humans, a correction procedure was supplied.

Basic Correction: "The correct answer is football players."

Elaborated Correction: The smallest class is named once and is named at the beginning of an all statement. Football players is named once and is named at the beginning of an all statement, so football players is the smallest class.

This type of correction may appear contrived at first, yet is, we believe, essential to providing students with practice on the appropriate form of a logical argument.

The other major objective of the program was to teach students to identify unsound arguments. This is discussed in detail in Collins (1984). For logically unsound arguments, students were taught to specify one of three reasons why an argument is unsound.

In the study, the Elaborated Correction treatment used an unaltered copy of the Reasoning Skills program. The Basic Correction treatment used a modified version of the program. In both conditions, students worked individually on an Apple IIe microcomputer. In the Basic Corrections program, if students made errors, they were given the correct answer, as specified above. This was the only difference between the two conditions.

One feature that was the same for both treatments was a provision for review of missed items. When students missed a question, the program presented the question again later in the lesson. Thus, all missed questions were answered correctly before the end of a lesson.

Measures

Screening Measure and Pretests

The group-administered screening test was designed to determine whether subjects had any background in specifying which objects fit into a particular class. For example, which of the following are vehicles (circle two) a) cars; b) toy dolls; c) airplanes; d) windmills. The test also assessed whether students had mastered the concepts of larger and smaller classes (e.g., which class is larger, the class of trucks or the class of vehicles?). Students who scored less than 90 percent on this measure were excluded from the study. Knowledge of classification is a prerequisite skill for the reasoning skills program, because large and small class information is used in the deductive process.

The Word Comprehension subtest of the Woodcock Reading Mastery Test was given individually to each student prior to the study. The subtest contains 70 items in analogy form. Split-half reliabilities range from .83 to .96, depending on grade level. The median concurrent validity correlation (with other subtests in the instrument) is .86. This subtest contains 70 items in analogy form. Since this subtest required an understanding of classification, a skill requirement for the CAI program, subjects were matched according to the scores on this subtest prior to random assignment to one of the two treatment groups.

Students were also given Form A of the Test of Formal Logic (see below) as a pretest.

Dependent Measures

These included:

a) a test of formal logic measuring acquisition of the CAI progam's content,

b) a transfer test, designed to assess generalization of subjects' skills to new material, and

c) an attitude survey, measuring students' attitudes toward the computer-assisted instruction and the reasoning program.

Test of Formal Logic (Collins, 1984). This was the primary dependent measure in the study. The purpose of this test was to measure a student's ability to construct and analyze syllogistic arguments. Two alternate forms of the test were designed. Form A was used as a pretest measure and then readministered two weeks after treatment as a maintenance test. Form B was given to all subjects immediately following the treatment as a post-test measure.

Fifteen items tested the skills for drawing conclusions from stated evidence. The following item is an example:

TE2

Here's the evidence:

All trains are vehicles.

Some trains are steel objects.

1. What will be the first word in the conclusion (all, some or no)?

_____

2. Write the conclusion:

____

In the remaining eight items, students analyzed a syllogism and determined if it was "sound" or "not sound," and if "not sound", explained why.

The content validity of the two forms of the Test of Formal Logic was assessed by four university instructors and 15 teachers in a graduate level college class. These persons were chosen because they were considered potential "users" of the program. These teachers examined the items and indicated those items which they felt were inappropriate and their reasons for exclusion. Based on their comments, these items were either dropped or revised.

The internal consistency reliability (coefficient alpha) for the instrument was .90 for Form A and .91 for Form B. This was based on a sample of 28 students not included in the experimental study. Parallel form reliability between Forms A and B was .84.

Transfer Test

This 15-item test evaluated subjects' abilities to generalize what they had learned on the computer to similar analytical tasks, but in prose paragraph form. The transfer test was devoted to the more difficult objective of the program -- deciding whether arguments were sound, and, if not sound, giving a reason. Two examples of these

TE2

arguments appear below:

A. _________ Michael and Sam were sitting in the Science Lab at school. Michael had a bottle of arsenic in his hand. Sam said, "Hey, you know that all poisons are dangerous!" Michael looked at the bottle of arsenic. Michael said, "It says on this bottle that arsenic is a poison. So, arsenic must be dangerous!" "Yes, I guess so", replied Sam.

B. _________ It is very easy to describe high school students in America. All high school students are people who want to drive cars. They always talk about borrowing their parents' car and driving around town. We can also say that all high school students are people who have to take tests. I wonder why all people who have to take tests also want to drive cars.

The transfer test was given to subjects on the day following the completion of training on the CAI program. Both invalid and valid arguments were included, with a heavier emphasis on invalid arguments (5 valid and 10 invalid arguments). The argument forms paralleled those arguments taught in the reasoning skills program.

Results

Test of Formal Logic

Table 2 presents the descriptive statistics for the pretest, the post and

Table 2 here

maintenance administrations of the Test of Formal Logic. The Elaborated Corrective Feedback group had a slightly lower mean pretest score than the Basic Corrective Feedback group on the pretest; this difference was not significant.

TE2

The posttest mean for the Elaborated Correction group was 18.0 and for the Basic Correction group 15.2. On the maintenance test, respective means were 18.7 and 13.2.

A 2x3 analysis of variance with one between subjects factor (Type of Correction) and one within subjects factor (Time of Test) was performed on the data. A planned comparison was utilized, contrasting performance of the two groups on both the post-treatment measures (immediate posttest and maintenance test). The ANOVA indicated a significant difference favoring the Elaborated Corrections group; $F_{(1,78)} = 9.87$; $p < .001$. This result indicates a significant effect for the use of elaborated corrective feedback on both immediate posttest and maintenance test.

Transfer Test

The mean score was 12.07 for the Elaborated Group and 10.29 for the Basic Group. Standard deviations were 2.97 and 2.56 respectively. These results indicate a significant difference between the two training groups; $t_{(25)} = 1.70$, $p < .05$, again favoring the Elaborated Corrective Feedback group.

Time Per Lesson

Data were collected on the time students took to complete each of the five lessons. The purpose of this analysis was to see whether students in the Elaborated Corrections group took more time to complete the lessons. The average daily lesson time was 24.86 minutes for the Elaborated Corrective feedback group and 23.93 for the basic group. A 2x5 analysis variance (ANOVA) with repeated measures was performed on the time-per-lesson data. The between-subjects factor was type of corrective feedback training method. The within-subject factor was time per lesson for five lessons. The analysis did not show any significant difference between the time the Elaborated group and the Basic group took to complete lessons.

Student Attitudes Toward Instruction

A survey was administered after the lessons to all students to determine whether any difference existed between the two groups on attitudes toward instruction. The items and mean scores on a Likert Scale are presented in Table 3.

Table 3 Here

There was only one significant difference between the two groups, on item 4, $t(26) = 5.14$, $p < .001$. This item asked students how well they felt they could detect unsound arguments. Students in the Elaborated group felt they were better in picking out a sound argument. Thus, the self-report parallels the performance data.

Discussion

This study was the first to explore experimentally the effectiveness of elaborated corrective feedback in teaching a complex cognitive skill to handicapped learners. These results, unlike those of Meyer (1982), indicate this is an effective instructional procedure. In the Meyer study, students were familiar with the phonic rules. The difference in the content of the two studies suggests elaborated correction procedures may be necessary only when students are learning new or complex rules or strategies. The merit of elaborated correction procedures is confirmed by the significantly higher scores on the transfer test and the greater confidence in identifying sound arguments for the students in the Elaborated Corrections group.

At first glance, though, the roughly equivalent times for the two treatments to complete the lessons seem anomalous. With more text to read in elaborated corrections, that treatment would seemingly take longer to complete the lessons. Completion times were not significantly greater for the elaborated corrections group, however.

The extra time required to read the elaborated corrections may have been compensated for by faster acquisition of the material. As mentioned earlier, the computer would return a student to items that were missed earlier in a lesson. If elaborated corrections resulted in fewer mistakes, students would spend less time returning to missed items. This interpretation suggests that taking more time early in a complex instructional sequence to offer elaborated corrections may in fact lead to savings in instructional time later in a program.

Both the basic and elaborated correction groups improved their reasoning skills as measured by the dependent variable. The groups demonstrated a mean score of 68-70% on the posttest (a dramatic gain from the mean scores of 25 to 34 percent on the pretest). The systematic design of the instruction -- particularly through a series of carefully controlled rules--may have contributed to this gain.

Well-designed computer programs could increase students' academic engaged time in two ways -- by replacing some time spent in seatwork with interactive instruction, which tends to have higher engagement rates (Rosenshine, 1983) and by having students receive interactive instruction outside of school. These applications are particularly significant because the program was a true tutorial. Reasoning skills were acquired without any instruction from the teacher. Typically, CAI programs merely provide drill and practice exercises to supplement teacher instruction. Here the program did all the instructing. This finding goes against prevalent research (Winkler, et al., 1984) suggesting that properly designed CAI best serves as a supplement to conventional instruction.

In this study, CAI succeeded at providing initial instruction. With suitable content (e.g., rule based) and viable instructional design (Engelmann & Carnine, 1982), CAI can play a much larger role in instruction.

TE2

Table 1. Examples of Syllogistic Arguments

1. Example of an argument with "All" Statements of evidence:

All men are human.

All humans require oxygen.

(So) All men require oxygen.

2. Example of an argument with a "No" statement of evidence:

All men are human

No birds are human.

(So) No birds require oxygen.

3. Example of an argument with a "Some" statement of evidence:

All men are human.

Some men are over six feet tall.

(So) Some humans are over six feet tall.

TE2

Table 2

Means, Standard Deviations and Mean Percent of Items Correct on Criterion Referenced Measure (Test of Formal Logic)

	Time of Test		
	Pre	Post	Maintenance
Elaborated Corrections (N = 14)			
M	6.0	18.0	18.7
SD	3.67	4.69	3.97
Mean Percent Correct	25.1	78.3	81.3
Basic Corrections (N = 14)			
M	7.0	15.2	13.2
SD	3.55	6.14	6.81
Mean Percent Correct	30.4	55.1	57.4

Table 3

Student Attitudes Toward the Program

Questions	Group			
	Elaborated Corrections		Basic Corrections	
	M	SD	M	SD
1. Was this program interesting to you? (1-yes 0-no)	.79	.43	.86	.36
2. How well did you enjoy the program? (2-I enjoyed it 1-I feel OK about the program 0-I did not enjoy the program)	1.57	.51	1.36	.63
3. Do you feel you understood the rules and examples in the program? (2-all the time 1-sometimes 0-never)	1.29	.47	1.20	.58
4. Do you feel like you can now pick out a sound argument? (2-all the time 1-sometimes 0-never)	1.43	.51	1.07	.27

TE2

References

Carnine, D. (1977). Phonics versus look-say: Transfer to new words. Reading Teacher, 30(6) 636-640.

Carnine, D. (1980). Phonic versus whole-word correction procedures following phonic instruction. Education and Treatment of Children, 3(4), 323-330.

Cherkes, M. (1979). Using logic in special classrooms. Academic Therapy, 15(2), 165-171.

Collins, M. (1984). The Effectiveness of Computer-Delivered Correction Procedures on Low-Performing Secondary Student's Reasoning Skills. Doctoral Dissertation, University of Oregon, Eugene, Oregon.

Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis issues for field studies. Chicago: Rand McNalley.

Engelmann, S., & Carnine, D. (1982). Theory of Instruction. New York: Irvington.

Engelmann, S., Carnine, D., & Collins, M. (1983). DIAL Reasoning Skills Program. Eugene: Engelmann-Becker Corp.

Englert, C. (1984). Effective direct instruction practices in special education settings. Remedial and Special Education, 5(2), 38-47.

Gersten, R., Carnine, D., & Williams, P. (1982). Measuring implementation of a structured educational model in an urban school district: An observation approach. Education Evaluation and Policy Analysis, 4, 67-74.

Gleason, M., Gersten, R., Glang, A., Kelly, B., Kinder, D., & Gurney, D. (1985). Observational Study of Beginning Special Education Teachers Performance. Unpublished manuscript. Eugene, OR University of Oregon.

Good, T., & Grouws, D. (1977). Teaching effects: A process-product study in fourth grade mathematics classrooms. Journal of Teacher Education, 28, 49-54.

Grant, L., McAvoy, R., & Keenan, J.B. (1982). Prompting and feedback variables in concept programming. Teaching of Psychology, 9(3), 173-177.

Lysakowski, R., & Walberg, H. (1982). Instructional effects of cues, participation and corrective feedback: A quantitative synthesis. American Educational Research Journal, 19(4), 559-578.

Meyer, L.A. (1982). An experimental study of the effects of word analysis and word supply correction procedures during word attack. Reading Research Quarterly, 4, 544-555.

Reith, H.J., Polsgrove, L., & Semmel, M.I. (1981). Instructional variables that make a difference: Attention to task and beyond. Exceptional Education Quarterly, 2(3), 61-71.

Rosenshine, B. (1983). Teaching functions in instructional programs. Elementary School Journal, 83, 335-351.

Siegel, M.A., & Crawford, K.A. (1983). Two-year follow-up study of discrimination learning by mentally retarded children. American Journal of Mental Deficiency, 88(1), 76-78.

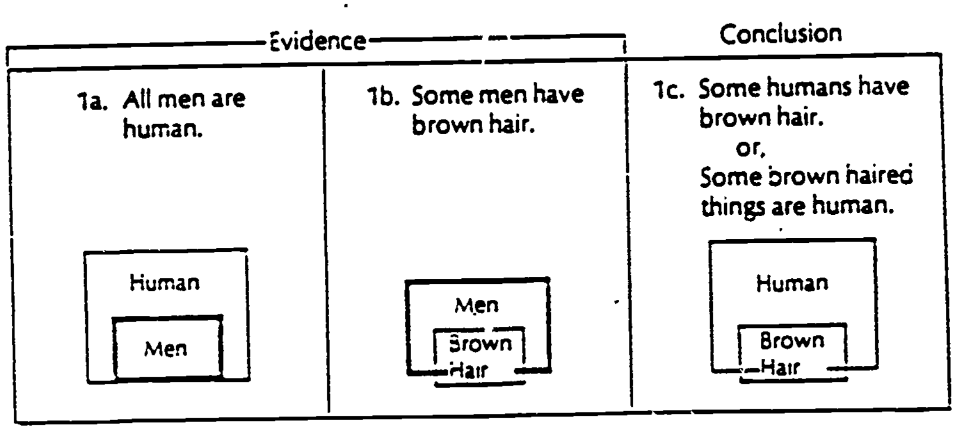
Stallings, J. (1975). Implementation and child effects of teaching practices in Follow Through classrooms. Monographs of the Society of Research in Child Development, 40.

Stevens, R. & Rosenshine, B. (1981). Advances in research on teaching. Exceptional Education Quarterly, 2, 1-9.

Winkler, J., Shavelson, R. Stasz, C., Ro[illegible], A., & Feibel, W. (1984). How effective teachers use microcomputers for instruction. Paper presented at the meet--- of the American Educational Research Association, New Orleans, LA.

Woodcock, R.W. (1978). The Woodcock Reading Mastery Test. Teachings Resources Corp. Hingham, MA.

TE2


Evidence
Conclusion
1a. All men are human.
Human
Men
1b. Some men have brown hair.
Men
Brown Hair
1c. Some humans have brown hair.
or,
Some brown haired things are human.
Human
Brown Hair

Teaching Problem Solving Through Computer Simulations

John P. Woodward

Douglas W. Carnine

Russell M. Gersten

University of Oregon

American Educational Research Journal (in press)

Secondary students spend a considerable amount of their time completing application-oriented activities. In performing these tasks, students are asked to make a variety of inferences about a subject area by prudently using facts, concepts, and strategies or problem solving skills. Unfortunately, it is easier to just teach students rote memory information and procedural knowledge (i.e the literal algorithms used in solving a problem) than the comprehension and problem solving skills called for in many application items (Doyle,1983).

Some writers (Cherryholmes, 1966; Greenblat & Duke, 1975; Cruickshank & Tefler, 1980; Budoff, Thormann, & Gras, 1984) have suggested that one way to enhance these kinds of cognitive skills is through educational simulations. Simulations are thought to increase student participation (Boocock & Schild, 1968; Farran, 1968; Stembler, 1975) and allow low achieving students the much needed practice in applying what they've learned to new situations (Cohen & Bradley, 1977).

Yet the results of research on educational simulations, on the whole, have been discouraging. After an extensive review of studies conducted in the early sixties, Cherryholmes (1966) found that the effects of simulations were no greater than conventional instruction. Students neither learned more facts and concepts than they did in conventional instruction, nor did they show the anticipated increases in critical thinking and problem solving. A more recent review of simulation research by Pierfy (1977) reached conclusions similar to those of Cherryholmes. At best, the results have been mixed regarding the effects of educational simulations (McHenry, 1969; Livingston, 1971; Greenblat, 1973; DeNike, 1976; Pierfy, 1977; Jackson, 1979; Bredemeier & Greenblat, 1981; McKenzie & Padilla, 1984). Current research in computer simulations and other computer enrichment activities, although limited, also indicates little support for these techniques (Bangert-Drowns, Kulik, & Kulik, 1985; Waugh, 1986, Much of the research on simulations has been plagued by fundamental weaknesses in research design. Several interventions have been much too brief, usually with only one play of the simulation (Boocock & Schild, 1968; Fletcher, 1971; Pierfy, 1977). Quite a few studies (Livingston, 1970; Emery and Enger, 1972; Fennessey et al., 1973; Brenestuhl, 1975) used rather crude quasi-experimental designs involving intact classes being assigned to treatments on a non-random basis. In some cases, the intent of the simulation games, and hence the research hypotheses, were poorly formulated (Williams, 1980).

Finally, there are problems with criterion measures used in many simulation studies. Of the 22 studies reviewed by Pierfy (1977), virtually all of them were investigator-developed criterion measures, with very little detail about the construction of the measures or their reliability and validity. By failing to thoroughly analyze the instructional goals of simulations, some researchers did not design measures to capture everything taught by the games (Megarry, 1977).

The problems with simulation research, however, go well beyond research design and instrumentation problems. Fletcher (1971) noted that many of the simulations used in the research had never been field tested, and thus were of unknown quality. He cited the great variation in the quality of the games used (in terms of complexity, levels of sophistication, and interaction among participants) as a major source of the weak results reported.

The study reported here attempted to remedy many of the problems and issues cited above. The obvious problems regarding the length of intervention and random assignment were relatively easy to avoid. Also, validity and reliability issues of the measures used are addressed. We attempted to created an instrument that reflected the problem solving skills actually taught in the simulation.. The intent and instructional goals of the simulation Health Ways were formalized into a set of specific strategies, all monitored by a meta-strategy. Figure 1 represents the problem solving strategies used in the simulation.

Insert Figure 1 about here

The simulation itself was field tested in several junior high school classes and with students of varied abilities. Problematic features of the simulation were refined after each field test. Other components of the intervention, which will be discussed in the methods section, were pilot tested before the study.

The present study also attempted to address a curious feature of most past simulation studies. Virtually all studies have directly contrasted simulations to conventional teaching methods. Only on a few occasions have the combination of conventional instruction and a simulation been compared to a conventional method alone. This is understandable with non-computer simulations, as time considerations (i.e., the number of days or weeks needed to fully implement a simulation) usually prohibited a combined intervention.

With computer simulations, the situation is different. The effects of different variables can be demonstrated quickly (Doob, 1972), and less instructional time is required to demonstrates causal relationships and essential concepts. Unfortunately, very few experimental studies involving educational computer simulations have been documented. None had compared conventional instruction and a computer simulation with just conventional instruction.

A final aspect sets the present study apart from earlier research. Conventional instruction was delivered according to principles derived from recent teacher effectiveness research (cf. Brophy & Good, 1986). Teacher modeling, high rates of teacher-student interaction, guided practice, and structured seatwork were used in teaching basic facts and concepts to both groups in this study. This enabled us to better gauge the additional effects the simulation had on basic facts and concepts.

A secondary purpose of the study was to determine the extent to which a simulation could assist secondary learning disabled students in special education classrooms in acquiring factual information and problem solving skills. In addition to evaluating these students' performance in absolute terms, their performance was compared to that of their nonhandicapped age mates who were enrolled in health education courses.

Method

Subjects

All of the subjects were learning disabled high school students eligible for special education services by federal and local standards. Because of the reading requirements of the simulation, students who scored lower than the sixth grade reading level, as measured by the Metropolitan Achievement Test (1970), were excluded as subjects. The thirty students who met these criteria participated in the study; an equal number of students in three classes were randomly assigned to either the conventional or simulation condition.

Materials

Health Ways is a commercial software program developed by Carnine, Lang, and Wong (1985) for the Apple II and IBM PC computers. As a simulation, Health Ways requires learners to manipulate several variables (e.g., life threatening diseases, stress levels) in order to achieve an optimal life expectancy. Learners are presented a basic health profile which can be examined further by selecting various options. An example of such a profile is presented in Figure 2. For example, learners can inquire about eating habits by selecting the nutrition option. By doing this, they can see how much cholesterol, sugar, salts, etc., the profile character consumes. Learners must make changes in the most life threatening health habits (e.g., heavy smoking) and control other variables to be successful at a Health Ways game.

Insert Figure 2 about here

We also developed the Health Ways Supplementary Curriculum - an accompanying written curriculum (Woodward & Gurney, 1985) that extended information presented in Health Ways and the original Health Ways teachers guide. Information was taken directly from two widely used junior high school health textbooks. All of the information was rewritten to control for vocabulary and amount of new information. Clippings from newspapers, news magazines, journal articles, and health pamphlets were also used in the supplementary curriculum. Even though the supplementary materials relied heavily upon texts and other sources, considerable preparation time was required. Researchers estimate that it took approximately 70 hours to create the modified curriculum. The reading level of the curriculum, as determined by the Fry (1977) Readability Test, was approximately sixth grade.

Procedures

All students were instructed for 40 minutes per day for twelve days. The lesson consisted of two parts. The first part, called structured teaching, was identical for subjects in both conditions. A structured teaching method, following the model proposed by Rosenshine and Stevens (1984) and Brophy and Good (1986), was used for this segment.

Structured Teaching. Instruction was conducted in a large group of 12 to 15 students for this part of each lesson. Instruction began by reviewing essential information from previous lessons. Students were then presented with a list of vocabulary words, which were essential to the day's lesson (e.g., cholesterol, diabetes). For the next fifteen minutes, each student independently read the two to three pages of text for that day's lesson and answered written comprehension questions. The teacher then discussed the answers with the group and presented a series of review questions covering the main points of the lesson. At the end of the initial instruction, students separated into two groups: one which worked with the computer simulation (the Simulation Group) and one which worked on traditional application activities (the Conventional Group).

The Conventional Group worked in the resource room under the supervision of the resource room teacher, who presented these students with a variety of application or review activities. These exercises, typical of a high school health education class, were reviewed for representativeness by two health teachers at the high school where the study took place. For example, the Conventional Group students kept track of their diets for three days and analyzed their cholesterol levels. Other exercises included analyzing one paragraph profiles of different individuals and diagnosing poor health habits. Students completed the exercises during the last twenty minutes of the period.

Simulation Group students were taught in a computer lab, each student working individually at a microcomputer. Students in this condition worked on the Health Ways simulation for twenty minutes each day. Instruction over the twelve day period was broken into three phases: initial modeling of the simulation tutorial and one simulation game (three days), guided practice on three simulation games (two days), and independent practice with individual feedback from the instructor (seven days).

In the <u>initial modeling phase</u>, the instructor modeled each component of an effective strategy for playing the <u>Health Ways</u> games. The teacher modeled working on the most important health problems first, to control stress immediately if it rose to a level that was too high, and to be sure to maintain changes by using the maintenance option. Each component of th strategy was modeled in isolation and then integrated in later instruction. The most important instruction in the initial modeling phase involved the prioritizing of health problems. Students were first taught to look at current disease and hereditary information. If the profile indicated a heredity of hypertension, this meant that the student should first look at the individual's diet and check the level of salt consumption. If it was high, this led to a correlated change (i.e., an heredity of hypertension implies the need for a low salt diet, thus the student must change the character's salt intake). This habit was the first to be changed. The remaining habits were identified in their order of priority (i.e. the second and third most important habits).

In the guided practice phase of instruction, students were briefly taught in a group with one microcomputer. They were shown the initial screen in a Health Ways game profile, and individual students were asked to prioritize the first three health problems or habits and their correlated changes. Once correctly identified, the students were shown another profile. After three or four profiles, each student went to his or her computer in the lab and played the Health Ways games.

The independent practice phase allowed the students to play Health Ways games continuously for twenty minutes. The teacher circulated among the students, observing and commenting on a student's "play" of a game (i.e., how well he or she employed the strategy taught in the modeling phase of instruction).

All teaching was done by the researcher and a certified special education teacher. Assignment of teachers to treatment was counterbalanced, with the researcher and the teacher changing groups half way through the experiment. This was done to control for the effects of the teacher, a common problem that has flawed many computer based instruction studies (Bangert-Drowns, et al., 1985). The amount of total instructional time was controlled in this study; both groups received the same amount of teaching and independent work.

Measures

Students were assessed one day, two days, and two weeks following instruction. On the first day, acquisition of basic facts and concepts taught in the curriculum was measured by the Nutrition and Disease Test. The Nutrition and Disease Test was a 30 item test designed to measure students' retention of the important information contained in the written curriculum. Questions on the test were fill-in the blank, usually requiring only one or two word answers. The first 20 questions were solely from the written curriculum. The remaining 10 were questions over material that appears in both the written curriculum and the Health Ways simulation. Internal consistency reliability (coefficient alpha) of this measure is .84 based on a sample of 42 students. The Nutrition and Disease Test was given again two weeks after the instruction as a retention measure.

The second measure, the Health Diagnosis Test, was a set of three written profiles administered two days after the instruction. This test measured the student's ability to detect important health problems facing an individual, identify and change related health habits, and control stress as it increased due to the health changes. Central to the Health Diagnosis Test was prioritizing health problems. For example, the test measured the student's ability to not only identify, but order the health problems in terms of their importance to the individual's longevity. The Health Diagnosis Test has a test-retest reliability of .81. Table 1 presents an example of one of the test's three profiles.

Insert Table 1 about here

Scoring Procedures. For the Nutrition and Disease Test, only answers contained in the written curriculum or acceptable synonyms were considered correct. Subscale scores were obtained for two sections: a) those items reinforced and b) not reinforced by the Health Ways simulation.

Special procedures for scoring the Health Diagnosis Test were developed. Three different areas were assessed in the measure: a) identifying important health problems, regardless of order, and making the appropriate correlated change b) the ability to prioritize health problems and c) attending to stress when it was at a high level.

Current health facts and statistics were used to develop criteria for correct prioritizing. The criteria state that the learner should attend to current disease and hereditary information in determining which health habits are most detrimental, hence, which habits need to be changed first. A strict and a moderate criteria were used to measure students' ability to prioritize. To score at the strict criteria, a student must change the three most important health problems in a specific order (i.e. the habit associated with the current disease first, the one associated with the hereditary disease second, and the remaining detrimental habit third). These criteria were established by a committee consisting of the experimenter, a professional health educator, and two special education researchers. Students who scored at the moderate criteria simply changed, in any order, the habits associated with the current disease and hereditary problems within the first three changes.

In addition, both tests were also given to a random sample of non-handicapped high school students from health classes. Tests were given to tenth, eleventh and twelfth graders in introductory and advanced health classes. Their scores were compared with those of the two groups that participated in the study.

Results

The Nutrition and Disease Test

A 2 x 2 (treatment by time of test) analysis of variance with repeated measures on one factor was performed on the total number of correct responses on the Nutrition and Disease Test. Table 2 provides the descriptive statistics for the correct number of responses for the post and maintenance tests for each group. Means were also converted to a percent correct.

Insert Table 2 about here

The analysis shows a significant main effect for instructional method; $F_{(1,28)} = 5.30$, $p < .03$. For both treatment groups there was a significant drop in scores from post to maintenance test; $F_{(1,26)} = 16.23$, $p < .001$. No significant interaction was found. The simulation had a significant effect on mastery of key concepts in the unit; this effect was maintained over a two week period.

Subscales analyses. The 30 item test was broken into two subscales: a) items reinforced by the Health Ways simulation and b) items taught in the curriculum and not reinforced by the simulation. Separate 2 x 2 ANOVAs with repeated measures were performed on each subscale. The effect on items reinforced by Health Ways was significant; $F_{(1,28)} = 40.02$, $p < .01$. The effect for items not reinforced, however was nonsignificant; $F_{(1,28)} = 3.73$, $p < .06$. This demonstrates that the simulation was an effective vehicle for reviewing material that had already been presented in the written curriculum.

Health Diagnosis Test

Scores for the conventional and simulation groups were compared on a) the total test score and b) the total test score without stress as a factor. The reason for this was that the simulation group was explicitly taught the relationship between a health change and an increase in stress through Health Ways. Students in the conventional group were never taught about this relationship, thus it would be unlikely that these students would immediately control the stress level in the Diagnosis Test. Therefore, the factor of stress was removed from the analysis of the total test scores, which appears at the top of Table 3.

The three essential problem solving skills for the Health Diagnosis Test were independently compared: (1) prioritizing health habits (2) stress management and (3) identifying health problems and making correlated changes. The t-tests demonstrate a significant difference between the two groups according to all analyses demonstrating that the intervention had a consistent, significant effect on problem solving skills of the simulation students.

Insert Table 3 about here

The correlation between Metropolitan Achievement Test (MAT) Reading scores and total scores on the Health Diagnosis Test was non-significant (.12), much weaker than the correlation of .44 between MAT Reading and the Nutrition and Disease Test, a more conventional academic measure of facts and concepts. The non-significant correlation suggests no relationship between traditional academic measures (such as a standardized achievement test) and the problem solving skills measured by the Diagnosis Test.

In addition to the generally superior performance by the simulation students on the Diagnosis Test, there is some indication of a moderately strong relationship between their scores on this test and their understanding of the simulation. At the end of the third profile, the examiner asked each student to state his or her reasons for making the first, second, and third changes on the profile. In other words, the examiner asked why were the changes made in the order specified by the student. Responses to this question were categorized as 1) the student guessed (didn't know, didn't care, didn't know why) 2) the student was working on health problems but in no apparent order (i.e. no prioritizing strategy was used) or 3) the student worked on the most important health problem first (i.e. some kind of prioritizing strategy was used). The correlation between a student's score for this response and his or her total test score on the Diagnosis Test was .69. This suggests a moderately strong relationship between the strategies that students thought they were using in the test and those that they actually used.

Secondary Analyses: Comparison with Non-Handicapped High School Students

In a supplemental analysis, a one way analysis of variance (ANOVA) was used to compare the test performance of the conventional and simulation groups with non-handicapped students from regular health classes who did not participate in the study. The purpose of this quasi-experimental comparison was to extend the post test analysis to students of a comparable age group who were also receiving health instruction. Again, scores from each section of the Health Ways Nutrition and Disease Test and the Health Ways Diagnosis Test were analyzed. These results are presented in Table 4.

Insert Table 4 about here

Total score on the Diagnosis Test showed significant differences between the groups ($F_{(2,42)} = 27.36$ $p < .001$). A Tukey post-hoc comparison indicated significant differences favoring the handicapped simulation students in comparison to regular classroom students ($p < .01$). There was an equally significant difference favoring the regular classroom students over the learning disabled students in the conventional group ($p < .01$). The learning disabled students in the simulation group had problem solving skills on the health profiles superior to those of non-handicapped students in regular health classes. The non-handicapped students, in turn, out performed the handicapped students in the conventional group.

A significant difference also appeared between the groups on the reinforced items on the Nutrition and Disease Test, $F_{(2,42)} = 5.35$, $p < .01$. Tukey post-hoc comparison showed a significant difference between the special education simulation group and the two other groups ($p < .05$), favoring the handicapped students taught by Health Ways. Differences on the non-reinforced subscale items were non-significant.

Discussion

The Health Ways simulation was an effective tool in teaching material not easily taught by traditional means. In this study, we used direct instruction techniques in the initial phase of instruction to teach material rewritten from widely used health textbooks. The results indicate the use of computer simulations can effectively complement traditional instruction.

In addition to providing the experimental students with a simulation, we taught an explicit strategy that enabled them to be successful (see Figure 1). The results support the view that a structured approach in simulations, one where learners' tactics are specified and guided, does have significant educational effects.

The analyses indicated that the students in the simulation group performed at a significantly higher level on health fact and concept items that the simulation reviewed. The simulation also had a significant effect in developing problem solving skills in health. Simulation students were significantly better at prioritizing specific health habits, ones that needed to changed in the game's profile character in order to improve his health and longevity.

The superior performance by the learning disabled students in the simulation group over non-handicapped students from regular health classes suggests the extent to which explicit strategy instruction can be successful in teaching problem solving skills. The two non-handicapped students who had the highest scores on the Diagnosis Test articulated a prioritizing strategy comparable to that given by several of the special education students. Thus, many of the special education students in the simulation group showed a conscious awareness of the strategies that they were using, as did the two untaught, non-handicapped students, who may have achieved their awareness in a more intuitive manner.

Both the instruction in basic health concepts and the explicit strategy for the simulation were based on instructional design principles described in the Theory of Instruction (Engelmann & Carnine, 1982). Instruction began with models by the teacher of both successful and unsuccessful strategies that could be used with Health Ways. Next, students practiced the strategy over a range of profiles with feedback from the researchers. Gradually, the explicit reminders or prompts about steps in the strategy were removed, and the simulation students worked on Health Ways profiles independently.

The simulation itself was also designed to foster the acquisition of a strategy. The Health Ways simulation was preceded by a tutorial containing three simpler versions of the simulation profiles, each one slightly more complex than the preceding one. This gradual progression from simple to complex allowed aspects of the overall strategy to be introduced and practiced one at a time. The contribution of detailed, explicit strategy instruction (i.e., the methods described by Engelmann & Carnine, 1982; Palincsar & Brown, 1982) has been investigated with handicapped students in reading comprehension (Carnine & Kinder, 1985), content area instruction in science (Darch & Carnine, in press) and logical reasoning (Collins, 1984). Some research on computer simulations suggests that when contrary procedures are followed (e.g., informal instruction, particularly where clear strategies and corrective feedback are absent), student learning is insignificant (Waugh, 1986).

The present results indicate that simulations combined with instruction in strategies for successful use of the simulation can contribute to a student's learning of both factual information and problem solving skills. However, the results say nothing about the use of computer simulations as "stand alone" activities. Since the two treatments different in several respects, we cannot isolate a specific variable that accounted for the results. That 15 learning disabled students exhibited problem solving skills and that this was true for only two non-handicapped students underscores the potential of combining instruction and simulations. Future research could address the simulation alone by comparing Health Ways to Health Ways accompanied by explicit strategy instruction. These findings would help articulate the context in which computer simulations can be of the most benefit to students. A separate line of research will continue to evaluate the various components of strategy instruction, whether mediated by a teacher or by a computer.

References

Bangert-Drowns, R., Kulik, J., & Kulik, C. (1985). Effectiveness of computer-based education in secondary schools. Journal of Computer Based Instruction, 12(3), 59-68.

Boocock, S. & Schild, E. (Eds.). (1968). Simulation Games in Learning. Beverly Hills, CA: Sage.

Bredemeier, M. & Greenblat, C. (1981). The educational effecti eness of simulation games: A synthesis of findings. Simulation & Games, 12, 307-332.

Brenestuhl D. (1975). Cognitive vs. affective gains in computer simulation. Simulations & Games, 6, 303-311.

Brophy, J. & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.) Third handbook of research on teaching. New York: MacMillan.

Brophy, J. & Evertson, C. (1976). Learning from teaching: A developmental perspective. Boston: Allyn and Bacon.

Budoff, M., Thormann, J., & Gras, A. (1984). Microcomputers and speical education. Cambridge, MA: Brookline Books

Carnine, D., & Kinder, D. (1985). Teaching low performing students to apply generative and schema strategies to narrative and expository material. Remedial and Sepcial Education, 6(1), 20-30.

Carnine, D., Lang, D., & Wong, L. (1985). Health Ways. Unpublished Computer Program, Eugene, OR.

Cherryholmes, C. H. (1966). Some current research on effectiveness of educational simulations: Implications for alternative strategies. American Behavioral Scientist, 10, 4-7.

Cohen, R. & Bradley, R. (1978). The use of simulation games to create real worlds for special education students. Reading Improvement, 15, 275-278.

Collins, M. (1984) The effectivenes of computer-delivered correction procedures on low-performing secondary students' reasoning skills. Unpublished docotroal dissertation, University of Oregon.

Cruickshank, D., & Telfer, R. (1980). Classroom games and simulations. Theory into Practice, 1980, 19, 75-80.

Darch, C., Carnine, D., & Kameenui, E. (in press). The role of visual - displays and social structure in content-area instruction.

DeNike, L. (1976). An exploratory study of the relationship of educational cognitive style to learning from simulation games. Simulation & Games, 7, 65-74.

Doob, P. (1972). Prospects for simulation gaming in health planning and consumer health education. ERIC ED 070 596.

Doyle, W. (1983). Academic work. Review of Educational Research, 53, 159-199.

Emery, E. & Enger, T. (1972). Computer gaming and learning in an introductory economics course. Journal of Economics Education, 3, 77-85.

Engelmann, S. & Carnine, D. (1982). Theory of Instruction: Principles and applications. New York: Irvington.

Farran, D. (1968). Competition and Learning for Underachievers. In S. Boocock and E. Schild (Eds.). Simulation Games in Learning. Beverly Hills, CA: Sage.

Fennessey, G., Livingston, S., Edwards, K., Kidder, S., & Nafziger, A. (1975). Simulation, gaming, and conventional instruction: An experimental comparison. Simulation & Games, 6, 288-302.

Fletcher, J. (1971). The effectiveness of simulation games as learning environments. Simulation & Games, 2, 425-554.

Fry, E. (1977). Fry's Readability Scale. Providence, RI: Jamestown Publishers.

Greenblat, C. (1973). Teaching with simulation games: A review of claims and evidence. Teaching Sociology, 1, 62-83.

Greenblat, C., & Duke, R. (1975). Gaming simulation: Rationale, design, and application. New York: Halsted.

Jackson, M. (1979). An antipodean evaluation of simulation in teaching. Simulation & Games, 10, 99-138.

Livingston, S. (1970). Simulation Games and Attitude Changes: Attitudes Toward the Poor. Baltimore: The Center for the Study of Social Organization of Schools, Johns Hopkins University (Report No. 63).

McHenry, W. (1969). A study of the use of the Life Career Guidance Game in junior high school group guidance. Unpublished doctoral dissertation, George Washington University.

McKenzie, D. & Padilla, M. (1984). Effects of laboratory activities and written simulations on the acquisition of graphing skills by eighth grade students. Paper presented at the annual meeting of the National Association of Science Teaching, New Orleans, April, 1984.

Palincsar, B. & Brown, A. (1982). Inducing strategic learning from texts by means of informed, self-control training. Topics in Learning and Learning Disabilities, 2(1), 1-17.

Pierfy, D. (1977). Comparative simulation game research: Stumbling blocks and stepping stones. Simulation & Games, 8, 255-268.

Rosenshine, B. & Stevens, R. (1984). Classroom instruction in reading. In D. Pearson (Ed.). Handbook of research on teaching (pp 745-798). New York: Longman.

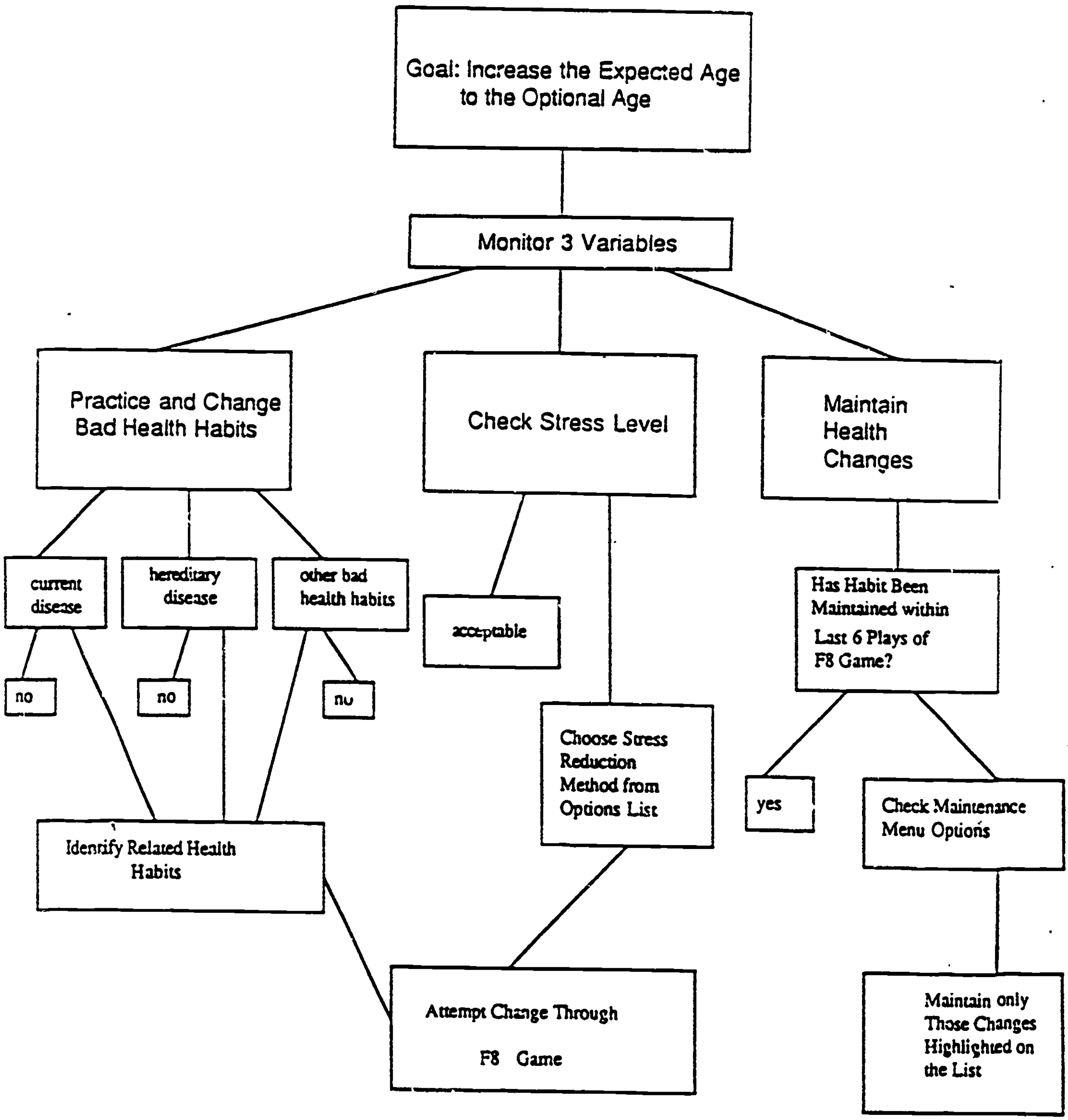
Stembler, W. (1975). Cognitive effects of a programmed simulation. Simulation & Games, 6, 392-403.

Waugh, M. (1986). The effect of teacher involvement on student performance in a computer-based science simulation. Paper presented at the National Education Computing Conference, San Diego, June, 1986.

Williams, R. (1980). Attitude change and simulation games. Simulation & Games, 11, 177-196.

Woodward, J., & Gurney, D. (1985) Health Ways Supplementary Curriculum. Unpublished curriculum, Eugene, OR.

Figure 1



The Healu Ways Problem Solving Strategies: Move Left to Right
Move Down the Branch if Conditions Warrant it
Return Up to Main Node
When All the Way to the Right, Return Left Again

Figure 2

Health Ways Simulation Profile

Ages:	Today's Year	Week	Day	Expected	Winning
	50	00	0	65	75

Will Power = 180 Stress = 35

Name: James
Heredity: Lung cancer
Diseases: No current ailments

A) Weight 20 pounds overweight
B) Tobacco Moderate smoker
C) Alcohol Non-drinker
D) Exercise Moderate 5 times/week
E) Nutrition See Sub-menu
F) Lifestyle See Sub-menu
G) Maintenance Menu
H) Stress Reduction Menu
I Help

Table 1

Health Facts on Julie Hill

Age: 34

<u>Heredity: diabetes Current Diseases: liver disease</u>

1. Weight and diet: 7 pounds overweight

   a. Eats breakfasts
   b. Eats a lot of food with cholesterol
   c. Eats a lot of empty calorie sweets
   d. Eats very little food with sodium
   e. Drinks very few beverages with caffeine
   f. Eats a lot of food with fiber
   g. Eats balanced meals

2. Tobacco: non smoker

3. Alcohol: light drinker (3 drinks a week)

4. Exercise: exercises 5 times each week

5. Stress: average stress in her life

Table 2

Means (M), Medians (Mdn), and Standard Deviations (SD) of Number of Total Correct Answers on the Nutrition and Disease Test by Instructional Group

Instructional Group	Post Test					Maintenance Test				
	N	M	Mdn	SD	Mean % Correct	N	M	Mdn	SD	Mean % Correct
Simulation	15	22.00	21.5	3.72	73.3	15	19.97	20.5	5.08	66.5.
Conventional	15	17.93	18.5	5.86	59.7	15	15.47	16.5	5.44	51.6

Table 3

Summary of t-Tests for the Diagnosis Test

	Simulation		Conventional				
	M	SD	M	SD	t	df	p
Total Test Score	27.7	6.2	12.47	4.9	7.52	28	<.001
Component Problems Solving Skills Involved in the Test							
Prioritizing Alone	8.4	4.7	1.47	1.9	5.27	28	<.001
Stress Management	4.9	1.8	1.73	2.3	4.27	28	<.001
Identifying Health Problems & Making Correlated Changes	14.4	2.1	9.27	2.8	5.66	28	<.001

Table 4

Means (M) and Standard Deviations (SD) for Handicapped and Non-Handicapped Students on the Two Academic Measures

	N	M	SD
Nutrition and Disease Test: <u>Total Score</u>			
Mildly Handcapped Students Taught By:			
Simulation	15	22.00	3.72
Conventional	15	17.93	5.86
Non-Handcapped Students:	15	19.47	4.94
Nutrition and Disease Test: Items Reinforced by <u>Health Ways</u>			
Mildly Handicapped Students Taught By:			
Simulation	15	7.33	1.35
Conventional	[illegible]	5.60	2 20
Non-Handicapped Students:	15	5.53	1.46
Health Ways Diagnosis Test: <u>Total Score</u>			
Mildly Handicapped Students:			
Simulation	15	27.73	5.89
Conventional	15	12.47	4.88
Non-Handicapped Students:	15	18.07	6.03

The Effectiveness of Videodisc Instruction in Teaching Fractions to Learning-Disabled and Remedial High School Students[2]

BERNADETTE KELLY[1]

DOUGLAS CARNINE

RUSSELL GERSTEN

and

BONNIE GROSSEN

University of Oregon

Abstract

This study compares the effectiveness of a videodisc curriculum that incorporates principles of instructional design (including discrimination practice and cumulative review) with a traditional basal program designed to teach basic fractions skills. Twenty-eight high school students, including 17 mildly handicapped students, qualified for the study by showing (a) mastery of whole number operations and (b) less than 50 percent mastery of the fractions skills to be taught. The students were matched in pairs based on a pretest score and math scores from the California Achievement Test, and then randomly assigned to one of the treatments. During the ten-day intervention, observers collected data on levels of treatment implementation and student on-task behavior. A criterion-referenced posttest and two-week maintenance test were administered. The videodisc curriculum resulted in significantly higher posttest and maintenance test scores. Levels of on-task behavior were significantly higher in the videodisc sessions, although levels in both conditions were above 80 percent. An analysis of student error patterns indicated that differences in instructional design features contributed to the relative effectiveness of the two curricula.

The National Assessment of Educational Progress reported that, nationally, "performance of fractions computation is low, and students seem to have done their computation with little understanding" (Lindquist, Carpenter, Silver & Matthews, 1983, p. 16). For example, the assessment found that only one-third of U.S. seventh-graders can add 1/2 and 1/3. The problem is even more pronounced for handicapped students.

[1]Requests for reprints should be addressed to the first author.

[2]The research was supported in part by Grant No. G008400660-02 from the U.S. Department of Education

Research on effective instructional practices with special education students gives some indication of how to improve instruction. Englert (1984) measured mildly handicapped students' growth on a range of basic skills measures and correlated this growth with observed teacher performance. More effective teachers (classified on the basis of high student academic gain) provided appropriate academic feedback to student errors more frequently than did less effective teachers. The more effective teachers also maintained higher pacing and student success rates throughout each lesson. This set of variables has been found to be effective with low performing students in regular classroom settings (Good & Grouws, 1979; Gersten, Carnine & Williams, 1982; Rosenshine, 1983).

However, improved teacher training and improved teacher presentation techniques may not be enough. The curriculum itself is being called into question. Ten years ago, a report from the National Council of Teachers of Mathematics (Carpenter, Coburn, Reys & Wilson, 1976) delineated problems in the way fractions were taught in conventional curricula. More recently, the California Department of Education declared that all 14 of the major textbooks were deficient in their treatment of fractions, decimals and problem solving.

Clearly, empirical investigations of the instructional design of a curricula in an area such as teaching fractions needs careful investigation. This study compares two curricula designed to teach basic fractions skills—a traditional basal curriculum, and an innovative curriculum based on Engelmann & Carnine's (1982) theory of instructional design. The basal program was selected from the four most widely adopted texts in the United States. It features the following instructional design components: clearly stated objectives, practice examples relating directly to lesson objectives, review tests with remediation pages specified, reteaching worksheets, and the provision of step-by-step strategies. It also includes practice with concrete objects (manipulatives), a feature recommended by many math educators. We believe it represents one of the better basal texts available.

The *Mastering Fractions* program (Systems Impact, 1985) has been developed from a research based theory of instruction (Engelmann & Carnine, 1982) and incorporates sophisticated principles of instructional design. It uses interactive videodisc to replace conventional text, although worksheets use conventional print formats. The primary question under investigation is whether the *instructional design features* incorporated into the videodisc program can increase student performance. It may be argued that the videodisc medium may contribute to increased student performance. While the medium can *facilitate* the implementation of effective instructional procedures, there is evidence that the instructional design features incorporated into a curriculum result in higher student performance, rather than the instructional medium per se.

A recently-conducted study (Hasselbring, Sherwood, & Bransford, 1986), compared the effectiveness of the *Mastering Fractions* program with a curriculum incorporating the same instructional design features, *but not involving the videodisc medium*. Instead, teachers presented the examples and exercises from the *Mastering Fractions* curriculum on an overhead projector. They found no difference between students' scores on a criterion-referenced posttest between these two conditions. In other words, the medium of instruction appeared to have no effect. The third experimental condition utilized the fractions curriculum currently adopted by the school district. Student performance in both *Mastering Fractions* conditions (with and without the videodisc) were significantly higher than student performance in the basal condition. This finding suggests that the difference in performance can be attributed to the instructional content of the *Mastering Fractions* program, and that any novelty effects produced by the videodisc medium are not significant.

In a review of research evaluating a range of instructional media, Clark (1983) argued that instructional technologies are "mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes change in our nutrition" (p. 445). Clark recommended that future research focus on curriculum design rather than medium of delivery, as the curriculum design seemed to be a major variable in determining the effectiveness of an instructional program.

Comparison Between the Videodisc and Basal Curricula

In order to elucidate the principles of instructional design incorporated in the experimental curriculum, this section compares the *Mastering Fractions* program and the basal program used in the current investigation. The section covers the following curriculum dimensions: review procedures, discrimination practice, example selection, and explicit strategy teaching.

Review Procedures

In the basal program, a skill is introduced and practiced but then "disappears" for several days. For example, *Mathematics Today* teaches multiplication of fractions in one lesson. In subsequent lessons, other skills are introduced, including multiplication of whole numbers and fractions, and multiplication of mixed numbers. However, in the next three lessons students work with word problems, reciprocals and division, after which students are expected to perform the multiplication of fractions independently on review and test lessons.

In *Mastering Fractions*, the skill of multiplying fractions is introduced and then practiced on *every subsequent lesson in the program*. Each new skill that is taught is reviewed cumulatively, or else incorporated into more complex skills.

Discrimination Practice

Students who learn to carry out certain steps again and again on the same type of problem may have difficulties when they encounter different problem types mixed together on a test. For example, a 14-day unit in the basal program introduces adding and subtracting fractions. In the next unit, students learn the strategies for multiplying and dividing fractions. No practice is given on *discrimination* between the strategies (e.g., multiplication and addition). In the review and test lessons, the problem types are still separated. Students never receive discrimination practice between strategies. After the two units, fractions operations do not appear again in the text for the remainder of the school year.

In *Mastering Fractions*, a skill is introduced, practiced, and within a few lessons mixed with other types of problems. For example, exercises in the lesson presentation specifically address the differences between addition and multiplication strategies. If students have difficulty making the discrimination, specific remediation is given, after which students are required to work a set of problems involving both operations. The skills are then integrated with other types of problems on every worksheet.

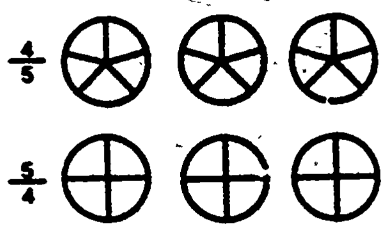
Darch, Carnine and Gersten (1984) compared the effectiveness of a regular basal mathematics curriculum with a curriculum program similar to *Mastering Fractions* in that it incorporated systematic discrimination practice. Students who received discrimination practice performed significantly better than students who did not, on a criterion-referenced posttest and maintenance test. Englert (1984) also emphasizes the importance of discrimination practice for mildly handicapped students, to avoid confusion between related concepts.

Example Selection

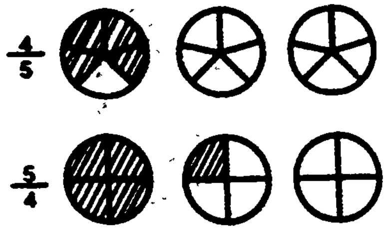
Range of Examples. In the basal program, when students first encounter pictures of fractions, all examples are less than one. In the next grade level, mixed numbers are introduced as a whole number and a fraction, reinforcing the misconception that fractions can only represent qualities less than one. Improper fractions do not appear until the next grade level. A common error occurs when improper fractions are finally introduced; students represent these fractions as less than one; e.g., for 5/4 students write:

Mastering Fractions teaches students a strategy for reading and writing both proper and improper fractions from the beginning of the program:

(a) The denominator tells the number of parts in each unit:



(b) The numerator tells the number of parts used or shaded:



The wide range of examples prevents students from forming misconceptions and gives students a more complete understanding of what a fraction represents.

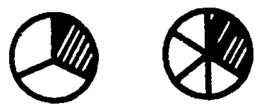
In a carefully controlled experiment, Carnine (1980) demonstrated how a limited range of examples can cause students to form misconceptions. The instructional task was to write hundredths fractions as decimals. One group of students was presented with a wide range of examples, with numerators of one, two or three digits (e.g., 185/100, 2/100, 75/100). The other group was presented with a limited range of examples; all numerators comprised of two digits (e.g., 28/100, 84/100, 55/100). Carnine hypothesized that students in the limited range group would learn the misconception that the decimal point is always placed directly in front of the digits in the numerator (i.e., 4/100 = .4, 185/100 = .185). His prediction that these students would not be able to generalize to other examples was verified. Students in the limited range groups scored 0% and 7% respectively on the problem types X/100, XXX/100 on the immediate posttest. Students who had received the full range of examples scored 89% and 93% respectively.

Easily Confused Labels. When highly similar terms (e.g., the terms numerator and denominator) are introduced at the same time, there is an increased likelihood that students will become confused. In the basal program, the terms numerator and denominator were introduced in the same lesson. In subsequent fraction examples, the teacher referred to the terms numerator and denominator, and the labels appeared on some worksheets, but no systematic teaching ensured that students could successfully apply the labels to the appropriate parts of a fraction.

In the *Mastering Fractions* program, the terms numerator and denominator were separated by several lessons, so that students were facile with one label before the other, similar label was introduced. This procedure decreases the likelihood that students will become confused and make reversals.

Explicit Strategy Teaching

In the basal program, students are not always given an explicit strategy to solve a problem. This could lead to student misunderstandings. Equivalent fractions serve as an example. In the first set of basal exercises, pictures of the two equivalent fractions, and three of the four fraction numbers are given; the students count the number of shaded parts to complete the problem:



$$\frac{1}{3} = \frac{?}{6}$$

Students can write the fourth number and complete the equation without understanding equivalent fractions. The students count the shaded parts and write the numerator. In the final set of exercises given that day, the pictures are removed.

$$\frac{3}{4} = \frac{?}{8}$$

The student workbook states, "You may draw a picture to help you." At least some students will not be sure how many parts to draw or shade; unless, of course, they already know how to write 3/4 as 6/8.

In *Mastering Fractions*, the strategy for equivalent fractions emphasizes this rule: when you multiply by one you don't change the value. When a fraction is multiplied by a fraction equal to one, the original fraction is equivalent to the new fraction; i.e.,

$$\frac{1}{2} \times 1 = \frac{1}{2}$$

$$\frac{1}{2} \times \frac{4}{4} = \frac{4}{8}$$

so, $$\frac{1}{2} = \frac{4}{8}$$

With this conceptual basis for equivalent fractions, students are introduced to the strategy for determining the missing number when given a problem; e.g., 2/3 = ?/6. First, students identify the fraction of one they multiply 2/3 by, so as to end up with 6ths in the denominator. The denominator of the fraction inside the parenthesis is 2, so the fraction equal to one is 2/2: 2/3 × (2/2) = ?/6. Thus, the missing numerator is 4: 2/3 = 4/6.

Kameenui, Carnine, Darch and Stein (1986) compared a basal approach to introducing fractions with a strategy-based approach similar to that found in the *Mastering Fractions* curriculum. For the explicit rule-based strategy group, the teacher demonstrated concepts and skills in a step-by-step fashion. Teacher guidance was gradually and systematically faded until students were performing independently. Correction procedures directed students to the explicit instruction they had received. In contrast, the basal approach was much less structured. Emphasis was placed on activities using student discussion and the use of manipulatives. Students in the explicit strategy group performed significantly higher on a criterion-referenced posttest and on a transfer test of related fractions skills.

Method

A study was conducted to determine whether the instructional design features incorporated into the videodisc program would have a significant effect on student performance. The study compared the relative effects of *Mastering Fractions* and a traditional basal program on student acquisition of skills in a unit on fractions. Classroom behaviors known to be correlates of learning (academic engagement and success rate during the lesson) were also measured and an analysis of students' error patterns was made. Student attitudes were also assessed, and information on obtained levels of implementation were recorded.

Many factors that research has identified as components of effective teaching (Brophy & Good, 1986) were constant across the two conditions: e.g., daily feedback on independent assignments, guided practice prior to independent work, uninterrupted successful practice, and quantity of review. The major differences between the two conditions were the range and sequencing of the instructional examples.

Subjects

Prior to training, subjects from two high school math classes were screened for: (a) mastery of the preskills necessary to learn basic fraction concepts and operations, and (b) prior knowledge of the specific skills to be taught. Two classes, both containing mainstreamed students, participated in the study. One of the classes was a remedial math class containing 22 students. Eleven were classified as learning-disabled (LD) ninth and tenth graders; the remaining 11 students were not classified as handicapped. The other general math class contained 12 ninth graders in need of remedial math, along with six ninth-, tenth-, and eleventh-grade LD students. Students were classified as LD based on the State of Oregon recommendations. That is, any student scoring more than three years below grade level on two different standardized tests in the same skill area qualified for an individualized education program, and met the criterion for the Learning Disabled classification.

In each classroom, qualifying students were randomly assigned to the basal text (BT) or interactive videodisc (IV) treatment. This resulted in four instructional groups. In the remedial class, nine students were assigned to each treatment. In the general math class, eight students were assigned to each treatment.

Out of 34 subjects, only 28 completed the study and took the posttest; 26 students took the maintenance test. Subject attrition resulted from a variety of sources; five students were absent for more than 50% of the intervention days, and one student had recently arrived from Asia with insufficient language skills to benefit from the instruction. One student could not take the maintenance test as he was in a detention center at the time of test administration. Another student was found to have cheated on the test, so his score was not included in the maintenance test data. Table 1 shows the number of subjects qualifying in each group who completed the study.

Table 1
Distribution of Subjects in the Four Instructional Groups[a]

	Class		
Treatment	Remedial Math	General Math	Total
Basal	8	8	16
Videodisc	7	5	12

[a]Five students were dropped due to excessive absence.

Measures

Preskills Screening Test

A screening test, developed by the experimenter, was administered to ensure that students had mastered the requisite whole number skills for a unit in fractions (i.e., facility with basic addition, subtraction, and multiplication facts). The first part of the test comprised ten of the more difficult facts. All students who were tested achieved at least 80% and were eligible for the study based on this criterion.

The second part of the screening test was criterion-referenced to the skills to be taught in the fractions unit. Students who scored above 50% on this part were ineligible for the study. Ten students were excluded based on this criterion. Eligible subjects were grouped in pairs, matched on total math scores from the California Achievement Test (CAT) and on pretest scores. Individual students within each pair were then randomly assigned to the two treatment conditions. The mean scores on a 6-item pretest for the videodisc and basal groups were 2.4 (SD = 1.16) and 2.1 (SD = .90), respectively. There were no significant differences between the groups. Mean scaled scores (expanded standard scores) on the CAT were 511 for the videodisc students and 504 for the basal students. Standard deviations were 56.25 and 60.6, respectively Again, these differences were not significant. The mean scale scores corresponded to percentile score equivalents of 29th percentile and 26th percentile respectively. (CAT scores were not available for all students; $N = 10$ for the IV group, $N = 14$ for the BT group.)

Measures of Achievement

The principal measure for the study was a criterion-referenced test (CRT) developed by the experimenter. Two parallel forms were developed as a posttest and a two-week maintenance test The test included the following skills, taught in both the IV and BT conditions: writing fractions from pictures, vocabulary (e.g., denominator), addition and subtraction of fractions with like denominators, multiplication of fractions, and multiplication of a fraction and a whole number.

Reliability and Item Analysis. Field test versions of the CRT were given to thirty fourth- and fifth-graders who had had some fractions instruction. Internal consistency reliability was assessed for each form; coefficient alpha reliability was .98 for post and .98 for maintenance. Alternate form reliability was also evaluated; the Pearson correlation coefficient between the two forms was .96

Measures of Classroom Variables

Two classroom variables associated with higher student achievement are (a) total time students are *actively engaged* in instructional activities; time *'on task'* (Rosenshine, 1983) and (b) student success rate while doing independent seatwork (Fisher et al., 1980).

Active Engagement. An observational recording form was designed to measure the extent to which students were actively engaged during instruction. Each group of students was observed either three or four times during the study. Student behaviors were recorded with a six-second momentary time sampling procedure. 'On-task' behaviors included answering questions, writing, and watching the teacher or the monitor during the lesson presentation. Behaviors recorded as 'off-task' included gazing out of the window, sleeping, or chatting to another student. Other behaviors (e.g., passing out papers, waiting for teacher assistance) were recorded as transitional activities. Observers recorded all on-task behaviors as plus (+), all off-task behaviors as a minus (-), and all transitional activities as a zero. Thus it was possible to determine the proportion of instructional time spent in each of the three behavior categories for each of the two conditions.

Success Rate. Students' independent seatwork was collected at the end of four observation lessons. The percent of problems attempted and the percent that were successfully completed was calculated.

Measures of Implementation

Implementation checklists were used to identify those elements of the teaching models that were consistently implemented and those implemented at lower levels. The checklists were similar to the form developed by Good, Grouws & Ebmeier (1983). All items on the checklists were operationally defined. Below are two sample items that were applicable to both instructional models.

	YES	NO	NA
i) Did the teacher award points for independent work done on the previous day?			
ii) Did the teacher circulate during independent work reinforcing appropriate behavior?			

Items relating specifically to the IV model (e.g., whether the teacher checked student performance at the specified points in the lesson, or administered a daily review quiz) were developed using the videodisc teacher's guide. Items applicable only to the basal text method (e.g., whether the teacher provided an opportunity to use manipulatives, or whether the teacher supplied examples in addition to those presented in the text) were developed using the basal text teacher presentation book. Each item scored in the 'yes' category by the observer was tallied, and the percent of total checks possible was calculated for each lesson observed.

Measures of Student Attitudes

A questionnaire was developed, based on the work of Fennema and Sherman (1976). Students were asked their opinion on a 3-point scale in response to a series of statements that related to students' evaluation of their math ability and of the relevance of fractions for daily life. For instance,

1. I think I could handle more difficult fractions.
2. Learning fractions is a waste of time.

Items were read to students one at a time and the question asked, "Is this true for you?" Students responded to each item with: Yes, No or Don't Know.

Materials

Interactive Videodisc

The materials required for implementation of the IV fractions curriculum were: a videodisc player, the videodiscs, consumable student worksheets and teacher answer keys.

Lesson Format. Each videodisc lesson took approximately 30 minutes to complete. Lessons typically began with a brief quiz covering the essential skills introduced in the previous lesson. The lesson presentation followed next—an explanation followed by written problems for each of several skills. After completing the lesson, students were assigned independent problems for seatwork. The worksheets comprised 20 to 40 items, including a variety of skills that students had learned thus far.

Unit Mastery Tests. In the IV curriculum, every fifth lesson was a test. Teachers used the tests to determine whether a review of particular skills was necessary from any of the four lessons preceding the test lesson.

Basal Text

The materials required for implementation of the BT fractions curriculum were: a teacher presentation book (with answer keys), student textbooks and consumable worksheets. In some lessons, manipulatives were also used, e.g., paper strips or fraction pie models.

Lesson Format. Each 30-minute lesson was designed to teach a single objective. Each lesson began with an introduction, in which the teacher used discussion and demonstrations to develop ideas. Next, the teacher guided students through several examples in the student textbook before assigning in-class problems. After completing the lesson, follow-up activities, usually involving manipulatives, were used to consolidate the concept developed in the lesson. Students were then assigned independent problems for seatwork. The worksheets comprised 20-40 items focusing on the student objective introduced that day.

Review and Unit Tests. Review tests were provided at the end of the unit, sampling each of the major skills and concepts that had been introduced. Teachers used the results of the review test to reteach concepts and skills that students had not mastered. The unit test was presented the next day. The review and unit tests sampled the same skills in the same order, and had a standardized test format.

Procedures

Teachers were the experimenter and a research assistant from the University of Oregon. Each teacher taught one condition for one-half of the study, then changed conditions for the remainder of the study, to minimize teacher-student effects.

Monitoring Implementation

The teachers were observed on four occasions to assess the level of implementation in each classroom. Teachers received specific feedback on their performance, using the Implementation Checklist (discussed under **Measures**). Throughout the study, teachers discussed any problems associated with the implementation of the two approaches.

Observers

Two trained observers recorded students' time on-task and percent correct responses on independent worksheets, on three or four occasions for each group of students. Before collecting the experimental data, the observers practiced using the instruments until inter-observer reliability exceeded 85 percent.

Administration of Measures

Criterion-referenced tests were administered to all students participating in the study immediately following the completion of the unit (posttest), and two weeks after completion of the unit (maintenance test).

Students' on-task behavior and success rate, and levels of implementation were measured on the second, fourth, seventh and ninth days of the intervention. The experimenter conducted student attitude surveys before and after completion of the study.

Results

The primary dependent variable was student performance on the 12-item criterion referenced test (post and maintenance). A 2 × 2 analysis of variance (ANOVA) was performed on the CRT scores. The between-subjects factor was the instructional method (videodisc vs. basal text); the within-subjects (repeated) factor was the time of test (post and maintenance). Significant main effects were found for the instructional method [$F = 17.28, p < .001$] and for the time of test [$F =$ 4.53, $p < .05$]. Figure 1 graphically depicts the mean scores for students in each condition on the posttest and maintenance test. Table 2 presents the descriptive statistics for the comparison

Students in the videodisc and basal conditions were on-task 96% and 84% respectively of the total instructional time during observation periods. A Mann Whitney U Test indicated a significant difference between the two conditions ($U = 3.5, p < .005$). Students' performance on independent seatwork was 96% correct for the BT group and 91% correct for the IV group.

Levels of implementation were extremely high in both conditions; 93% of the possible implementation behaviors were observed in the BT condition, and 92% in the IV condition. The item that was weakest in both conditions referred to the teacher awarding points for the previous day's independent work.

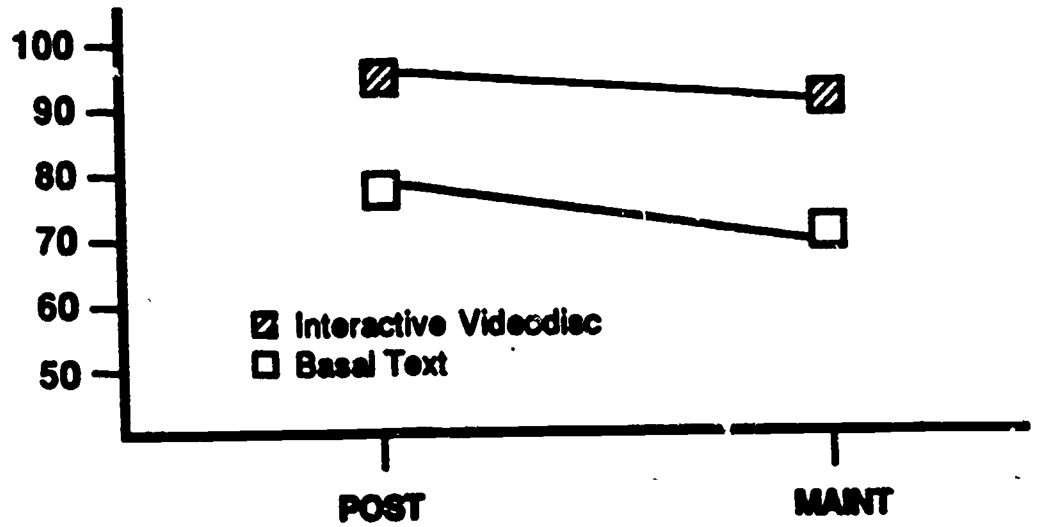


Figure 1. Mean Percent Correct Scores on the Post and Maintenance Tests for Interactive Videodisc and Basal Treatments.

Table 2
Means, Standard Deviation, and Mean Percent Correct Scores on the Posttest and Maintenance Test

	Posttest				Maintenance Test			
Instructional Method	N	M	SD	M%	N	M	SD	M%
Interactive Videodisc	12	11.4	1.0	95.1	12	11.3	1.4	93.8
Basal Text	14	9.5	2.3	79.1	14	8.4	1.9	70.2

Responses from the student questionnaires were summarized and assigned a score ranging from −1 (all negative responses) to +1 (all positive responses) for the students' perception of (a) their competence in working with fractions, and (b) the relevance of fractions for daily life. Students in both conditions made similar gains in perceived competence and relevance. The results are summarized in Table 3.

An analysis of covariance was performed, with pretest scores serving as the covariate. There were no significant differences between the two instructional groups for either type of question (ability or relevance). This includes no significant difference in the change in attitude between the two groups. Both groups showed growth in perceived competence. This parallels the observed growth in skills as measured on the criterion-referenced test. Only trivial increase in perceived relevance were found.

Table 3

Mean Responses on Pre- and Post-Study Attitude Measures of Perceived Competence In, and Relevance of Fractions

Type of Question	Pre[a]	Post[a]	Type of Question	Pre[a]	Post[a]
Competence			Relevance		
IV	.10	.81	IV	.48	.57
BT	-.21	.54	BT	.15	.36

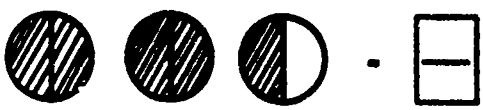
[a]Scores range from -1 (very negative) to +1 (very positive).

Discussion

The results of the experiment suggest that the different instructional design features in the two curricula produced different levels of student mastery of the content covered. The students receiving the videodisc curriculum scored significantly higher, both on the criterion-referenced posttest and on the maintenance test. The videodisc scores also dropped less dramatically over time—a nonsignificant drop of 1% compared to a drop of 7% for the basal text students.

While a significant difference was found between the two conditions for students' on-task behaviors, it should be noted that levels of on-task were high in both conditions. Students receiving the basal lessons were well motivated and actively involved during the lessons. Similarly, student success rates on independent worksheets were very high in both conditions: 96% and 91% correct for the basal students and videodisc students, respectively. These findings imply that the instructional quality of the IV curriculum (rather than other process variables) was largely responsible for the differences in student performance.

Patterns of student errors also confirm the importance of the specific differences between the programs. For example, a large proportion (75%) of students in the basal treatment made errors when asked to write the fraction for a diagram representing a fraction greater than one. Given the diagram



56% of the basal students wrote 5/6, even though all students could identify

as 1/2. The inability of 75% basal text students to *extrapolate to* fractions greater than one is a predictable consequence of all examples being less than or equal to one during the treatment intervention. In contrast, only 8% of the videodisc students, who had been exposed to fractions greater than one, exhibited this error on the posttest. This parallels the results of the Carnine study (1980) cited earlier. For further discussion on the relationship between student error patterns and instructional design, see Kelly (1986).

Videodisc Technology

The potential of videodisc technology in the classroom lies in its ability to assist the teacher in the consistent implementation of sound instructional procedures. The *Mastering Fractions* program takes advantage of the videodisc medium to demonstrate concepts clearly. For example, when equivalent fractions are taught, students see a fraction on a balance beam. The side with the fraction tips down. When an equivalent fraction is placed on the other side, the balance becomes level. The video sequence shows what equality means in a vivid, compelling manner.

These procedures would be extremely labor intensive to achieve using more traditional methods. For example, in the Hasselbring et al. study (1986) cited earlier, the preparation and organization of materials in the condition emulating the *Mastering Fractions* program necessitated the employment of a half-time teacher's aide.

A well designed program can also improve the quality of instruction provided by less confident teachers. The videodisc program can provide clear initial demonstrations and also provide frequent checks on student performance, helping teachers to diagnose student errors. Appropriate remediation procedures can also be specified in a disc program, providing the teacher with a means for giving students immediate corrective feedback. In addition, the videodisc presentation frees the teacher from demonstrating at the front of the classroom and enables the teacher to move among the students, monitoring their performance.

The most obvious disadvantage of the videodisc medium—as with any new technology—is the cost. However, the cost of hardware has already dropped substantially. The discs are also highly durable. Surface scratches do not hinder the video or sound quality when the disc is played. The quality of the disc does not deteriorate over time, since the laser beam reads grooves that lie below a heavy coating of plastic.

Presenting videodisc lessons to groups of students makes the technology even more affordable. The combined cost of the hardware and software for a program such as *Mastering Fractions* is approximately the same as two Apple microcomputers and one or two inexpensive math software programs. If the videodisc is used five periods each day with classes of 20 students, one hundred students are served each day. In contrast, two microcomputers used for five periods each day serve only ten students.

This study demonstrates how the interactive videodisc can facilitate the implementation of effective mathematics instruction. The capability of the videodisc medium to incorporate instructional design features, together with its cost effectiveness, demonstrates its potential as a powerful instructional tool.

References

Brophy, J., & Good, T. (1986). Teacher behavior and student achievement in Wittrock, M. (Ed.). *Handbook of Research on Teaching (3rd Ed.).* New York: Macmillan.

Carnine, D. W. (1980). Relationships between stimulus variation and the formation of misconceptions. *Journal of Educational Research*, 74(2), 106-110.

Carpenter, T. P., Coburn, T. G., Reys, R. E., & Wilson, J. W. (1976). Notes from national assessment: addition and multiplication with fractions, *The Arithmetic Teacher*. 23(6), 137-142.

Darch, C., Carnine, D., & Gersten, R. (1984). Explicit instruction in mathematics problem solving. *Journal of Educational Research*, 77(6), 351-359.

Engelmann, S., & Carnine, D. (1982). *Theory of Instruction: Principles and Applications.* New York. Irvington.

Englert, C. S. (1984). Effective direct instruction practices in special education settings. *RASE*, 5(2), 38-47.

Fennema, E., & Sherman, B. (1976). Fennema-Sherman mathematics aptitude scales. *KSAS Catalogue of Selected Documents* in Psychology, 6,(1),31.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1980). Teaching behaviors, academic learning time and student achievement; an overview. In C. Denham & A. Lieberman (Eds.), *Time to Learn.* Washington, D. C., USOE National Institute of Education

Gall, M., Gersten, R., Grace, D. & Erickson, D. (in preparation, 1987). *Determinents of effective algebra instruction.* Paper presented at American Educational Research Association, Washington, D.C.

Gersten, R., & Carnine, D. (1984). Direct instruction mathematics: a longitudinal evaluation of low income elementary school students. *The Elementary School Journal, 84*(4),395-407.

Gersten, R. M., Carnine, D. W., & Williams, P. B. (1982). Measuring implementation of a structured educational model in an urban school district: an observational approach. *Educational Evaluation and Policy Analysis, 4*(1), 67-69.

Good, T. L., & Grouws, D. A. (1979). The Missouri mathematics effectiveness project: An experimental study of fourth-grade classrooms. *Journal of Educational Psychology, 71* 355-362.

Good, T. L., Grouws, D. A., & Ebmeier, H. (1983). *Active Mathematics Teaching.* New York: Longman.

Hasselbring, E., Sherwood, R., & Bransford, J. (1986). *An Evaluation of the Mastering Fractions Level One Instructional Videodisc Program.* Unpublished manuscript, George Peabody College of Vanderbilt University, the Learning Technology Center, Nashville.

Kameenui, E., Carnine, D., Darch, C., & Stein, M. (1986). Two approaches to the development of mathematics instruction. *Elementary School Journal, 86*(5), 633-651.

Kelly, B. F (1986). *The relative effectiveness of instructional design features in two fractions curricula with mildly handicapped and remedial high school students.* Unpublished doctoral dissertation, University of Oregon, Eugene.

Lindquist, M., Carpenter, T. P., Silver, E., & Matthews, W. (1983). The third national mathematics assessment: Results and implications for elementary and middle schools. *Arithmetic Teacher, 31*(4), 14 19.

Rosenshine, B. (1983). Teaching functions in instructional programs. *The Elementary School Journal, 83,* 335-351.

Silbert, J., Carnine, D., & Stein, M. (1981). *Direct Instruction Mathematics.* Columbus: Merrill

Stein, M., Jenkins, J. R., & Arter, J. (1983). Designing successful mathematics instruction for low performing students. *Focus on Learning Problems in Mathematics, 5*(2), 79-99.

Systems Impact, Inc. (1985). *Mastering Fractions.* Washington, D.C.: Author.

Bernadette Kelly is Research Assoc., Douglas Carnine and Russell Gersten are Assoc. Professors, and Bonnie Grossen is a Graduate Asst.; all with the Follow Through Project, University of Oregon, 1751-C Alder, Eugene, OR 97403-1215.

Closing the Performance Gap in Secondary Education

John Woodward

Douglas Carnine

University of Oregon

Maria T. Collins

Northern Illinois University

Quite commonly the differences between novices and experts are attributed to general intelligence, superior problem solving skills, or imagination (Larkin, McDermott, Simon & Simon, 1980). A closer examination of experts, however, reveals something else. Recent work in cognitive science has underscored the importance of subject area knowledge as a main factor that distinguishes novices from experts. Chi (1978), Chase and Simon (1973), Jeffries, Turner, Polson, and Atwood (1981) have conducted several interesting studies which indicate differences in these knowledge levels are mainly due to substantial practice, a thorough familiarity with the subject area, and a facile use of reliable strategies for solving problems in the particular subject area. Experts can rapidly retrieve items relevant to the problem at hand (Chase & Simon, 1973). As a result, experts solve problems considerably faster and more accurately than novices. These studies show that the primary difference is not innate ability, but expert's superior use of content specific facts, concepts, and problem solving strategies - a state arising from instruction and extensive practice or structured experience.

This research has considerable implications for special education, particularly for mildly handicapped secondary students who need to learn more than just basic skills. Some educators (e.g., Goodlad, 1983) note that these students, when mainstreamed, are expected to graduate from high school with their non-handicapped peers. Yet secondary special education instruction often amounts to simple drill on elementary facts and concepts. Although automaticity in math facts, for example, is important, these students also need to learn information that is more in keeping with what is taught at the secondary level. If these students are to master many of the same basic requirements as their non-handicapped peers, as some educators have suggested (e.g., Meyen, Alley, Scannell, Harnden, & Miller, 1982), then they need to be taught complex concepts and problem solving strategies as well as elementary facts.

This view is not intended to imply that mildly handicapped students will or should become experts in a subject area. Rather, mastery or competence in specific knowledge -- a less thorough, but adequate understanding of the material -- is a more reasonable goal. In essence, students need to be taught how think more "effectively" about a subject area. Students should learn core facts, concepts, and rules at an automatic level and then use explicitly taught strategies to solve a range of challenging problems.

Designing Instruction for the Secondary Level Through CAI

Teaching these skills involves well designed instruction within a specific content area. Such instruction should thoroughly describe the different stages that lead the mildly handicapped student from a novice state to a level of competence. A description of this process, which is always guided by the content as well as by instructional design principles, should detail how best to organize different kinds of knowledge in a content area. This description should also include the interrelationship between different kinds of knowledge (e.g., how knowledge of basic facts and concepts relate to problem solving strategies) and the optimal means for teaching each kind of knowledge. Some instructional designers (Case & Bereiter, 1984; Engelmann & Carnine, 1982) have already begun to articulate the many steps that move a student toward competence in a subject area.

For the last two years we have been studying the effects of CAI on mildly handicapped secondary students. CAI was chosen because of its increasing popularity in the schools and more importantly, because it is useful medium for embedding and then testing many empirically based instructional design principles. Our research has been conducted in three different areas: fact instruction in vocabulary, concept instruction in elementary logic, and problem solving in health education. Each of these approaches represents a different level of teaching in the novice to "competence" continuum.

As curricular approaches, the programs are designed to be tutorial or compensatory (Deshler, Schumaker, Lenz & Ellis,1984). That is, the programs are designed to either teach the student secondary level information in the quickest, most effective means possible (e.g., vocabulary instruction) or the curriculum is altered in such a fashion that students are taught comparable information in a different way. Technology, when used selectively, can assist this process by relieving teachers of time consuming, relatively low level teaching and, in some cases, as a way of conveying information not easily presented by conventional means.

In our experimental studies, the major focus in the data analysis has been the comparative performance of handicapped students randomly assigned to various experimental conditions. However, we have also tested samples of non-handicapped secondary students on the criterion referenced post test measures used in the study. This quasi-experimental procedure was used in order to gauge the progress of our experimental students in becoming competent in a content area. Our goal was for the subjects in the experimental treatment to a) surpass mildly handicapped subjects in the comparison group and b) perform at a level similar to non-handicapped students on the skill.

The items contained in our measures are well within the range of typical instruction at the secondary level. Where in some cases non-handicapped students may not have been directly taught the exact content prior to test administration (e.g., a particular vocabulary word or how to derive a conclusion from two premises), it is not unreasonable to expect that many of these students could have gained this knowledge on their own. Performance levels of non-handicapped students were significantly above a chance level of responding, enough to justify this assumption.

Insofar as the differences between the experimental groups and their non-handicapped peers diminishes after instruction, we are better able to understand the combined effects of instructional design principles and technology on the acquisition of domain specific knowledge. Furthermore, we can judge the extent to which mildly handicapped students are meeting basic academic requirements. What follows is a description of each program, the results of our quasi-experimental analyses with normal high school students, and the implications for knowledge development within a specific content area.

Teaching Vocabulary: Instruction at the Fact Level

Teaching vocabulary is regarded as important instructional activity, particularly as word knowledge is highly correlated with reading comprehension skills. In light of this, many researchers (Anderson & Freebody, 1981; Pearson & Gallagher, 1983; Tierney & Cunningham,1984) have looked for effective ways to teach vocabulary. Unfortunately, those methods which were most successful also required a considerable amount of instructional time. For example, a study by Beck, Perfetti, and McKeown (1982) attempted to teach only 104 words in 75 thirty-minute lessons. At the end of the study, students knew an average of 85 words that they did not know prior to the program, but this took 2,250 minutes of instruction or approximately 26 minutes per word. This amount of time is considerably more than that typically devoted to vocabulary instruction in the middle grades (Durkin, 1979; Roser & Juel, 1981).

Computer assisted instruction, it would appear, offers the advantage of increasing instructional time on such a low level task without placing increased demands on a teacher's already limited time. Students should be able to master the words more effectively if they are given extra practice on difficult words and cumulative review throughout the program. In the end, it is hoped that this knowledge will be used later in reading and writing activities.

We compared two methods of computer assisted instruction (CAI) for teaching vocabulary to mildly handicapped secondary students (Johnson, Carnine, & Gersten, 1986). The study examined the impact of two instructional design variables: the effect of size of the teaching sets and provisions for daily and cumulative review on the acquisition and maintenance of word meaning. Two CAI vocabulary programs were used to present the same 50 words and definitions.

The experimental software program used in the study, the Small Teaching Set program (Carnine, Rankin, & Granzin, 1984), begins by testing students on a set of 50 words. Lessons are created using only words student could not define on the pretest. The program then provides instruction on a "teaching set" of no more than three new words. After initial instruction, these words are then added to a "practice set" consisting of a maximum of seven words. The student must meet a specific mastery criterion on each word (i.e., two consecutively correct responses in each of two lessons) before it is removed from the practice set. Once the student has mastered ten words, the program tests the students on these words. Missed words are placed in the practice set and retaught. Figure 1 is a visual representation of the practice and review schedules embedded in the program. The figure shows how a word moves from an initial test item through a practice set to a final cumulative review lesson.

[Insert Figure 1 about here]

The comparison program, the Large Teaching Set program, teaches words in sets of 25 words (Davidson & Eckert, 1983). The student may choose to learn the words in any of four types of formats: (a) a teaching display which shows the word, its definition, and one example sentence; (b) a multiple choice quiz format; (c) an exercise in which a definition is displayed and the student must spell the correct missing word to complete a sentence; and (d) an arcade-type game in which the student matches words to their definitions. No cumulative review format was incorporated into the program.

Comparative Performance of Mildly Handicapped Students

Twenty-four mildly handicapped high school students were randomly assigned to one of the two CAI programs. The same set of 50 words were used in both programs. Students worked individually on their assigned program 20 minutes a day until they reached mastery. These words, 25 verbs and 25 adjectives, were considered important by two secondary special education teachers.

All students were given a 50 item, multiple choice test on the definitions as soon as they achieved mastery (i.e., 90 percent). Ten of the twelve subjects (83%) in the Small Teaching Set program met mastery criterion by the end of 11 sessions, while this was true for only eight of the twelve subjects (67%) in the Large Teaching Set program. The study was terminated after 11 sessions because the remaining subjects were experiencing frustrations. The mean number of sessions to mastery (for those who reached mastery) was 7.6 for those in the Small Teaching Set and 9.1 in the Large Teaching Set program. Results of a t-test indicate this difference is significant ($p < .05$). Hence, subjects in the Small Teaching Set program mastered the 50 words in significantly less time. In addition, more students in the Small Teaching Set program reached mastery within 11 lessons. Given that the groups achieved equivalent levels of performance on the multiple-choice tests, their difference in acquisition rates becomes even more meaningful. Subjects taught with the Small Teaching Set program required less time to meet mastery criterion on the words, yet their posttest performance was equal to that of subjects in the other treatment who took longer reaching mastery.

Comparison of Mildly Handicapped with Non-Handicapped Students

The same 50-item multiple choice vocabulary test was administered to a sample of 30 non-handicapped 10th-grade students in a regular English class. As Table 1 demonstrates, the posttest mean scores of the mildly handicapped subjects were slightly higher than the non-handicapped students' mean score. After a maximum of 11 sessions of computer-assisted vocabulary instruction, the performance of mildly handicapped subjects on the multiple choice test was very similar to that of non-handicapped 10th grade students.

[Insert Table 1 about here]

Implications for Software Design

Two issues arise from this study. First, the size of the teaching set and schedules for review led to an significant difference in learning rates between the two handicapped groups. Example set size and review schedules are comparatively subtle instructional design principles, yet they are essential for tasks where a considerable amount of practice and memorization are required.

Second, a minor finding in the study had to do with the arcade-type game contained in Large Teaching Set program. During the study, some of the Small Teaching Set students occasionally asked the experimenter why they didn't get to play a game like the one in the other program. However, after the study, students in the Small Teaching Set were asked what they specifically did not like about the program. Not one subject mentioned the lack of a computer game format.

This finding, though preliminary, is important for CAI software designers who apparently believe that for educational software to be motivating, it must approximate computer games that are popular in video arcades. Focusing on these kinds of surface features -- rather than the instructional design considerations relevant to the task -- may very well lead to software programs that are insufficiently structured for success.

Elementary Logic: Instruction at the Concept Level

An understanding of elementary reasoning and logic typically precedes a student's further training in analytic thinking. Once a student has a firm grounding in basic reasoning skills, teachers are in a better position to show students how to spot faulty arguments, identify false conclusions, and detect unwarranted generalizations. Zetlin and Bilsky (1980), however, suggest that educators create a self-fulfilling prophecy by not routinely teaching reasoning skills to special education students. These students consistently perform poorly on logical problem solving tasks (Spitz & Borys, 1977) and as a consequence, teachers often believe that these students cannot be taught reasoning skills.

The Reasoning Skills program (Engelmann, Carnine, & Collins, 1983) was designed to teach students to: a) to draw conclusions from two statements of evidence and b) to determine whether a three-statement argument was logical or illogical. The program taught students the three possible key words (some, all, no) that can begin any statement in an argument; their relationship to inclusive, overlapping and non-overlapping classes; and relevant rules for constructing arguments. Students were also taught to identify unsound arguments by citing one of three reasons (e.g., inappropriate key word in the conclusion, the appropriate class size is not named in the conclusion).

The major strength of the Reasoning Skills program is the teaching of an explicit, step-by-step strategy based on a series of carefully controlled rules. Figure 2a represents the skill of drawing a conclusion from two statements of evidence. This requires the student to first read the evidence statements and check for key words that begin each statement. On this basis, the student is able to use a set of rules to first determine the key word in the conclusion and next, to complete the rest of the conclusion based on an examination of the classes in the evidence.

[Insert Figure 2a about here]

Figure 2b portrays a more complex task: critiquing an argument. A student must read both the evidence and conclusion and determine if the conclusion follows from the two evidence statements. To critique our argument, the student must consider more features than when constructing a conclusion (e.g., implications of the key word in the conclusion for class membership and order in the evidence statements). As in the previous task, the student must look at key words and classes. However, he or she must now make this evaluation by using a set criteria (i.e., the multiple choice items) that force the student to apply all previously learned knowledge about arguments.

[Insert Figure 2b about here]

The advantage of the Reasoning Skills program over more traditional introductions to elementary logic is that the program teaches concepts with a minimum of verbage. Concepts such as major and minor premises, middle terms, distribution of terms, and subject and predicate distinctions are avoided. Even further, the reflexive relationship between the statements of evidence (i.e., their order or position can be interchanged with no effect on the conclusion) are demonstrated in the program rather than the typical method where the major premise is conventionally written first (Black, 1952). For example, consider the argument,

All French presidents are bald.
Some socialists are French presidents.

Some socialists are bald.

It would be common for the major premise (All French presidents are bald) to appear first, even though this is unnecessary. It is likely that students, particularly mildly handicapped students, who continually see only this kind of ordering will have difficulty drawing conclusions when the statements of evidence are reversed.

Comparative Performance of the Mildly Handicapped Students

The main interest in our study was to examine the effects of different correction procedures on two groups of remedial and mildly handicapped students (Collins, Carnine & Gersten, in press). Thirty-four students were randomly assigned to one of two groups: the Basic Correction or the Elaborated Correction group. When students in the Basic Correction group made an error, they were told the response was incorrect and were provided the correct answer. When a student in the Elaborated Correction group made a mistake, he or she was immediately corrected and an explanation related the explicit strategy that the student had learned earlier. This was the only difference between the two conditions. In both conditions, students worked individually on a microcomputer. Students worked on their respective version of the program until they completed five lessons.

Student learning was measured on a criterion referenced test. The data analysis indicated a significant difference favoring the Elaborated Corrections group ($p < .001$) on both the immediate posttest and a maintenance test administered two weeks later. There was also a significant difference between the two groups on the transfer test, again favoring the Elaborated Correction group ($p < .05$). The transfer test used arguments embedded in prose passages.

The two groups took roughly an equivalent amount of time to complete the five lessons, indicating the extra time required to read the elaborated corrections may have been compensated for by faster acquisition of the material. This interpretation suggests that taking more time early in a complex instructional sequence to offer elaborated corrections may, in fact, lead to savings in instructional time later in a program.

Comparative Performance of the Mildly Handicapped Students

Following this study, the program was revised and presented to another sample of mildly handicapped secondary students. The Test of Formal Logic was also administered to three non-handicapped groups: a tenth grade honors class, a college level logic class, and college level education students. Table 2 shows the ANOVA results Part 1 of the Formal Logic Test. This section of the test measured the students' ability to identify the key word in the conclusion and write the remainder of the conclusion based on two evidence statements. Tukey post hoc comparisons showed only one significant difference between the first three groups (i.e., the instructed handicapped students, the honors class, the logic class) and the education students ($p < .05$). The education students from the university scored significantly lower than the instructed handicapped students and the other two groups.

[Insert Table 2 about here]

Table 3 shows ANOVA results Part II of the test. This section requires students to determine whether or not an argument is faulty and if so, select a reason. A Tukey post hoc comparison showed a significant difference between logic class and the other three groups (i.e., the handicapped, honors class, and education students) ($p < .05$). There were nonsignificant differences between the last three groups. This finding indicates that on sophisticated reasoning skills, only the university logic students are competent. In contrast, on the easier reasoning skills, all the groups are comparable except for the lower performing university education students. Most important, there were non-significant differences between the university logic class and the high school handicapped students on the easier reasoning skills.

[Insert Table 3 about here]

Implications for Software Design

As previously described, the Reasoning Skills program contains several instructional design features that allow the student to achieve competence in a complex area of knowledge. Most important is a generalizable strategy that applied to all arguments except ones containing double negatives. Once we devised an overall strategy, the program was divided into distinct components. Necessary skills for each component were pretaught. For example, students were taught how to evaluate evidence statements to see if they were appropriate before they applied rules for using key words in determining the logical soundness of the conclusion.

Each component was chained to the next one. As the learner moved from one component to the next, prompts were faded. That is, added instructional elements -- ones that would trigger the execution of a compor nt -- were gradually removed. For example, once a student learned how to evaluate the appropriateness of evidence statements they were prompted to look at the first word in the conclusion. Eventually, this prompt was faded and instruction focused instead on the two classes in the conclusion that followed the key word. Introduction of new components and prompt fading continued until all types of arguments (i.e., all, some, no) were gradually integrated. Students were given discrimination practice between the different types of arguments for the remaining lessons in the program.

By minimizing the verbage traditionally associated with the subject and concentrating, instead, on class size, the student is able to "reason" about arguments. The program demonstrates that a CAI tutorial can teach these skills without added teacher instruction. What is required is a careful preliminary analysis of the content by a curriculum designer. The next step, which has not been completed yet, is to link the program to further instruction in reasoning and logic (e.g., analyzing longer arguments or detecting improper generalizations in short paragraphs).

Health Knowledge: Instruction in Problem Solving

Secondary students spend a considerable amount of their time completing application-oriented activities. These academic tasks often involve higher order cognitive skills, and students are asked to make a variety of inferences about a subject area by prudently using facts, concepts, and content related strategies or problem solving skills. Some writers (Doob, 1972; Greenblat & Duke, 1975; Budoff, Thormann, & Gras, 1984) have suggested that one way to enhance the higher order skills of students is through educational simulations. In addition, simulations have been suggested as a way of increasing the participation of lower achieving and inattentive students (Farran, 1968; Boocock & Schild, 1968; Stembler, 1975).

One of our interests in studying simulations was to investigate how they could be used to enhance -- rather than replace -- secondary level instruction, not only in terms of their effect on basic fact and concept retention, but as they related to problem solving. We chose a health simulation because it was designed to foster the acquisition of particular strategies. Health Ways was preceded by a tutorial containing three simpler versions of the simulation profiles, each one slightly more complex than the preceding one. This gradual progression from simple to complex allowed aspects of an overall monitoring strategy to be introduced and practiced. Health was also a good subject area because it is rich in facts and concepts.

Figure 3 gives a visual representation of the strategies the students needed to use to succeed at Health Ways. Students monitor three separate strategies (i.e., prioritizing and changing bad habits, checking the stress level, and maintaining health changes) through a monitoring or meta-strategy. While playing a Health Ways game, the student first prioritizies and changes a bad health habit, moving down through the tree until an appropriate action can be taken. If there is no current disease, he or she next looks at the hereditary diseases. If there is one, a related health habit (e.g., eating foods with too much sugar for a person with an hereditary history of diabetes) is identified and the student attempts to change the habit through the F8 "computer." F8, essentially, simulates fate or chance. It displays four random numbers, each between five and twenty-five. Number values are associated with successful changes and the score on F8 determines whether or not the habit can be changed.

[Insert Figure 3 about here]

Regardless of the success on the F8 game, the student must return to the upper level of the the tree and move to the right to the check-stress- level strategy. Again, the student descends in the tree, this time in the check-stress branch, to determine the appropriate action. Next, the student returns to the upper level, moves to the maintain-health-changes strategy and, if necessary, descends in that branch. The process of descending and traversing the tree (i.e., going back to the far left once the right most branch is checked) is repeated until the student succeeds or fails at achieving the main goal (i.e., increasing expected age to winning age).

Comparative Performance of the Mildly Handicapped Students

To measure the effects of the simulation, thirty students were randomly assigned to either the conventional or simulation condition. Direct instruction techniques were used to teach a typical health curriculum to all students for 20 minutes per day for twelve days. This was the first part of each day's lesson.

At the end of the initial instruction, students separated into two groups - one which worked on application activities (the conventional group) and the other with the computer simulation (the simulation group). The conventional group worked in the resource room under the supervision of the resource room teacher, who presented these students with a variety of application or review activities.

Simulation students, on the other hand, were taught in a computer lab, each student working individually at a microcomputer. The twelve day course of instruction for these students was broken into three phases: initial modeling (three days), guided practice on three simulation games (two days), and independent practice with individual feedback from the instructor (seven days). During the initial modeling phase, students were taught an explicit strategy for using the simulation. The effects of appropriate and inappropriate strategies were demonstrated. Students were first shown how to prioritize health problems by using information they had learned in the direct instruction portion of the lesson. As the researcher demonstrated progressively more difficult games or profiles, students were shown how to monitor and change two other variables: stress level and maintenance of health changes. During the guided practice phase, students were then able to practice different strategies with feedback from the researcher.

Students were assessed one day, two days, and two weeks following the instruction. On the first day, student's acquisition of basic facts and concepts about health taught in the curriculum was measured by the Nutrition and Disease Test. The first 20 questions of this test were solely from the written curriculum. The remaining 10 covered material that appeared in both the written curriculum and the Health Ways simulation. Internal consistency reliability (coefficient alpha) of this measure is .84. On the second day, the students were given the Health Diagnosis Test, an individually administered test that measured prioritizing skills. This test was a set of three written profiles and measured health related problem solving skills (i.e., the student's ability to detect important health problems facing an individual, identify and change related health habits, and control stress as it increased due to the health changes). The Health Diagnosis Test has a test - retest reliability of .81. Two weeks after the instruction the students were again given the Nutrition and Disease Test. This served as a retention measure.

The 30 item Nutrition and Disease test was broken into two subscales: (a) items reinforced by the Health Ways simulation, and (b) items taught in the curriculum and not reinforced by the simulation. The effect on items reinforced by Health Ways was significant ($p < .01$) and nonsignificant for those items not reinforced ($p < .06$). This indicates that the simulation was an effective vehicle for reviewing material that had already been taught in the written curriculum.

t-tests performed on the Diagnosis Test demonstrate a significant difference between the two groups ($p < .001$) in problem solving skills. Simulation students were better able to diagnose health problems, prioritize them as to their effects on an individual's longevity, and prescribe appropriate remedies.

Comparison of the Mildly Handicapped with Non-Handicapped Students

A one way analysis of variance (ANOVA) compared the test performance of the conventional and simulation groups with a random selection of students from regular health classes who did not participate in the study. Again, scores from each section of the Health Ways Nutrition and Disease Test and the Health Ways Diagnosis Test were analyzed. Table 4 shows a significant difference between the three groups on the Diagnosis Test ($p < .001$). A Tukey post-hoc comparison indicated significant differences between the handicapped simulation students and those in the regular classroom ($p < .01$) as well as a significant difference between the regular classroom students and the mildly handicapped students in the conventional group ($p < .01$).

A significant difference also appeared between the groups on the reinforced subscale of the Nutrition and Disease Test ($p < .01$). Tukey post-hoc comparison showed a significant difference between the mildly handicapped simulation group and the two other groups ($p < .05$), favoring the handicapped students taught by Health Ways, but no difference on items not reinforced.

We infer from the results that a combination of direct instruction in basic facts and concepts with a computer simulation was successful in teaching problem solving in a content area. Further, the superior performance by those in the simulation group over non-handicapped students from regular health classes shows that this kind of problem solving is by no means an automatic by-product of regular high school instruction. Instead, teaching competence in health requires a careful orchestration and integration of facts, concepts, and strategies.

Implications for Software Design

The success of the Health Ways study was a direct product of a careful analysis of simulation interventions. As Figure 3 indicates, a student must use many skills. In order to execute appropriate actions, a student must have a firm grasp of both facts (e.g., what is cholesterol? What disease is related to cholesterol?) and strategies (e.g., The stress level is going up and I haven't changed an important bad habit yet. What do I do?). In such a network of information is is easy for a student to act in many ways that lead to serious errors. For mildly handicapped students, the effect of this is usually frustration and a failure to learn anything from the simulation.

This is why an explicit strategy is essential. As with Reasoning Skills, components of the strategy are progressively introduced and then chained together. Here, students first learned about prioritizing and then were prompted to execute specific actions under certain conditions (e.g., The character's current disease is lung cancer. What related habit should you look at? Does alcohol have anything to do with lung cancer? Does smoking?). When the next component (i.e., stress management) was introduced, prompts for students for prioritizing were gradually faded. The fading, which lasted through the guided practice phase, allowed students to maintain a high level of success while learning essential skills.

Integrating software with traditional curriculum and using an explicit strategy for using the simulation had a very significant effect on problem solving ability and hence, student competence in the content area. In health, as with many science and social studies areas, there are wide range of goals, many of which are discretionary. This study shows that both the curriculum and the software can be adapted to meet important instructional goals, ones that lead to increased competence in the subject matter.

Finally, linking traditional practices to computer instruction allows for the optimal use of each medium. Group instruction is an efficient way of teaching and firming basic fact and concept knowledge. It is particularly appealing where schools only provide enough computers for an entire class in a computer lab. With the high demand placed on labs, computer time must be used judiciously. In this study, Health Ways was used to teach problem solving skills that could not be easily be demonstrated by conventional means. Thus, computer use was restricted to an area where it optimized instruction.

Conclusions

The three studies suggest that instructional design principles and computer technology can effectively work together in teaching mildly handicapped students to think more effectively about a content area. We believe that successful programs -- either with or without the use of technology -- begin with a careful analysis of how best to teach the content. This requires an understanding of how different kinds of knowledge in the content area are related as well as how the knowledge can be effectively sequenced. From this content analysis, an integral part of the instructional design process, comes the use of empirically based principles for sequencing and presenting the material. Finally, we consider whether or not technology is the most efficient or optimal means of instruction.

The three studies reviewed above bear out this curriculum process. The ability to define words is fact level knowledge that requires considerable practice. An above average amount of practice is required for mildly handicapped students. To increase the efficiency of this practice, an optimal example set size and cycles of review were employed. Finally, we used a CAI program incorporating these design principles in order to relieve teachers of this time consuming and relatively low level task.

In analyzing elementary reasoning skills (our example of concept teaching), we noted that traditional instruction often does not provide explicit, step-by-step strategies and tends to be laden with too many terms and definitions. An elaborated correction was used because the content was rule based and thus, when students erred they were reminded of the procedure for deriving the answer.

We used a CAI program to test whether or not such knowledge could be adequately taught as a tutorial. Typically, CAI programs merely provide drill and practice exercises to supplement teacher instruction. Here the program was a true tutoriai-- it did all the instructing.

Finally, our simulation instruction evolved out of an examination of problem solving instruction. As with social studies, health is a discretionary content area, allowing for various instructional goals. Typical health instruction often treats many diseases and bad health habits in an indepth, but undifferentiated fashion. That is, students are rarely given the opportunity to comparatively examine and prioritize the relative impact of different habits on a particular individual's heredity and lifestyle. Nor do they integrate the implications growing out of this prioritization with stress management and maintenance of habit changes.

By combining direct instruction in basic facts and concepts with an explicit strategy for using a simulation, we were able to teach more advanced forms of knowledge in health. The explicit strategy enabled the mildly handicapped students to focus on the essential features of the simulation and not be misled by distracting information. We chose the computer simulation because it allowed us to dynamically display changes in the three main areas of each profile (i.e., prioritizing, stress management, and maintenance). Each change had repercussions on all other system variables, thus forcing the student to manipulate several factors at once. Although it might have been possible to demonstrate these kinds of interactions through other means (e.g., role playing, board games), we are convinced that a computer is the optimal medium for demonstrating dynamic change.

It is tempting to try to infer too much from the results of the quasi-experimental component of our studies. A truer reflection of the combined effects of instructional design principles and technology would come from a series of aptitude treatment interaction studies. It is worth noting, however, that our studies do give us some indication of the effect of these two variables on mastery of material. Results of the vocabulary study, for example, show that handicapped students achieved performance levels comparable to non-handicapped peers. This was a function, no doubt, of increased and efficient practice as well as direct instruction in the material. Furthermore, it is unlikely that the lack of difference between the two populations was due to some change in general intelligence or distractability.

Data from the logic study are even more compelling. Significant differences on Total Test show a superior knowledge of logic by the handicapped secondary students over college level education students. One might attribute this difference to the general difficulty of the skill. But it is the nonsignificant differences between the college level logic students and the handicapped students that is the most important comparison.. These data indicate that mildly handicapped students can be taught complex material to a level that is comparable to older, non-handicapped students who have received different instruction over the same content.

The mildly handicapped students who received the health simulation scored higher than both the handicapped control group and the non-handicapped peers from regular health classes on the problem solving measure. Admittedly, much of this difference can be attributed to specific instruction in this area; these skills are not a ready by-product of typical health instruction. However, these are desirable problem solving skills, as noted by three secondary health teachers who reviewed the measure. Furthermore, these skills reflect an above average level of competence.

We base this comment on the performance of two non-handicapped students who were given the measure. Both students, rated by the teacher in a later discussion as being two of the best students in the class, had the highest scores on the Diagnosis Test. When asked by the researcher why they had completed the exercises as they did, each student articulated a set of strategies that were fully consistent with those contained in the explicit strategy. Thus, the handicapped students in the simulation group were taught to use strategies highly comparable to ones used by the two non-handicapped students. Performance by these two students reveals an integration of health knowledge that is at a higher level than most of their peers yet comparable to many of the handicapped students in the simulation group.

Our results do not imply that a concentrated effort in instructional design and technology will erase all differences between mildly handicapped and non-handicapped students. There are simply too many other variables that account for the difference between the two groups. Instead, by pushing these two factors to the forefront, we are better able to explore the limits of education for the mildly handicapped in a far more precise manner. In doing this, we come closer to enabling secondary mildly handicapped students to meeting the basic requirements expected of their non-handicapped peers.

Technology enables us to present certain kinds of instruction (e.g., the dynamic change in a health profile) in ways that we were incapable of doing in the past; the consequence being an integration of traditional and technology based curriculum. This point is critical. Software designers in the past have too often looked at technology based programs as stand alone products. Many times this has led to peculiar developments -- to game formats that will hold the students attention or to programs that are so broad (e.g., LOGO, Rocky's Boots) that the instructional goals are neither clear nor easily accomplished. A better understanding of the application of technology in special education is gradually emerging as it is linked to better instructional design principles.

References

Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), Comprehension and teaching: Research reviews. Newark, DE: International Reading Association.

Beck, I.L., Perfetti, C.A., & McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. Journal of Educational Psychology, 74, 506-521.

Black, M. (1952). Language and logic. Engelwood Cliffs, N.J.: Prentice Hall, Inc.

Boocock, S., & Schild, E. (Eds.). (1968). Simulation games in learning. Beverly Hills, CA: Sage.

Budoff, M., Thormann, J., & Gras, A. (1984). Microcomputers in special education. Cambridge, MA: Brookline Books.

Carnine, D. (1984). Mainstreaming computers. Educational Leadership, 41 (8), 77-82.

Carnine, D., Rankin, G., & Granzin, A. (1984). Learning vocabulary. Unpublished computer program, Eugene, Oregon.

Case, R., & Bereiter, C. (1984). From behaviourism to cognitive development: Steps in the evolution of instructional design. Instructional Science, 13, 141-158.

Chase, W. & Simon, H. (1973). The mind's eye in chess. In W.G. Chase (Ed.), Visual information processing (pp. 215-281). New York: Academic Press.

Chi, M. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), Children's thinking: What develops? (pp. 73-96). Hillsdale, NJ: Erlbaum.

Collins, M., Carnine, D., & Gersten, R. (In press). Elaborated corrected feedback and the acquisition of reasoning skills. A study of computer assisted instruction. Exceptional Children.

Davidson, J., & Eckert, R. (1983). Word attack! [computer program]. Ranchos Palos Verdes, CA: Davidson & Associates.

Deshler, D., Schumaker, J., Lenz, B., & Ellis, E. (1984). Academic and cognitive interventions for LD adolescents: Part II. Journal of Learning Disabilities, 17 (3), 170-179.

Doob, P. (1972). Prospects for simulation gaming in health planning and consumer health education. ERIC ED 070 596.

Durkin, D. (1979). What classroom observations reveal about reading comprehension instruction. Reading Research Quarterly, 14, 481-533.

Engelmann, S., & Carnine, D. (1982). Theory of instruction: Principles and applications. New York: Irvington Press.

Engelmann, S., Carnine, D., & Collins, M. (1983). DIAL reasoning skills program [computer program]. Eugene, Oregon: Engelmann - Becker Corp.

Farran, D. (1968). Competition and learning for underachievers. In S. Boocock & E. Schild (Eds.),

Goodlad, J. (1983). A place called school. New York: McGraw-Hill.

Greenblat, C., & Duke, R. (1975). Gaming simulation: Rationale, design, and application. New York: Halsted.

Jeffries, R., Turner, A., Polson, P., & Atwood, M. (1981). The process involved in designing software. In J. R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 255-283). Hillsdale, NJ: Erlbaum.

Johnson, G., Carnine, D., & Gersten, R. (1986). Effects of instructional design variables on vocabulary acquisition of LD students: A study of computer assisted instruction. Manuscript submitted for publication.

Larkin, J., McDermott, J., Simon, D., Simon, H. (1980). Expert and novice performance in solving physics problems. Science, 208, (20), 1335-1342.

Meyen, E., Alley, G., Scannell, D., Hamden, G., & Miller, K. (1982). A mandated minimum competency-based test program and its impact on the learning disabled (Research Report No. 40). Lawrence, KS: The University of Kansas Institute for Research in Learning Disabilities.

Pearson, P.D., & Gallagher, M. (1983). The instruction of reading comprehension. Contemporary Educational Psychology, 8, 317-344.

Roser, N., & Juel, C. (1981, December). Effects of vocabulary instruction on reading comprehension. Paper presented at the annual meeting of the National Reading Conference, Dallas.

Spitz, H., & Borys, S. (1977). Performance of retarded adolescents and nonretarded children on one and two-bit logical problems. Journal of Experimental Child Psychology, 23, 415-429.

Stembler, W. (1975). Cognitive effects of a programmed simulation. Simulation & Games, 6, 392-403.

Tierney, R.J., & Cunningham, J.W. (1984). Research on teaching reading comprehension. In P.D. Pearson (Ed.), Handbook of reading research (pp. 609-656). New York: Longman.

Zetlin, A. & Bilsky, L. (1980). Reasoning by trainable mentally retarded and young non-retarded individuals. Journal of Mental Deficiency Research, 24 (1), 65-71.

Table 1

Comparison of Mildly Handicapped with Nonhandicapped Samples:
Multiple Choice Vocabulary Test

Group	Test	n	M	SD	Mean Percent Correct
Small Teaching Set	posttest	12	42.0	4.0	84.0
Large Teaching Set	posttest	12	43.7	7.7	87.4
Nonhandicapped Comparison (10th grade)	--	26	40.3	4.9	80.6

Table 2

Part I of the Formal Logic Test

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	p
Between Groups	3	258.21	80.07	15.5	.001
Within Groups	143	792.44	5.54		
Total	146	1050.65			

Table 3

Part II of the Formal Logic Test

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	p
Between Groups	3	112.2	37.41	4.49	.005
Within Groups	143	1190.18	8.32		
Total	146	1302.42			

Table 4

Means (M) and Standard Deviations (SD) for Three Groups

	N	M	SD
Nutrition and Disease Test: <u>Total Score</u>			
Mildly Handcapped Students:			
Simulation	15	22.00	3.72
Conventional	15	17.93	5.86
Non-Handcapped Students:	15	19.47	4.94
Nutrition and Disease Test: Items Reinforced by <u>Health Ways</u>			
Mildly Handicapped Students:			
Simulation	15	7.33	1.35
Conventional	15	5.60	2.20
Non-Handicapped Students:	15	5.53	1.46
Health Ways Diagnosis Test: <u>Total Score</u>			
Mildly Handicapped Students:			
Simulation	15	27.73	5.89
Conventional	15	12.47	4.88
Non-Handicapped Students:	15	18.07	6.03

FIGURE 1

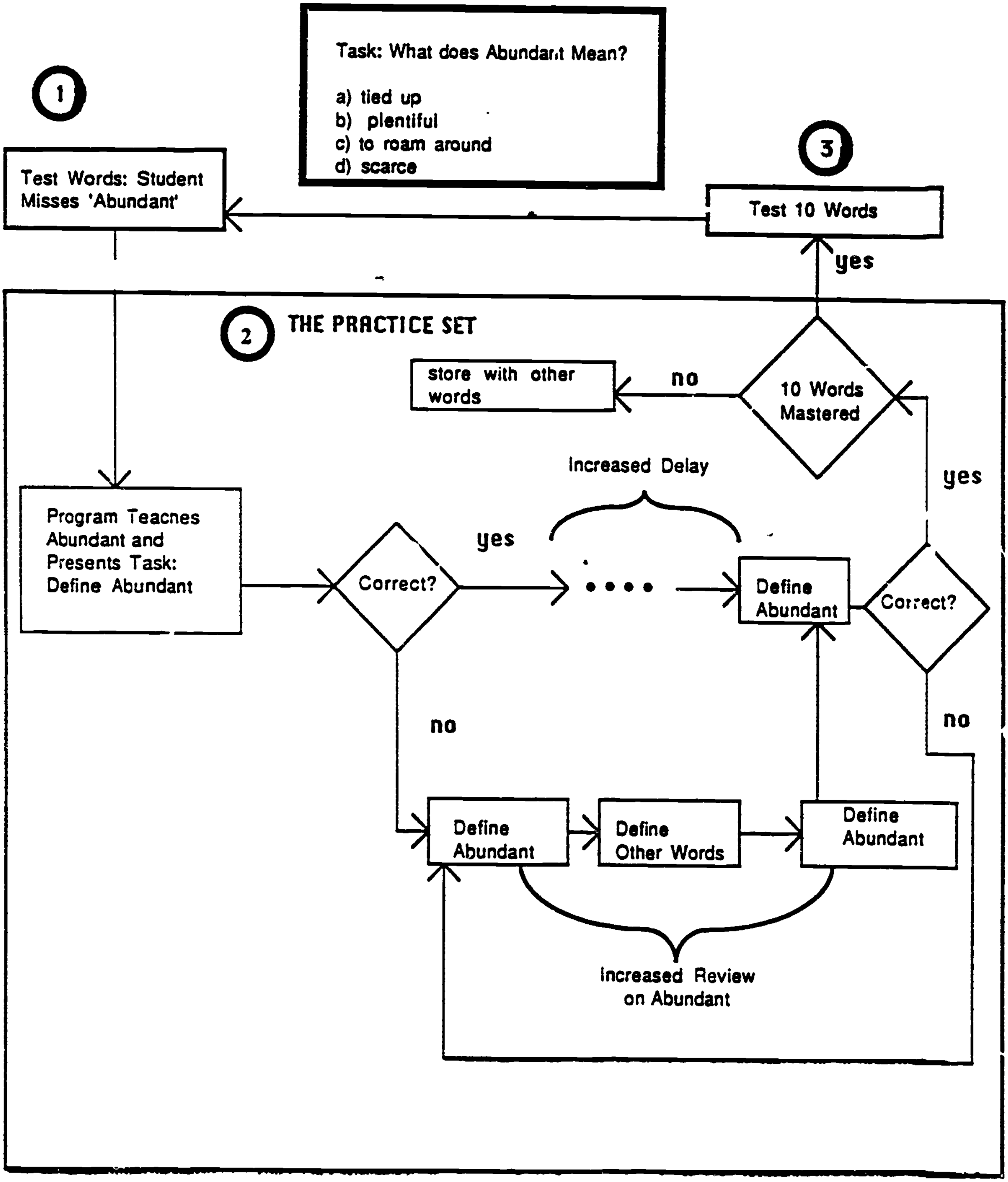

Task: What does Abundant Mean?
a) tied up
b) plentiful
c) to roam around
d) scarce
1
Test Words: Student Misses 'Abundant'
3
Test 10 Words
yes
2
THE PRACTICE SET
store with other words
no
10 Words Mastered
Increased Delay
yes
Program Teaches Abundant and Presents Task: Define Abundant
Correct?
yes
Define Abundant
Correct?
no
no
Define Abundant
Define Other Words
Define Abundant
Increased Review on Abundant

Figure 2a

Problem: All incisors are teeth
No teeth are muscles

Task 1: What will be the first word in the conclusion? (all, some, no)

Task 2: Write the conclusion on the line below.

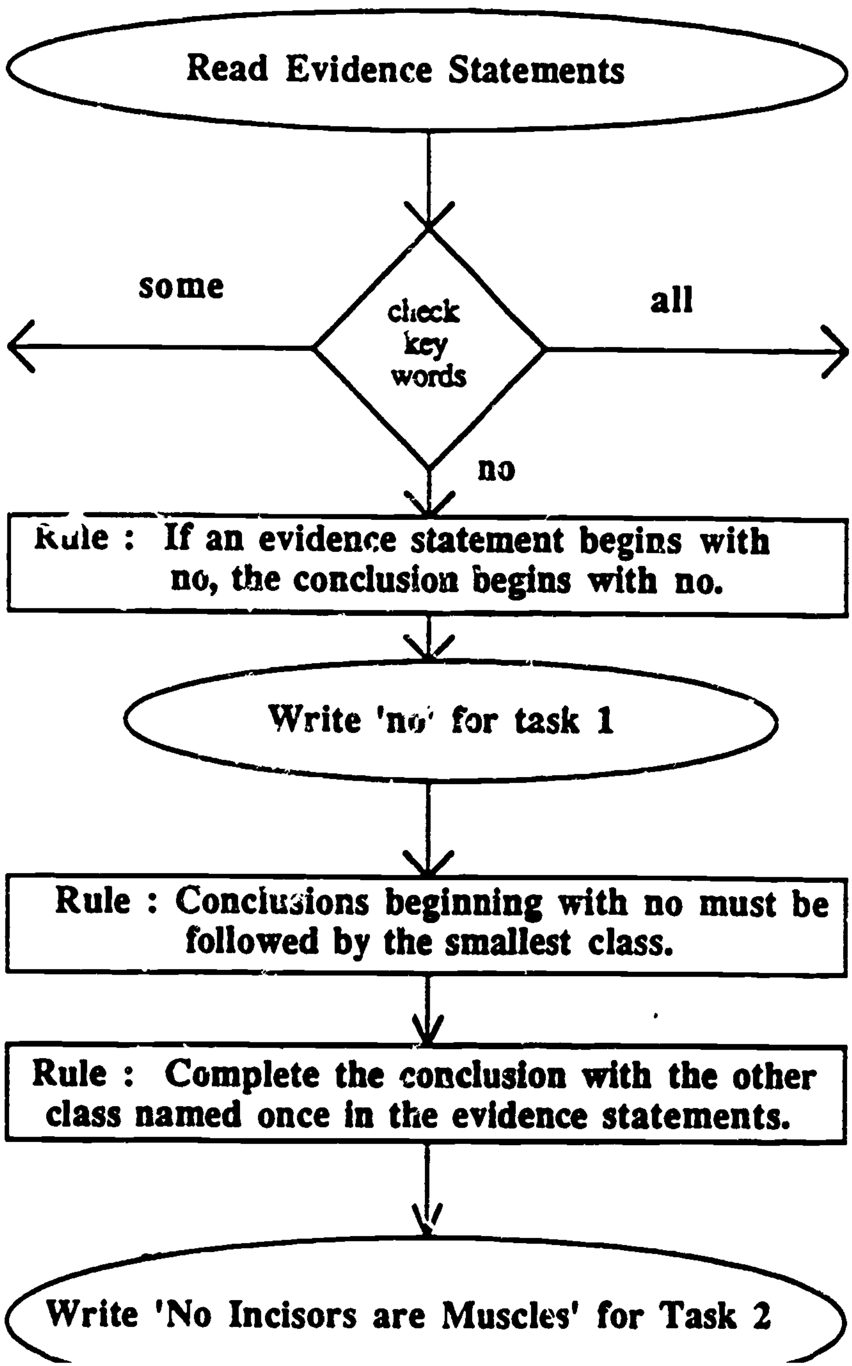


Figure 2b

Problem:

No metals are plants
All plants are living things
No metals are living things

Write the number below that best tells about this argument

1. The argument is sound
2. The conclusion does not name the smallest class
3. The conclusion does not name the largest class
4. The conclusion does not begin with the right word

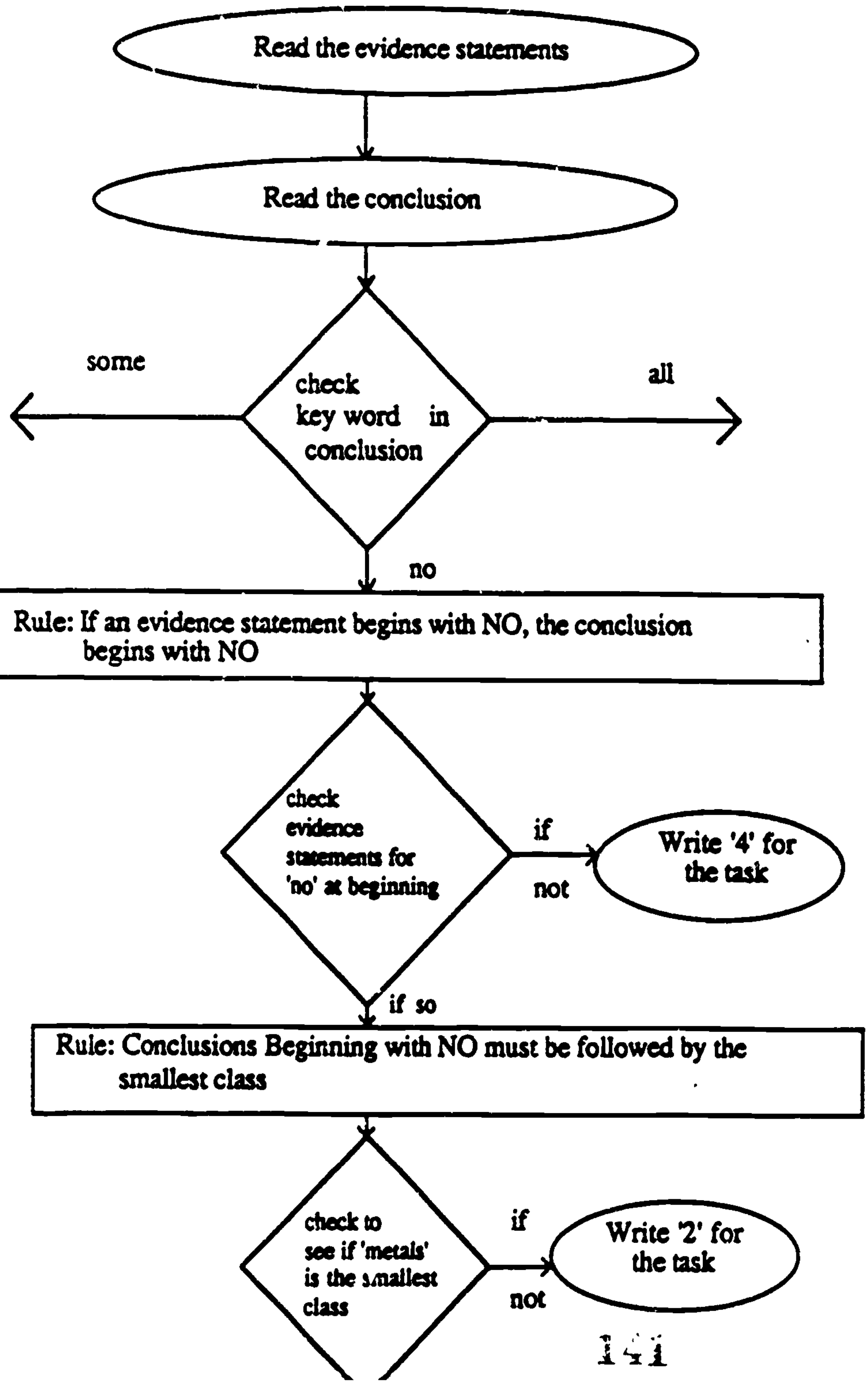
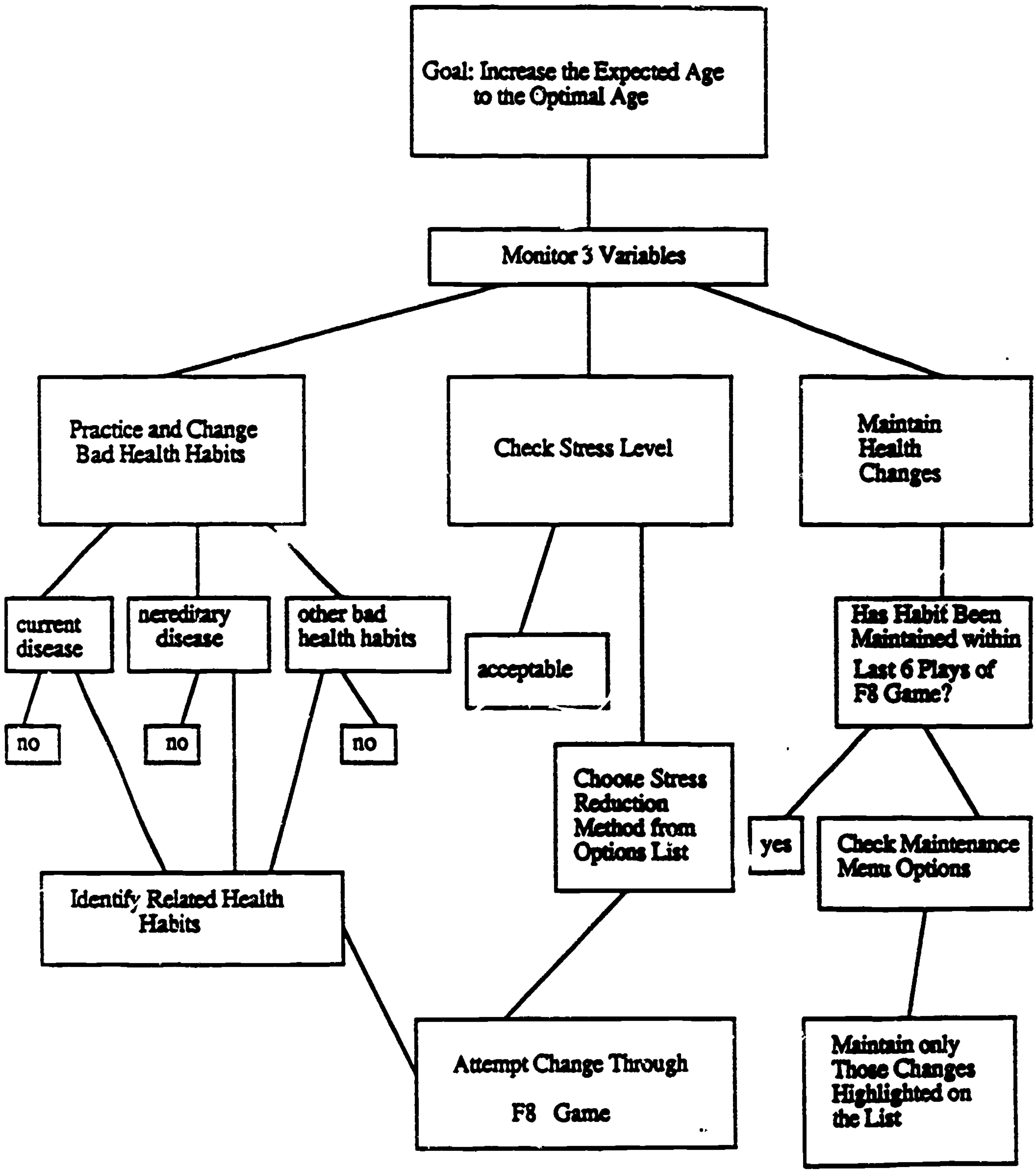


Figure 3



Goal: Increase the Expected Age to the Optimal Age
Monitor 3 Variables
Practice and Change Bad Health Habits
Check Stress Level
Maintain Health Changes
current disease
nereditary disease
other bad health habits
acceptable
Has Habit Been Maintained within Last 6 Plays of F8 Game?
no
no
no
Choose Stress Reduction Method from Options List
yes
Check Maintenance Menu Options
Identify Related Health Habits
Attempt Change Through F8 Game
Maintain only Those Changes Highlighted on the List