AL/HR-TP-1995-0031

INDIVIDUAL AND COOPERATIVE GROUP LEARNING WITH USER-CONTROLLED AND PROGRAM-CONTROLLED MATHEMATICS TUTORS

ARMSTRONG LABORATORY

Robert M. Yadrick
J. Wesley Regian

HUMAN RESOURCES DIRECTORATE
TECHNICAL TRAINING RESEARCH DIVISION
7909 Lindbergh Drive
Brooks AFB, Texas 78235-5352

Catherine Connolly Gomez
Linda Robertson-Schule

Galaxy Scientific Corporation
Bldg. 9020
1880 Carswell Avenue
Lackland AFB TX 78236-5507

November 1995

Interim Technical Paper for Period October 1993-December 1993

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS

NOTICES

This paper has been reviewed and is approved for publication.

ROBERT M. YADRICK
Project Scientist

R. BRUCE GOULD, Ph.D.
Technical Director

JAMES B. BUSHMAN, LT.COL., USAF
Chief, Technical Training Research Division
Human Resources Directorate

Please notify this office, AL/HRPP, 7909 Lindbergh Drive, Brooks AFB TX 78235-5352, if your address changes, or if you no longer want to receive our techinical reports. You may write or call the STINFO office at DSN 240-3853 or commercial (210) 536-3853.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
November 1995

3. REPORT TYPE AND DATES COVERED
Interim Technical Paper Oct 93 - Dec 93

4. TITLE AND SUBTITLE
Individual and Cooperative Group Learning with User-Controlled and Program-Controlled Mathematics Tutors

5. FUNDING NUMBERS
C - F41622-92-D-0006
PE - 61102F
PR - 2313
TA - T2
WU - 18

6. AUTHOR(S)
Robert M. Yadrick
J. Wesley Regian
Catherine Connolly Gomez
Linda Robertson-Schule

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Galaxy Scientific Corporation
1880 Carswell Ave, Bldg 9020
Lackland AFB TX 78235-5507

8. PERFORMING ORGANIZATION

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Armstrong Laboratory
Human Resources Directorate
Technical Training Research Division
7909 Lindbergh Drive
Brooks AFB TX 78235-5352

10. SPONSORING/MONITORING
AL/HR-TP-1995-0031

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We used a 2x2 design to compare the performance of remedial subjects learning to solve mathematics word problems. Subjects worked either alone or as a member of a dyadic cooperative group, and received up to 12 hours of instruction and practice, using either a "problem-solving environment" over which the user exercised considerable control or a directive, sometimes intrusive tutor. Individual subjects who worked with the problem-solving environment showed the largest average improvement between pretest and posttest scores. Group members showed moderate improvements after working with the directive tutor. There were no significant changes between pretest and posttest for either group members who worked with the problem-solving environment or for individuals who worked with the directive tutor.

14. SUBJECT TERMS
Cooperative group learning
Mathematics learning
Computer-based instruction

15. NUMBER OF PAGES
27

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Table of Contents

List of Tables

List of Figures

PREFACE

This report describes one of several experiments conducted in the TRAIN Cooperative Laboratory from October 1993 to March 1994. Funds for this research were provided by the U.S. Air Force Office of Scientific Research and the Armstrong Laboratory TRAIN Project, AL/HRTI, Brooks AFB, TX, Dr. Wes Regian, Director.

INDIVIDUAL AND COOPERATIVE GROUP LEARNING WITH USER-CONTROLLED AND PROGRAM-CONTROLLED MATHEMATICS TUTORS

INTRODUCTION

The present study originated with our interest in a number of initially independent concerns. First, we are interested in issues regarding the design, development, and evaluation of Computer-Based Instruction (CBI), primarily with a view to advancing Air Force training. We define "issues" to include both questions concerning instructional design and concerning the conditions which contribute to the effective implementation of CBI. For example, we have investigated applications of cooperative group training using CBI in other domains (Shebilske, Regian, Winfred, & Jordan, 1992), because of the potential for achieving efficiencies and economies in the use of Air Force training facilities and materials. One of our present purposes was to extend this research into additional domains.

Another purpose of the present study was to evaluate and compare two forms of a computer-based mathematics instruction system developed at the Air Force's Armstrong Laboratory, which incorporate different instructional approaches. One approach is comparatively unstructured and allows a considerable amount of user control (the Word Problem Solving Environment, or WPSE), while the other (Solver) is very structured and directive under certain circumstances, and sometimes places severe constraints on user actions.

Finally, the Air Force is concerned with remedial basic skills training for Air Force recruits. Determining efficient and effective approaches to delivering remedial training on fundamental academic skills is a matter of considerable practical importance to the Air Force.

We examined aspects of all these areas in this experiment, focusing on remedial subjects in a 2 x 2 factorial study. Subjects worked either as individuals or as members of dyadic cooperative groups, and used either the WPSE or Solver. In addition, the experimental procedure emulated the way in which Air Force technical training is sometimes conducted. For example, instruction and practice were concentrated in 3 days of intensive work, during which subjects in the grouping conditions were assigned to a team and shared instructional equipment with a partner they had not met before.

Cooperative Group Learning

There is substantial evidence that cooperative group learning results in greater student learning, relative to individual learning, although research has sometimes yielded inconsistent results (Johnson & Johnson, 1989). Similarly, research regarding the more specific issue of the relationship between CBI and cooperative group learning has yet to produce a completely clear picture. As Webb (1987) pointed out in a review of studies which compared individual and cooperative group learning in computer settings, the issue is clouded by the many differences between studies and the many factors which potentially could influence group effectiveness, such as the students' ages, the setting, the domain and subject matter, instructions regarding group interactions, achievement measures, group size, group ability mix, etc. Webb discussed nine studies which found no differences between cooperative group and individual work and five studies which reported differences in favor of cooperative groups, across a number of different subject-matter areas. She ultimately decided that it was impossible at the time to explain why some studies resulted in differences and others did not, but maintained that the important point was that "no study found greater learning among students working alone than students working in groups" (p. 195).

As Webb's (1987) observations about the various potentially important factors imply, investigators who wish to understand the reasons for outcome differences across studies must consider a number of influences simultaneously. There has been some investigation of the factors which may affect or limit the benefits of cooperative groups. For example, Nastasi and Clements (1991) suggested that group ability mixture appears to be a limiting factor: "research suggests... low ability students receive more explanations and learn more from heterogeneous than from homogeneous groups" (p. 121). Similarly, Hooper and Hannafin (1989) found a weaker relationship between interaction and achievement in homogeneous groups than in heterogeneous groups, even though there were more interactions among low-ability subjects than among high-ability subjects. It therefore seems likely that the nature, as well as the quantity of interactions between group members is important. In support of this notion, researchers (Hooper, 1992; Nastasi & Clements, 1991; Slavin, 1990; Webb, 1987, 1991) have identified a variety of group interactions which appear to foster increased learning.

Moreover, interactions between grouping and other variables have been observed. For example, holding each group member individually accountable for their own performance can be important (Hooper, Ward, Hannafin, & Clark, 1988). Perhaps the most pertinent for present purposes is a study by Mevarech (1991) which showed the potential relationship between grouping and instructional approach, although the study did not involve computer-based instruction. In a 2 x 2 design, subjects learned under either a mastery or a non-mastery instructional approach, and either worked alone or as members of small cooperative groups. Mevarech found that performance was best in the condition that combined both the mastery approach and grouping. However, relative to control group performance, he also found substantial and essentially equal positive effects for both the mastery approach alone and cooperative groups alone.

Learner Control of CBI

The effects of allowing learner control, as opposed to program control, over various aspects of how CBI is delivered have traditionally been investigated separately from the issue of group vs. individual learning. Early proponents of learner control intuitively expected that allowing students to sequence and pace lessons as they wished or access support features whenever they needed to would maximize their understanding, but early results were disappointing (Steinberg, 1978). As with the issue of grouping, results have been mixed (Kinzie, 1990) and differences between studies, CBI systems, populations, and domains make generalizing the findings in the literature difficult to do with any confidence.

One problem with understanding the effects of learner control is that studies have not focused systematically on control over particular aspects of CBI. For example, studies have allowed or not allowed learner control over continuing through a lesson if students are unable to answer questions along the way (Avner, Moore, & Smith, 1980); over sequencing of topics and sequencing of learning objectives within topics (Rubincam & Olivier, 1985); over the option of trying other alternatives after receiving feedback about an initial choice (Gray, 1987); as well as lesson pacing (Dalton, 1990), problem context (Morrison, Ross, & Baldwin, 1992), receiving feedback (Pridemore & Klein, 1991), adding or dropping instructional elements (Hicken, Sullivan, & Klein, 1992); and review options (Kinzie, Sullivan, & Berdel, 1988). In addition, some authors consider "adaptive control", which modifies instruction based on considerations like prior performance, as an alternative category (Park & Tennyson, 1983). As Milheim and Martin (1991) point out, the concept of learner control is actually a continuum, with total learner control at one end and total machine control at the other, while most implementations of learner control fall somewhere in between. We also note that the structure and complexity of the CBI system, including both the instructional approach and the number and variety of system features, largely determine what is available for allocation to program or learner control.

Another problem is that differences between learners appear to be important. Steinberg (1989) urged caution in making generalizations based on the handful of studies available at the time. However, she suggested that some generalizations regarding the advantages or disadvantages of allowing learner control "merit serious consideration" (page 120). One important generalization which she proposed is that beginning learners with little prior knowledge of a subject may not perform well if allowed control. For example, they may not manage their study time well (Tennyson, Tennyson, & Rothen, 1980), they may not sequence instruction properly (Gay, 1986); or they may not adopt a consistent learning strategy (Rubincam & Olivier, 1985). In general, beginners tend not to possess two skills that Steinberg (1989) considered important for making learner control effective. First, they may not discriminate accurately between critical and tangential information, and second, they often do not possess a suitable repertoire of domain-specific learning strategies. Lee and Lee (1991) added that beginners also frequently lack general, across-domain strategies. There is also evidence that learner control is used more effectively by high aptitude students than by low aptitude students (Ross & Rakow, 1981). On the other hand, it appears that learner control can be made more effective, even for beginners, by giving explanatory feedback (Steinberg, Baskin, & Hofer, 1986), by adaptively determining learner needs but making suggestions rather than requiring student actions (Tennyson, 1980), or by offering support features that more skilled learners find useful but allowing learners the option of using them or not as they prefer (Steinberg, 1989).

Lee and Wong (1989) provided evidence that prior domain knowledge is an important determinant of the effectiveness of learner control. They found that program control led to better performance than learner control for students who had low pretest scores, but that there was no difference for students with high pretest scores. The matter was explored further by Lee and Lee (1991), who crossed locus of control with learning phase. Students in their study worked with learner-controlled or program-controlled CBI either during an initial acquisition phase, defined as before traditional classroom instruction, or during a later review phase, defined as after traditional instruction. They found an advantage for program control during acquisition and an advantage for learner control during review, supporting the theory that learner control works best when students have prior domain knowledge. Lee and Wong point out that this could also explain why learner control can work well for simple tasks, since little prior knowledge is required (Tennyson & Rothen, 1979).

Finally, Hooper, Temiyakarn, and Williams (1993) recently studied cooperative learning and learner control simultaneously. Students in the learner control condition could vary the number of examples and practice items they received and could also decide whether to receive explanatory feedback. In the program control condition, students received all of the examples and practice items and were given explanatory feedback after each incorrect answer. However, there were no differences between these conditions on measures of achievement, attitudes, efficiency, or time on task. One finding of interest was that students in the group/ learner control condition decided to use only a limited amount of the instructional support available, paralleling similar findings for individuals (Steinberg, Baskin, & Matthews, 1985).

The Present Study

We intentionally attempted to select uniformly low-ability remedial subjects for this study. One prediction, therefore, seemed relatively clear. The literature regarding the benefits of working as a member of a cooperative group (Hooper & Hannafin, 1989; Nastasi & Clements, 1991) led us to expect that members of homogeneous low-ability groups would not benefit from grouping to the extent that other subjects might, including low-ability members of heterogeneous groups. We therefore expected to find no differences between individuals and group members using a given system.

In many respects this study was motivated more by practical issues of concern to the Air Force than by interest in resolving problems in the research literature, and it was not completely clear what we should expect to find. On the one hand, the literature indicates that beginners and people of low ability in a domain may benefit from program control (Lee & Wong, 1989; Lee & Lee, 1991; Ross & Rakow, 1981). However, it may be that Solver incorporates more program control than is beneficial, disrupting a user's train of thought or not allowing him or her to pursue a chosen problem-solving strategy. Further, as Hooper, Temiyakarn, and Williams (1993) point out, one should not assume that factors which affect individual instruction will necessarily have the same effects on group instruction. Neither is it clear how applicable previous research results actually are to our subjects, who had all been exposed to the domain before and were not really beginners. However, this exposure had taken place years before for most of our subjects, and the effects of prior instruction appeared to have largely decayed. We decided that the conservative approach would be to expect that these subjects would resemble beginners and would benefit more from using Solver than from using the WPSE.

Another tentative prediction was that there would be no interaction between the grouping and system variables. Thus, Solver users on average should perform better than WPSE users regardless of their status as individuals or group members.

METHOD

Participants

Subjects were recruited through several local temporary employment agencies. Although the exact amount varies slightly by agency, subjects were paid approximately $5.00/hour for their participation, a standard local wage for unskilled temporary workers. Subject characteristics were similar in some important respects to those of Air Force recruits, the primary group to which the results were to be extrapolated. (Air Force basic trainees are not available to serve as subjects in studies lasting more than a half day.) All were high-school graduates or had earned a high school equivalency certificate. Some had taken college courses, but none had a college degree. All were between the ages of 18 and 30, and had at some time successfully completed at least one mathematics course which covered the subject matter (for example, percentages) used in this study. This was determined in most cases by self-report, bolstered by the fact that completion of such a course is a requirement for high school graduation or equivalency in Texas. We began by selecting subjects in need of remediation, that is, who no longer could work domain problems reliably. Table 1 gives additional demographic information about these subjects.

Table 1
Subject Demographics by Group

		Ind/ WPSE	Ind/ Solver	Group/ WPSE	Group/ Solver
Sex	Male	9	8	9	7
	Female	7	7	9	9
Ethnicity	Black	2	1	3	2
	Hispanic	8	10	7	10
	White	6	4	8	4
College Attendance	No	14	10	13	14
	Yes	2	5	4	2
	No response			1	
Algebra	Yes	13	13	12	11
	No response	3	2	6	5

Our sample appears roughly to reflect the local population. For example, there is a larger proportion of Hispanics and a smaller proportion of blacks than one might find in a sample drawn from many other American cities. The sample probably is not very representative of the entire population of

Air Force recruits, either. Our goal, however, was to identify and study people who resembled Air Force recruits who might benefit from remedial training in practical mathematics. The composition of our target population is therefore unknown, although our subjects presumably resembled our target population in important ways, such as having relatively low language and quantitative skills.

Potential subjects were recruited by the agencies according to the age and education criteria, and were told that they might be selected to serve in any of a number of different studies. Groups of up to 30 subjects reported to our laboratory on each of seven successive Monday mornings. They were assigned to participate either in this experiment or in another study being conducted elsewhere in the laboratory. Selection was based on performance on a screening test, which is described in detail later in this article.

The experiment was run over the course of several weeks. The WPSE conditions were run during the first three weeks, and the Solver conditions were run during the second block of three weeks. Subjects were run in all four conditions during final "makeup" week, however, as we attempted to equalize the number of subjects in each condition by replacing individuals and groups who had dropped out or been discarded in previous weeks. Apart from these constraints, subjects selected for participation were assigned randomly to groups. The screening test was developed to identify suitable subjects quickly, but we did not consider it sufficiently discriminating to use for matching subjects. A low of five subjects and a high of 15 subjects qualified from the weekly pools of potential subjects.

A total of 65 subjects finished the study. In addition, a total of 23 other subjects began but dropped out or were discarded when the other member of their group dropped out. We contacted the appropriate employment agency and tried to determine why each dropout did not return. In most cases subjects offered legitimate reasons for not returning, such as car trouble or a child's illness, although a few said candidly that they disliked spending the day working math problems. Unfortunately, each time a group member dropped out, it was necessary to discard the data from his or her partner. We did not discard any subject's data for any reason apart from this.

Attrition was not concentrated in any particular condition. There were five dropouts in the Individual/WPSE condition, six total attritions (dropouts and dropped partners) in the Group/WPSE condition, four dropouts in the Individual/Solver condition, and eight total attritions in the Group/Solver condition.

Materials and Equipment

The study was conducted in our laboratory at Lackland Air Force Base, Texas, which consists of 30 networked Compaq 486/33L computers with NEC/Multisync VGA monitors, standard keyboards, and Logitech three-button MouseMan mice. The computers were situated in five rows of carrels. The carrels easily accommodated two people each, and offered some degree of protection from outside sounds and other distractions.

The Tutors -- Both versions of the tutor were developed by Armstrong Laboratory and contractor personnel using the Toolbook software construction set (Toolbook 1.0, 1989.) The problem pool was developed by mathematics teachers from San Antonio area middle schools and high schools, who were

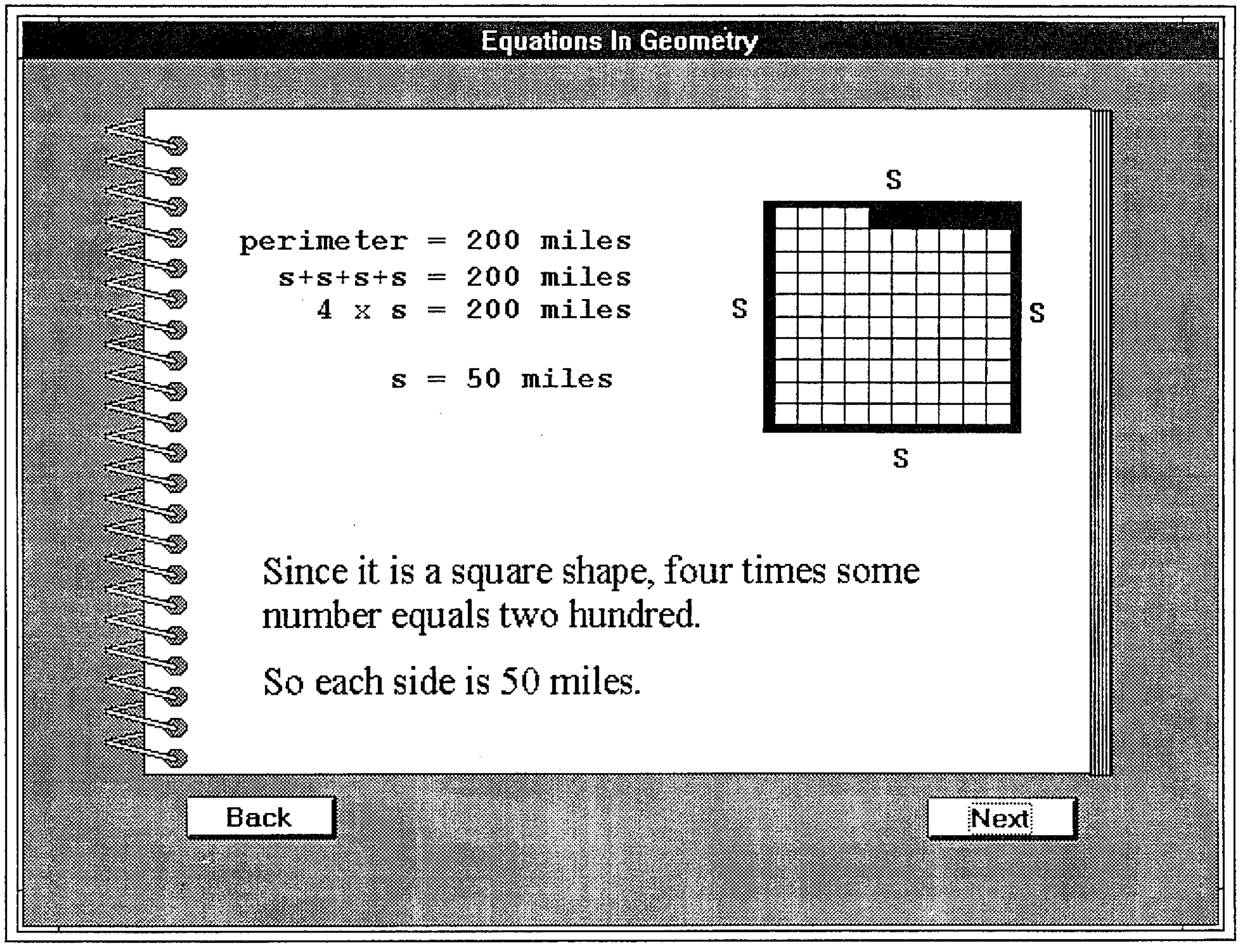

Figure 1. Sample Instructional Screen

hired as consultants. In the teachers' judgment, the problems are representative of those taught in eighth- and ninth-grade-level mathematics courses.

The WPSE makes considerable instructional capacity and support, such as hints, basic formulas, scale conversions, and the like available to the subject. However, the onus is on the subject to learn to select and execute the correct problem-solving steps, learn the appropriate sequences of interface manipulations, decide when to use various system features or to ask for help, and, most of all, learn how to solve problems at the same time.

The curriculum is modular. Each module concerns a different topic and begins with CBI which describe the concepts and principles of the topic using graphics, animation, and examples keyed to the text to illustrate important points. Figure 1 shows an instructional screen from the module on geometric equations.

Each CBI session is followed by a short multiple-choice quiz over the basic concepts of the module. Subjects who fail the quiz must go through the instructional sequence again until they can pass.

After passing the quiz, users are presented with a series of practice problems arranged in an ascending order of difficulty. Figure 2 shows a Level-2 problem from the module on geometric equations. A problem's difficulty level depends on such factors as the number of variables the problem includes, whether variable values are given directly or expressed in relation to the value of another variable, and the number of steps required to solve the problem (these factors are, of course, not necessarily independent). Problems at each level within a module are "equivalent" in the terminology of Reed, Dempster, and Ettinger (1985), that is, they share an underlying solution equation and have similar cover stories. For example, Level-1 problems from the module on geometric equations give two values from among a rectangle's length, width, and perimeter or area, and ask the subject to calculate the third value. The Level-2 problem shown in Figure 2 is more difficult than a Level-1 problem because, although the value for the perimeter is given expressly, the value for length is given in relation to the value for the width. The highest-difficulty-level problems in this module present the subject with to be found.

Subjects using the WPSE work problems by selecting in turn one of five "Problem Solving Steps" from a pull-down menu that appears when one clicks on the menu bar (see Figure 2): "Identify Goal", "Make Variables", "Make Equation", "Solve Equation", and "Answer Question". The user must first select "Identify Goal", then click on a word in the goal sentence ("What is the width...").

Next, the user must select "Make Variables", then provide labels for necessary variables (e.g., "Perimeter", "Longer than Width") and assign them values by clicking on numbers in the Problem Window or by entering numbers. The next step is to construct a word form of equation by clicking in turn on variables in the Variables Window and operators on the keypad to the left of the Variables Window. This equation appears in the Equation Window as it is constructed. For the problem in Figure 2, such an equation might read "Perimeter = (2 X Width) + (2 X (Width + Longer than Width))".

The next step, "Solve Equation", causes the numerical form of the equation and the solution to appear in the Equation Window: "66 = (2 X 15) + (2 X (15 + 3))". The point is to focus on the process of understanding a problem, developing a solution, and building an equation that embodies the solution, but there is no instruction or practice on the mechanics of computation. The last step, "Answer Question", involves entering the numerical answer and units ("15 inches") in a window that appears when the step is selected. The Instruction/Advice Window at the bottom of the screen retains all the help the subject receives from the system, so that by scrolling the user can review whatever hints, formulas, definitions, etc., have been presented previously. complex figures which must be broken into a set of rectangles and triangles if the figure's total area is

Clicking on "Help" on the menu bar produces a menu that allows selection of a weights & measures conversion table, a table of basic formulas, a glossary, interface help, or hints. What hints are presented depend on where the subject is in the problem-solving process, that is, on the active problem

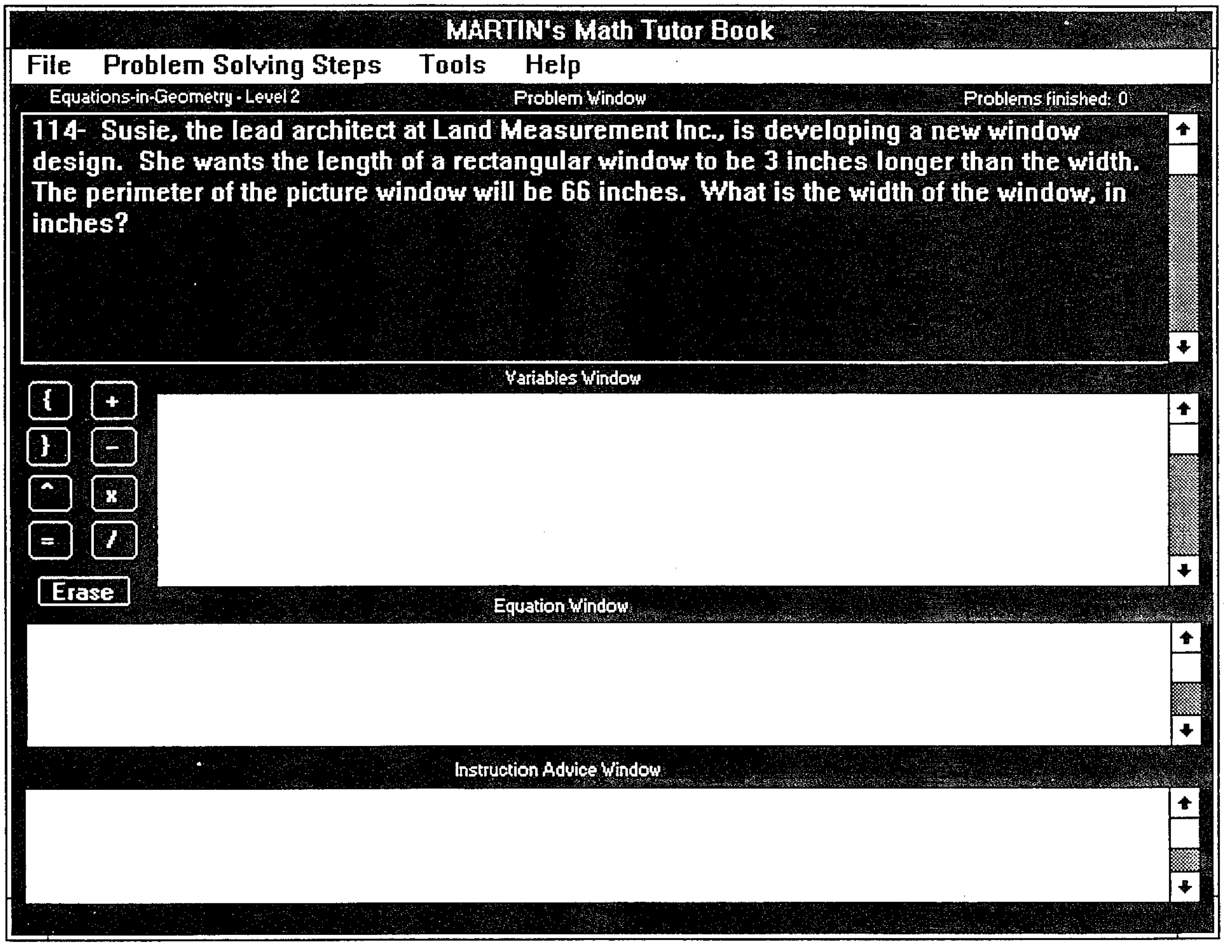

Figure 2. The Basic Environment Screen

solving step and what needs to be done to complete the active step. In addition, repeated requests for hints within a problem-solving step are answered with successively more precise and concrete suggestions. For example, an initial request for a hint during the "Identify Variables" step is answered by the rather nonspecific advice to "reread the question and determine what variables are important for solving the problem". However, if the request is repeated the WPSE suggests that the subject create a variable with a specific name. The system answers another request by presenting the value to be assigned for that variable. Eventually, the system will give the user all the variables needed to solve the problem, and will even suggest a correct equation. The user is free to accept or reject any or all of these hints.

Finally, clicking on "Tools" on the menu bar produces a menu that allows selection of the Notebook and the Plan, two features not used for this study (both are intended for use over an extended period). However, one potentially useful feature was accessed through the Tools menu. The "lesson review" feature allows one to stop working on a problem at any time, jump back to the CBI that begins the module, browse around and find particular information, then return to finish the problem. Subjects were encouraged to use this feature as needed.

The Solver system was developed by altering the WPSE. The two systems share the same curriculum and problem base, CBI, help and hints, and have much the same "look and feel" under many circumstances. The systems are different because Solver provides a more directive, guided approach. For example, Solver requires the user to follow a particular problem-solving approach patterned closely after that illustrated in the CBI, whereas the WPSE is flexible enough to allow virtually any approach that arrives at the correct answer. Differences are most apparent, however, if Solver determines that the user is "foundering". WPSE users have considerable freedom to solve a problem as they choose. The user may first discover an error when he or she performs the Solve Equation step and is told that the answer is incorrect. By contrast, a Solver user has this freedom only to a limited extent. If he or she does not complete a problem-solving step correctly, using no more than a predetermined maximum number of operations, the system intervenes to tell him or her specifically what variables to establish, what values to assign them, or what equation to build, depending on the active problem-solving step and the user's progress within that step. When this happens, the user must follow the system's instructions with regard to problem-solving operations until the step is complete. No other operations can be performed, although he or she can still use the "lesson review" and other help features.

Finally, if a user requires too many total operations to solve a problem, compared to an allowed maximum, Solver requires him or her solve the same problem again. The maximum-operations parameter is determined by multiplying the minimum number of operations required to solve each particular problem by a parameter selected in advance. For example, setting this parameter at 1.6 before the session begins sets the maximum number of operations at 1.6 times the minimum number of operations. The purpose of this is to assure that the user implements the system's preferred strategy at least once with reasonable efficiency for each problem, on the theory that learning complex skills is, in part, a matter of understanding and completing a correct sequence of actions (Singley & Anderson, 1989). The maximum number of operations allowed to complete either a step or an entire problem is determined by setting a system parameter by which the optimal number of operations is multiplied. This parameter was set at 2.3 for the present study, meaning that a subject could not use more than 2.3 times the optimal number of operations without "triggering" Solver either to provide guided tutelage through a step or to require the problem to be worked a second time.

Tutorials -- Each subject was given a tutorial booklet for either the WPSE or Solver, as appropriate. Each subject studied the CBI for the module, then followed the booklet, which led him or her through the process of solving three problems. All subjects worked through the tutorial alone and at their own pace, up to a maximum of three hours. Those who were later assigned to work as group members had not yet been told that they would be assigned to a group.

The same problems were used in the tutorial for each system. The problems were selected from a module on volumes, which was not used again in the study. Each tutorial was comprehensive and pedantic with regard to illustrating the system's features. For example, the process of solving one of the problems was highly elaborated, carefully explaining each step and illustrating important system features such as how to request hints, look up values from the conversion tables, and review the CBI. For WPSE users, the elaborated problem solution also explained an alternative approach to solving the problem. For Solver users, the elaborated solution illustrated how the system becomes directive when a particular problem solving approach is not followed or is not implemented efficiently. Solutions to the other two problems were comparatively straightforward, illustrating how to solve the problems using the minimum number of operations.

Practice Modules -- Subjects were given paper for scratch work and were also encouraged to make notes throughout the practice sessions. Each subject worked on a total of three practice modules, either as an individual or as a group member. Module 1 consisted of 20 problems about percentages, representing seven difficulty levels. Module 2 was actually a combination of two different short modules and included two CBI sessions, one about ratios (five difficulty levels) and one about writing algebraic equations (seven difficulty levels). There were a total of 21 problems in Module 2. Module 3 included 19 problems and covered elementary geometric equations, with problems representing seven difficulty levels. In general, there were two or three practice problems representing each difficulty level within each module. Problems were presented by ascending order of difficulty level, although presentation order within level was randomly determined by the system.

Tests -- All subjects were tested individually. Each subject took a total of three tests, all of which were given using paper booklets. The first, a screening test, was administered before the study proper began. Pilot studies had shown that, on average, more than half the potential subjects in each group recruited by the agencies could solve more than 70 percent of the pretest/posttest problems, leaving little room for either learning or for measuring improvement. We developed a brief screening test and pilot tested it until we had empirically determined performance criteria which, although not perfect, were usually satisfactory for identifying subjects who could not presently work the majority of problems in the problem set, but who were able to learn to solve at least some problems they could not solve initially.

The screening test had three parts. Part 1 consisted of eight fill-in calculation problems in addition, subtraction, multiplication, and division, to test whether prospective subjects could perform very basic mathematical operations accurately. It also included very simple algebraic equations such as solving the equation "$8x = 24$" for x. Part 2 presented a total of five word problems, each of which had been selected from among the Level-1 problems in the modules used in the study. That is, they were similar to the easiest problems that subjects would work with later. Each problem was followed by two multiple-choice questions, so that the maximum score for Part 2 was 10 points. One of the multiple-choice questions asked subjects to select the correct equation to solve the problem from among four alternatives, and the other asked them to select the correct answer to the problem. Part 3 also presented five problems with two multiple-choice questions each, but the problems had been selected from among the middle-difficulty-level problems in their respective modules. Subjects were allowed a maximum of 45 minutes to complete this test. Attached to the screening test was a brief questionnaire on which subjects were asked to give some background information, including gender, ethnicity, and to list all the mathematics and related courses they had taken in high school or college, including computer, statistics, and accounting courses, by title.

In order for a potential subject to qualify for the study, he or she had to answer at least six of the eight Part 1 problems and at least five of the 10 Part 2 problems correctly. He or she could not, however, answer more than three of the 10 Part 3 questions correctly. Note that chance performance on Parts 2 and 3, the multiple-choice parts, was to answer two questions correctly.

Overall, approximately 65% of potential subjects screened were not selected because they had too many correct answers on Part 3, and about 5% more were not selected because they did not correctly solve enough problems on Parts 1 and/or 2. These subjects served in unrelated experiments elsewhere in our laboratory.

There were two forms of the pretest/posttest, arbitrarily labeled "A" and "B" for this discussion. Roughly half the subjects in each group received each form as the pretest and the other form as the posttest. Each of the two forms consisted of 13 medium-difficulty-level problems. The two forms were constructed by selecting a problem from the problem pool and assigning it to one form, then selecting another problem that was equivalent (Reed, Dempster, & Ettinger, 1985) to the first problem and assigning it to the other form. No single problem was used on both test forms or on a test and as a practice problem. Later in this article we will present data concerning the reliability and equal difficulty of these two test forms.

Subjects were provided with scratch paper and calculators for the tests, but were not allowed to use notes or any other supporting materials. There were five multiple-choice questions for each of the 13 problems, and four alternative answers were listed after each question. The first question asked the subject to identify a statement of the goal of the problem. The second question required the subject to distinguish between necessary and extraneous information for purposes of solving the problem. The third question required identification of a correct equation for the problem, and the fourth question required the subject to select the correct answer to the problem. The fifth question asked for the correct label or unit (for example, gallons, miles) for the answer. Figure 3 gives an example of an actual test problem.

Design and Procedure

The study followed a 2 (individual vs. group) x 2 (WPSE vs. Solver) repeated measures (pretest/posttest) design. It was conducted over the course of the first three days of each of seven successive weeks. Subjects completed the screening test as one of a set of first-day intake procedures. About ten subjects typically qualified each week.

Those selected finished the rest of the intake process, then took the pretest, for which they were allowed up to 90 minutes, before leaving for lunch. Meanwhile, subjects had been assigned randomly to a condition simply by sorting screening tests into piles the order in which the tests were handed in, subject of course to the constraint that group conditions required an even number of subjects.

After lunch the subjects logged on to their assigned system and worked through the tutorial. Most subjects spent about two hours, apart from breaks, on the tutorial. Much of this time was spent on the elaborated example problem described earlier.

Bill's car gets 40 miles per gallon of gasoline. He invites John and Tim to travel across the country (about 3,000 miles) with him and they agree to split the cost of gas evenly. They figure gas will average about $1.20 per gallon. What will Tim have to plan for his share of the gasoline?

What is the problem asking you to do?

a. Find how much money Tim will spend on gas.
b. Find the cost of gas for the trip.
c. Find how much Bill will save on gas.
d. Find the average cost of expenses for the trip.

To solve this problem, what information is <u>not</u> important to know?

a. They will cover about 3,000 miles.
b. The cost of gas is $1.20 per gallon.
c. The car gets 40 miles per gallon.
d. The trip will take them across the country.

Which of the following is a correct equation to solve this problem?

a. 3,000 / (40 X 3 X 1.20)
b. (3 X 3,000) / (40 X 1.20)
c. 3,000 / 40 X 1.20 / 3
d. (40 X 1.20) / (3 X 3,000)

What is the correct answer?

a. 30
b. 40
c. 75
d. 90

What are the correct units for the answer?

a. Miles
b. Dollars
c. Gallons
d. Miles per gallon

Figure 3. An example pretest/posttest problem.

Subjects were told at this point that some of them would work alone and some as part of a group. Those working as group members were randomly paired and given instructions which stressed that subjects assigned as partners should work closely together. They were told to rotate working at the computer, in order to eliminate "free rider" effects, and proctors assured that this was done. They were also instructed to help their partners understand how to work the problems if they were the more proficient members of the group, and to ask questions and learn as much as possible from their partner if they were the less proficient members. All subjects were also told that they would be given an individual posttest, similar to the pretest, at the end of the study. The instructions stressed that each subject was responsible for his or her own posttest performance.

Each individual or group spent the bulk of the remaining experimental time working on the three practice modules. The modules were administered in the fixed order described previously. Subjects could work on each module for a maximum of four hours. They began working on Module 1 on the afternoon of the first day, and finished it on the second morning. They also completed Module 2 on the second day. They finished Module 3 on the third day and took the posttest later that afternoon. All subjects in a given week began work on each module at the same time. Subjects who finished a module before the time was up were allowed to leave the lab on break or simply sat quietly at their stations, where they could read books or magazines. Subjects were allowed a ten-minute break at the end of each hour of work, and an hour and a half for lunch each day.

At least one proctor was available at all times to assist subjects. Proctors were allowed to clarify the meaning and intention of both work and test problems if a subject found a problem statement ambiguous, but gave no other help.

RESULTS

In discussing the results, the word "problem" will be used to refer to a complete word problem, while "item" will refer to each multiple-choice question that followed each problem. Finally, "item type" will refer collectively to a particular sort of question. For example, all items which required identification of the correct equation, taken together, constitute an item type.

Pretest Differences

The two test forms yielded comparable pretest results. Scores for subjects given Form A ($\underline{n} = 33$, $\underline{M} = 27.45$, $\underline{SD} = 9.84$), were very close to those for subjects given Form B ($\underline{n} = 32$, $\underline{M} = 27.25$, $\underline{SD} = 10.00$). An independent t-test for the difference between these averages yielded $\underline{t}(63) = .08$, $\underline{p}$(two-tailed) = 0.93.

There were substantial initial differences between experimental groups, despite our attempts at random assignment within the constraints described previously. Pretest means and standard deviations, along with those for the posttest, are given in Table 2.

The pretest differences on overall scores between groups were examined using a 2 x 2 analysis of variance (ANOVA). There was no main effect for either individuals vs. group members, $\underline{F}(1,61) = .591$, $\underline{p} = .445$, or for WPSE vs. Solver, $F(1,61) = .854$, $p = .359$. However, the interaction between

Table 2
Pretest and Posttest Scores and Standard Deviations

Group	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD
Ind/WPSE	24.88	8.09	33.44	8.63
Ind/Solver	31.80	10.90	29.80	8.88
Group/WPSE	27.72	10.70	31.61	11.02
Group/Solver	25.25	8.71	29.94	7.18

these two variables approached significance, $F(1, 61) = 3.80$, $p = .056$. Additional 2 X 2 ANOVAs comparing experimental group differences for each item type showed that this difference in total scores arose chiefly from differences for item Types 1 (specifying the goal) and 3 (identifying the correct equation). Neither of these item types showed a main effect for either the grouping or system variable, but both showed significant interactions between the two. For item Type 1, the interaction $F(1, 61) = 10.59$, $p = .002$, and for item Type 3, the interaction $F(1, 61) = 4.56$, $p = .037$. Observed differences for the other three item types were not significant, although they were generally in the same direction. Table 3 shows pretest and posttest means and standard deviations by item type.

Further investigation revealed, however, that both the overall and item type-specific differences were largely confined to the data from week 7, the "makeup" week. Several subjects during this week had relatively high pretest scores (between 40 and 55). One could make a reasonable case that these subjects were unsuited for the study. It is not clear why they performed poorly on the screening test and were selected for the study, although we repeat our point that the screening test was less than perfect for its intended purpose.

We explored whether eliminating additional subjects, specifically, those scoring 40 or more on the pretest, would eliminate the initial group differences. This cutoff point is arbitrary, but it would discard everyone whose pretest score was more than 1.25 standard deviations above the overall group mean. The result would be to drop one subject from the Individual/WPSE condition, four subjects (not groups) from the group/WPSE condition, three from the individual/Solver condition, and one from the group/Solver condition.

The initial differences between groups appear largely to result from including these subjects. The effect of discarding them would be to eliminate most of the pretest differences between groups, both overall and per item type. For example, with the reduced dataset, there was no overall (across item types) main effect for either grouping, $F(1,52) = 1.26$, $p = .267$, or system, $F(1,52) = 1.49$, $p = .227$, nor was there a significant interaction between these two variables, $F(1, 52) = 0.74$, $p = .393$. Analysis by item type showed that only the item Type 1 interaction between grouping and system was significant for the reduced dataset, $F(1,52) = 6.05$, $p = .017$, an effect for which we have no explanation. The corresponding interaction was not significant for item Type 3, $F(1,52) = 1.31$, $p = .257$. Table 4 gives pretest means and standard deviations for the reduced dataset. We replicated all the pertinent analyses reported in this article using the reduced dataset, but discovered that none of our corresponding conclusions (which are based on posttest differences after pretest differences were statistically controlled) would change as a result of dropping these subjects. After weighing all this, we decided to report results for the full dataset rather than discard additional subjects.

Practice Session Differences

Practice module results for some subjects were lost due to an unrecoverable disk problem on one of the laboratory computers, and results for some others were lost when, for reasons that are not clear, the program apparently failed to write a complete report file. The following comparisons are based on the data from 11 subjects in the individual/WPSE condition, 14 in the individual/ Solver condition, 18 in the group/WPSE condition, and 16 in the group/Solver condition. These results must therefore be regarded with some caution, although we have no reason to suspect that substantial changes would result if the remaining data were available.

Table 5 gives condition means and standard deviations for the number of problems solved correctly for each practice module. There appears to be no characteristic or stable pattern of results across the three modules. For example, subjects in the two group conditions differed appreciably in the mean number of Module 1 problems worked correctly, but worked almost exactly the same mean number of Module 2 problems.

We analyzed these data using separate 2 x 2 ANOVAs for each module. The Module 1 grouping by system interaction was statistically significant, $F(1, 55) = 14.52$, $p < .001$, but neither main effect was significant. Neither main effect nor the interaction was significant for Module 2. For Module 3, the grouping main effect was significant, $F(1, 55) = 6.70$, $p = .012$, and so was the system main effect, $F(1,55) = 9.54$, $p = .003$, but the interaction was not.

Pretest/Posttest Differences

Pretest means and standard deviations by group have already been discussed. Corresponding values for the posttest are also given in Table 2, and are shown graphically in Figure 4. As can be seen, subjects in the individual/WPSE group improved by an average of 8.54 correct items, or about 34.3%, the equivalent of nearly two additional problems. The group/WPSE subjects improved by 3.91 items, or about 14.1%; the group/Solver subjects improved by 4.64 items, or about 18.3%, and the average score for the subjects in the individual/Solver group actually decreased by 2.00 items, a drop of about 6.3%. Because of the initial differences between groups, we examined posttest differences using a Multiple Analysis of Covariance (MANCOVA) (SPSS 6.0, 1992) with pretest scores as a covariate. This analysis showed a significant interaction between the repeated measure and system, $F(1, 61) = 6.94$, $p = .011$, and, more importantly, a significant three-way interaction between the repeated measure, system, and grouping, $F(1, 61) = 9.39$, $p = .003$.

Table 3
Pretest and Posttest Scores and Standard Deviations by Item Type

Type	Group	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD
1. Goal	Ind/WPSE	4.06	2.08	5.75	1.84
	Ind/Solver	5.53	2.07	4.87	1.81
	Group/WPSE	5.39	2.33	4.83	2.45
	Group/Solver	3.69	1.08	4.44	1.59
2. Inform.	Ind/WPSE	6.06	1.73	8.13	3.11
	Ind/Solver	6.86	2.50	7.00	2.48
	Group/WPSE	6.17	2.20	7.28	3.21
	Group/Solver	6.13	2.33	6.38	2.68
3. Equat.	Ind/WPSE	3.75	1.57	5.06	1.77
	Ind/Solver	5.23	2.58	4.60	1.99
	Group/WPSE	4.11	2.03	4.72	2.44
	Group/Solver	3.44	2.00	4.63	2.06
4. Answer	Ind/WPSE	4.13	1.75	5.38	2.39
	Ind/Solver	5.53	2.88	4.47	2.47
	Group/WPSE	4.61	2.06	5.33	2.83
	Group/Solver	4.88	2.36	5.38	1.75
5. Unit	Ind/WPSE	6.88	2.33	9.13	2.13
	Ind/Solver	8.60	3.07	8.87	2.92
	Group/WPSE	7.44	3.60	9.44	2.41
	Group/Solver	7.13	2.78	9.13	2.19

We used dependent t-tests to compare pretest and posttest scores separately for each individual group, in order to identify the source(s) of the overall differences. The between-test increase for subjects in the Individual/WPSE group was significant, $t(15) = 6.17$, $p < .001$, and those in the group/Solver condition also showed significant improvement, $t(15) = 2.93$, $p = .01$. The improvement for the group/WPSE condition was not significant, $t(17) = 1.69$, $p = .110$, and the small decline found for the Individual/Solver group also was not significant, $t(14) = -1.10$; $p = .288$. Item Types 1 (identify goal), 3 (correct equation) and 4 (correct answer) showed much the same pattern as the overall score (refer again to Table 3), that is, a substantial increase for the individual/WPSE group, a small decrease for the individual/Solver group, and moderate increases for both the group/WPSE and group/Solver conditions. Repeated-measures ANOVAs showed significant three-way interactions or interactions that approached significance between the pretest-posttest repeated measure, grouping, and system

Table 4
Pretest Scores and Standard Deviations -- Reduced Dataset

Group	Mean	SD
Ind/WPSE	23.80	7.09
Ind/Solver	23.29	7.28
Group/WPSE	27.92	7.78
Group/Solver	24.00	7.38

Table 5
Practice Module Problems Solved Correctly

	Group	Mean	SD
Module 1	Ind/WPSE	10.82	3.95
	Ind/Solver	13.50	5.13
	Group/WPSE	15.44	2.79
	Group/Solver	10.50	3.22
Module 2	Ind/WPSE	16.55	4.25
	Ind/Solver	13.86	6.61
	Group/WPSE	15.89	3.71
	Group/Solver	15.88	4.15
Module 3	Ind/WPSE	17.27	1.55
	Ind/Solver	15.86	2.38
	Group/WPSE	16.22	1.26
	Group/Solver	13.13	4.43

variables for item Types 1, 3 and 4. For item Type 1, $\underline{F}(1, 61) = 10.13$, $\underline{p} = .002$; for item Type 3, $\underline{F}(1, 61) = 7.34$, $\underline{p} = .009$; while for item Type 4, $\underline{F}(1, 61) = 3.30$, $\underline{p} = .074$. However, only the two-way interaction between the system variable and the repeated measure was significant for item Type 2, identifying unneeded information, $\underline{F}(1, 61) = 4.93$, $\underline{p} = .030$. There were no significant group differences for item Type 5, identifying the unit.

Test Reliability

We conducted an ancillary study to assess the parallel forms reliability of the two pretest/posttest forms. Over the course of 4 weeks, we administered both forms to 39 subjects who were participating

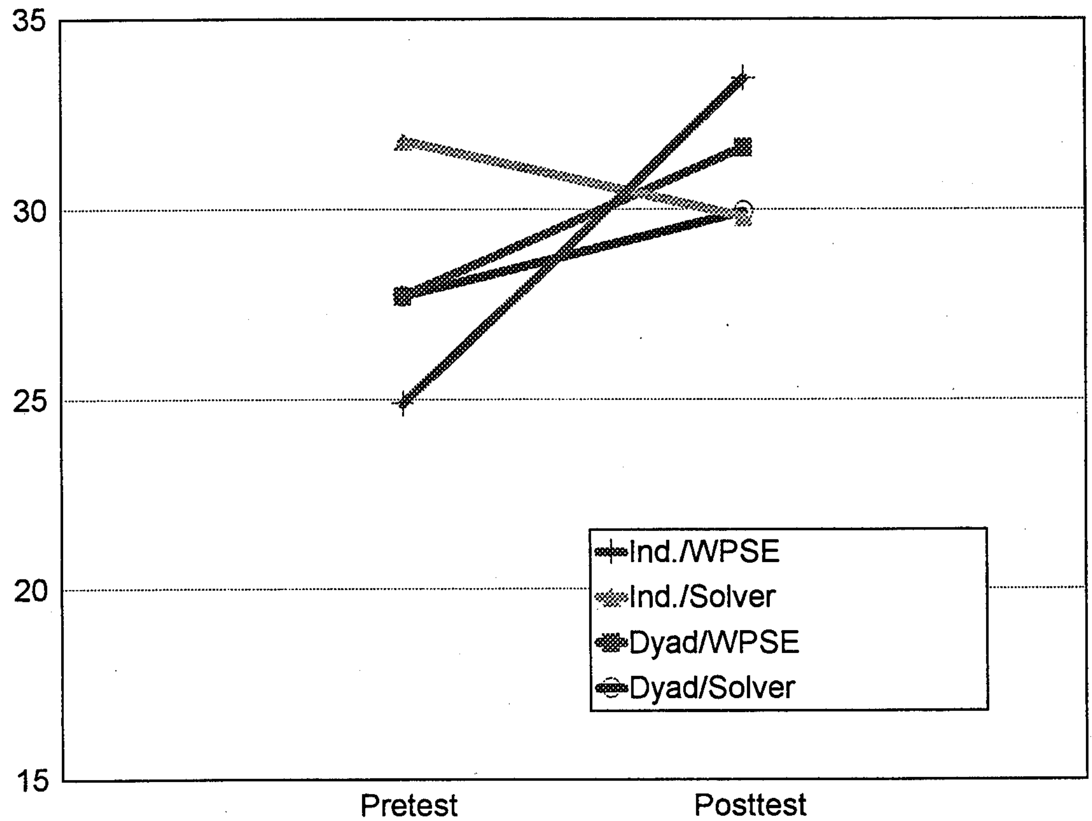

Figure 4. Pretest -- Posttest Differences

in unrelated experiments in our laboratory. These subjects were recruited from the same agencies according to the same criteria as those in the primary study, but were not screened in any way.

One form served as "pretest" and the other as "posttest" for each subject, although there was no mathematics instruction or practice of any sort between the two test sessions. The two forms were counterbalanced such that 19 subjects received Form A as pretest and Form B as posttest, while these roles were reversed for the other 20 subjects. Subjects were given calculators and scratch paper and were allowed up to 90 minutes per test. Proctors offered no help apart from clarifying anything a subject found ambiguous or confusing. After completing one form, subjects performed other tasks or went on break for at least half an hour and not more than an hour and a half, then worked on the other test form. The final sample consisted of 14 Hispanic males, four Hispanic females, two black males, four black females, 12 Anglo males, and three Anglo females.

The overall Pearson product-moment correlation between scores for the two forms was $r(39) = .81$, $p < .001$. The overall mean "pretest" score (that is, across all subjects and across both forms, whichever was administered first) was 44.28, with standard deviation 10.76, while the overall mean "posttest" score was 43.56, with standard deviation 12.15. There were some differences between subjects of different ethnic backgrounds, such that correlations ranged from a low of $r(18) = .77$, $p < .001$ for Hispanic subjects to a high of $r(15) = .89$, $p < .001$ for Anglo subjects.

Thus, the parallel forms reliability for these tests appears to be generally good. In addition, on the average there was essentially no change in scores between test sessions. The mean signed difference between scores (posttest minus pretest) was -.72, with standard deviation 7.13, providing a marker against which to compare effects for the different conditions in the main experiment. A dependent t-test showed that this change was not significant, $t(38) = -.63$, 2-tailed $p = .53$. It is also interesting to note that the average pretest score for this unscreened sample is about twenty points, or approximately two standard deviations, above that for the screened sample in the primary study. Indeed, the unscreened sample's average pretest score is about a standard deviation above the average posttest score for the screened sample, which represents their performance following several hours of instruction and practice. We interpret this as support for our contention that the screened sample represents a relatively homogeneous low-ability remedial population.

The overall pretest/posttest correlation for subjects in the primary study was $r(65) = .63$, $p < .001$. This is considerably lower than the reliability for the ancillary study sample, although one would expect lower correlation between pretest and posttest scores when different treatments result in different amounts of change between groups. On the other hand, it may be that the pretest/posttest correlation is simply lower for low-ability subjects, regardless of treatment.

To examine this possibility, we examined the scores of only those subjects in the ancillary study whose pretest score was 35 or less. There were only 10 subjects in this relatively small sample, but they may resemble more closely most of the subjects in our primary study. The pretest/posttest correlation for this sample was $r(10) = .57$, $p = .087$. Although the correlation between these tests is not high and is not significant, there appears to be little pattern to the differences. The mean posttest score ($M = 32.10$, $SD = 7.67$) was slightly higher than pretest score ($M = 29.9$, $SD = 7.05$). The 2.2 point increase was not significant, according to a dependent t-test ($t(9) = 1.01$, 2-tailed $p = .337$), and can be contrasted with gains for the Individual/WPSE and Group/Solver conditions in the primary study, which were both larger in magnitude and statistically significant.

DISCUSSION

Although we approached this study with few firm predictions, some aspects of the results are still a bit surprising. For example, although we did not necessarily expect to find an advantage for group membership, we did not expect to find superior improvement for subjects who worked alone. Nevertheless, the Individual/WPSE condition clearly yielded the best improvement of any group.

We were particularly unsure what to expect with regard to the system variable. Neither version of the tutor provides a "pure" user-controlled nor program-controlled environment. Even the more stringent Solver allows users much of the control that the WPSE does, so long as they do not "founder". Under most circumstances, users of either system can request hints or review previous instruction whenever they want. Most of the time the main difference between systems is that Solver allows only one method or approach to solving each problem, while the WPSE accepts any approach that yields a correct answer. We suspected that our subjects might benefit from the higher level of control that Solver provides.

Instead, one interpretation that is consistent with our results is that Solver provides too much program control, while the WPSE incorporates a generally beneficial mixture of user and program control. It also appears likely that Solver's interventions when a learner founders are not beneficial. This is consistent with evidence that system feedback is beneficial if and only if it is thoroughly explained (Pridemore & Klein, 1991). When Solver takes over it simply gives feedback without explanation and issues commands which must be obeyed.

Further, it seemed reasonable to expect that the group/Solver condition would produce superior performance, compared to that of the individual/Solver condition, and that the difference would stem from modest gains by the individuals and larger gains by the group members. Instead, we found modest average gains for group members and essentially no gain for individuals who used Solver.

Another possibility is that subjects in the different conditions were initially unequal in ability, and that resulting differences in learning rate either obscured the effects of grouping or actually represent the true effects of grouping. Although we found some initial differences in pretest scores, subjects in the individual/Solver condition showed the best pretest performance. This would imply that they had the highest initial ability, but they showed the no improvement between tests. The results do not likely reflect a "ceiling effect" concentrated in the individual/Solver condition, since the maximum possible score on the posttest was considerably above the average score for any condition. The considerably higher scores for subjects in the ancillary study also provide evidence against a ceiling effect.

Nevertheless, there is a relatively straightforward explanation for this pattern of results, which rests on two assumptions. First, we assume that working with another person counteracted the negative effects that using Solver had on learning. Although we do not know precisely why the Solver system apparently inhibits learning, we suspect that subjects find their lack of control over the system boring and the directive but unelaborated feedback and other system messages unhelpful. Grouping may alleviate boredom and help with generating explanations and understanding of system communications.

Thus, the superiority of groups using Solver, relative to individuals using Solver, may represent a true positive grouping effect which might not be specific to low-ability subjects.

Our second assumption is that homogeneous our groups of low-ability subjects who worked with the WPSE interacted ineffectively and weren't able to be helpful to each other. This notion is to some extent consistent with points made by Nastasi and Clements (1991) and Hooper and Hannafin (1989), which were mentioned earlier in this article. Still, these researchers' comments lead one to expect that low-ability cooperative groups will perform no better than individuals, but not necessarily worse than individuals. Our results indicate a possibility that low-ability cooperative group members may have hindered each other's learning. We might have anticipated this. Webb (1987) discussed the possibility that a "...detrimental effect is that students may provide each other with ineffective or incorrect explanations that steer each other wrong" (p. 203). In sum, it seems that whether grouping is beneficial or detrimental may depend on factors intrinsic to the design and quality of the CBI system.

It seems surprising at first that these effects were not related to the number of practice module problems solved correctly. In general, subjects who work more practice problems correctly should reasonably learn more and work more test problems correctly. Because the tests consisted of items representing each module, the most likely predictor of test performance might appear to be the overall score summed across all three practice modules. The most likely explanation rests on the fact that the tests consisted of middle-difficulty problems. Subjects only needed to complete the middle-difficulty problems in each module to in order to learn enough to work the test problems. Beyond this point, progress through the module involved working harder problems, which produced no additional measured benefit.

REFERENCES

Avner, R. A., Moore, C., & Smith, S. (1980). Active external control: A basis for superiority of CBI. Journal of Computer-Based Instruction, 6, 115-118.

Dalton, D. W. (1990). The effects of cooperative learning strategies on achievement and attitudes during interactive video. Journal of Computer-Based Instruction, 17, 8-16.

Gay, G. (1986). Interaction of learner control and prior understanding in computer-assisted video instruction. Journal of Educational Psychology, 78, 225-227.

Gray, S. H. (1987)j. The effect of sequence control on computer-assisted learning. Journal of Computer-Based Instruction, 14, 54-56.

Hicken, S., Sullivan, H., & Klein, J. (1992). Learner control modes and incentive variations in computer-delivered instruction. Educational Technology Research and Development, 40, 15-26.

Hooper, S. (1992). Effects of peer interaction during computer-based mathematics instruction. Journal of Educational Research, 85, 180-189.

Hooper, S., & Hannafin, M. J. (1989). Cooperative learning and computer-based instruction: Achievement, interaction, and performance. Proceedings of the 31st international conference of the Association for the Development of Computer-Based Instructional Systems (pp. 279-286). Washington, DC: ADCIS.

Hooper, S. Temiyakarn, C., & Williams, M. D. (1993). The effects of cooperative learning and learner control on high- and average-ability students. Educational Technology Research and Development, 41, 1042-1629.

Hooper, S., Ward, T. W., Hannafin, M. J., & Clark, H. T. (1989). The effects of aptitude composition on achievement during small group learning. Journal of Computer-Based Instruction, 16, 102-109.

Johnson, D. W., & Johnson, R. T. (1989). Cooperation and competition: Theory and research. Minneapolis, MN: Interaction Book Company.

Kinzie, M. B. (1990). Requirements and benefits of effective interactive instruction: Learner control, self-regulation, and continuing motivation. Educational Technology Research and Development, 38, 1-21.

Kinzie, M. B., Sullivan, H. J., & Berdel, R. L. (1988). Learner control and achievement in science computer-assisted instruction. Journal of Educational Psychology, 80, 299-303.

Lee, S. S., & Lee, Y. H. K. (1991). Effects of learner-control versus program control strategies on computer-aided learning of chemistry problems: For acquisition or review? Journal of Educational Psychology, 83, 491-498.

Lee, S. S., & Wong, S. C.-H. (1989). Adaptive program vs. learner control strategy on computer aided learning of gravimetric stoichiometry problems. Journal of Research on Computing in Education, 21, 367-379.

Mevarech, Z. R. (1991). Learning mathematics in different mastery environments. Journal of Educational Research, 84, 225-231.

Morrison, G. R., Ross, S. M., & Baldwin, W. (1992). Learner control of context and instructional support in learning elementary school mathematics. Educational Technology Research and Development, 40, 5-13.

Nastasi, B. K., & Clements, D. H. (1991). Research on cooperative learning: implications for practice. School Psychology Review, 20, 110-131.

Park, O., & Tennyson, R. D. (1983). Computer-based instructional systems for adaptive education: A review. Contemporary Educational Review, 2, 121-135.

Pridemore, D. R., & Klein, J. D. (1991). Control of feedback in computer-assisted Instruction. Educational Technology Research and Development, 39, 27-32.

Reed, S. K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 106-125.

Ross, S. M., & Rakow, E. A. (1981). Learner control versus program control as adaptive strategies for selection of instructional support on math rules. Journal of Educational Psychology, 73, 745-753.

Rubincam, I., & Olivier, W. P. (1985). An investigation of limited learner-control options in a CAI mathematics course. AEDS Journal, Summer, 211-226.

Shebilske, W. L., Regian, J. W., Winfred, A., & Jordan, J. A. (1992). A dyadic protocol for training complex skills. Human Factors, 34, 369-374.

Singley, M. K., & Anderson, J. R. (1989). The transfer of cognitive skill. Cambridge, MA: Harvard University Press.

Slavin, R. E. (1990). Cooperative learning: Theory, research and practice. Englewood Cliffs, NJ: Prentice Hall. SPSS for Windows [Computer program] (1992). Chicago, IL: SPSS, Inc.

Steinberg, E. R. (1977). Review of student control in computer-assisted instruction. Journal of Computer-Based Instruction, 3, 84-90.

Steinberg, E. R. (1989). Cognition and learner control: A literature review, 1977-1988. Journal of Computer-Based Instruction, 16, 117-121.

Steinberg, E. R., Baskin, A. B., & Hofer, E. (1986). Organizational/memory tools: A technique for improving problem solving skills. Journal of Educational Computing Research, 2, 169-187.

Steinberg, E. R., Baskin, A. B., & Matthews, T. D. (1985). Computer-presented organizational/memory aids as instruction for solving Pico-Fomi problems. Journal of Computer-Based Instruction, 12, 44-49.

Tennyson, R. D. (1980). Instructional control strategies and content structure as design variables in concept acquisition using computer-assisted instruction. Journal of Educational Psychology, 72, 525-532.

Tennyson R. D., & Rothen, W. (1979). Management of computer based instruction: Design of an adaptive control strategy. Journal of Computer-Based Instruction, 5, 63-71.

Tennyson, C. L., Tennyson, R. D., & Rothen, W. (1980). Content structure and instructional control strategies as design variables in concept acquisition. Journal of Educational Psychology, 72, 499-505.

Toolbook 1.0 [Computer Program] (1989). Bellevue, WA: Asymetrix Corporation.

Webb, N. M. (1987). Peer interaction and learning with computers in small groups. Computers in Human Behavior, 3,193-209.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. Journal for Research in Mathematics Education, 22, 366-389.