ARI Research Note 88-70

DTIC FILE COPY

AD-A199 022

Studies of Diagnosis and Remediation with High School Algebra Students

R. Martinak, D. Sleeman, A.E. Kelly, J. Moore, and R. Ward

Stanford University

for

Contracting Officer's Representative
Judith Orasanu

ARI Scientific Coordination Office, London
Milton S. Katz, Chief

Basic Research Laboratory
Michael Kaplan, Director

U. S. Army
Research Institute for the Behavioral and Social Sciences

July 1988

Approved for the public release; distribution unlimited.

U. S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the

Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

Stanford University

Technical review by

Steve Kronheim

DTIC
COPY
INSPECTED
3

Accession For	
NTIS GRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

This report, as submitted by the contractor, has been cleared for release to Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or other reference services such as the National Technical Information Service (NTIS). The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

Introduction

"Classical" CAI systems have been used to provide tutorial instruction and socratic or supportive problem solving. Tutorial systems aim to diagnose a student's errors and then to provide appropriate remediation. Supportive problem solving systems monitor the student's problem solving, and aim to provide help and advice whenever requested. The subfield of Intelligent CAI, or ITS (Intelligent Tutoring Systems), arose because workers felt that CAI was intrinsically limited, and in fact incapable of providing highly adaptive instruction (Hartley & Sleeman, 1973; Sleeman & Brown, 1982). The working hypothesis of the ITS field has been that to produce a highly successful tutorial system requires the ability to infer an accurate student model, and that it will then be relatively straightforward to use the model to direct a remedial dialogue. Remediation based on a student model is referred to in this paper as error-specific or model-based remediation (MBR). MBR provides feedback about specific error(s) in the student's procedure before reteaching a correct strategy. Its counterpart, Reteaching, simply reteaches the correct method.

The PIXIE system

PIXIE is a data-driven ITS shell (Sleeman, 1987), which attempts to diagnose and then remediate student errors. Knowledge about a particular domain is contained in the knowledge base. (At this point, knowledge bases for linear algebra, precedence in arithmetic, negative numbers, and fractions exist; the linear algebra database is the most complete.)

Each knowledge base includes:

- a set of (correct) domain rules

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS - -
2a. SECURITY CLASSIFICATION AUTHORITY - - 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE - -	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) - -	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 88-70
6a. NAME OF PERFORMING ORGANIZATION Stanford University 6b. OFFICE SYMBOL (If applicable) - -	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code) Palo Alto, California	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Same as 7a. 8b. OFFICE SYMBOL (If applicable) - -	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-84-K-0279
8c. ADDRESS (City, State, and ZIP Code) Same as 7b.	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: 6.11.02.B PROJECT NO.: 2Q161102B74F TASK NO.: n/a WORK UNIT ACCESSION NO.: n/a

11. TITLE (Include Security Classification)
STUDIES OF DIAGNOSIS AND REMEDIATION WITH HIGH SCHOOL ALGEBRA STUDENTS

12. PERSONAL AUTHOR(S)
R. Martinak, D. Sleeman, A.E. Kelly, J. Moore, and R. Ward

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim Report	FROM 6/84 TO 9/87	July 1988	100

16. SUPPLEMENTARY NOTATION Judith Orasanu, contracting officer's representative. Work performed by University of Aberdeen, AB9 2UB, Scotland as subcontract from Stanford University.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Intelligent Tutoring System; Remediation; Artificial Intelligence; Tutoring Systems; Computer-Aided Intelligence

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

This research note compares the effects of reteaching and different styles of error-based remediation. More research is needed to understand the factors which lead to successful remediation. Six studies discussed in this RN compare error-specific or model-based remediation (MBR) with reteaching in algebra. The results show that MBR and reteaching are both more effective than no tutoring, but MBR is not clearly more effective than reteaching. The results are discussed both in terms of stability of errors, and of their relevance to educational practice and the field of intelligent tutoring systems.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Judith Orasanu	22b. TELEPHONE (Include Area Code) 202/274-8722 22c. OFFICE SYMBOL BR

DD Form 1473, JUN 86 *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

- a set of incorrect variant rules
- sets of examples that are sufficient to discriminate between the models generated

The system generates a series of models, that include both the correct and all variants of the incorrect rules, in an offline phase. When the student works at the terminal, the system merely has to decide which, if any, of the set of predefined models fits the student's current answer. For example, if the student was presented with the task $3x + 5x = 9$ and the student gave $x = 8/9$ as a response, one explanation would be that the student had successfully added the two x-terms to get $8x = 9$, but then "inverted" the answer in the final step. For more details, see Moore and Sleeman (1987). As noted earlier, the assumption made by the ITS subfield has been that it would be relatively straightforward to remediate errors once they had been accurately diagnosed, and that MBR would be more effective than Reteaching. A further assumption is that students have (fairly) stable mental models for the task domain, and that these would be "used" consistently by the students.

The series of studies reported in this paper has attempted to empirically verify these assumptions, i.e., to test the null hypothesis of no difference between MBR and Reteaching.

Review of the literature on the learning of algebra

Diagnosing algebra errors was investigated quite extensively by Buckingham (1933), and others; these early investigators studied the types and frequencies of errors found in groups of students. Buckingham (1933) suggests two uses for the results from these studies:

> First, the findings may be utilised in planning the group teaching procedure. The teacher can guard against the errors which occur

> with greatest frequency and persistency by providing specific drills for the class as a whole in examples in which these errors are likely to occur
>
> Second, the findings of the investigation suggest the importance of attention to the individual needs of students. While, in a general way, certain types of errors predominate, there is no assurance that any particular student will make errors either in kind or frequency that correspond with those resulting from this study. Therefore, the second method of teaching would be to use tests of the type employed in this study to determine the difficulties of each student. Each student can then be provided with specific drill necessary to overcome the errors, (pp. 101 & 102).

Several reviews, three* of which are quoted below, indicate that error-specific remediation is known to be superior to reteaching:

> Brown and Burton foresee a time when schools might have a diagnostic specialist who would work with children having special difficulty in math. This diagnostician would conduct in-depth interviews in conjunction with specific computational tasks to detect possible procedural errors. The intent would be to gear instruction to specific procedural difficulties of an individual. (In Resnick and Ford, 1981, p. 88).

Secondly, Resnick (1984) also addresses the issue of what instructional response is appropriate for errors that are systematic. Although her investigation also centered on subtraction errors, her speculations seem appropriate for algebra errors. Resnick (1984) provides the following use of diagnosing systematic errors for instruction.

> ".... it may prove useful to tailor practice to specific kinds of buggy rules, either by choosing particular examples that are matched to a child's errors or by giving special attention to the parts of a procedure that evoke those errors." (p.13)

However, Resnick emphasises that one may derive the most benefit by determining more global misunderstandings rather than remediating each error. Again, however, this speculation is not supported by data.

*Although Sleeman agrees he has made similar claims orally, we are at present unable to find an appropriate written quotation.

Thirdly, Macnab and Cummine (1986) discuss the importance of showing the student both the correct procedure and pointing out the incorrect steps. They state:

> ".... demonstrating that there is a flaw in a pupil's method can be useful in situations where the pupil is aware of a correct procedure but prefers his alternative either because he thought it up himself, or it seems easier, or for some other reason. In such cases the unsound nature of the pupil's own method may have to be demonstrated before he will adopt a correct method" (p.125).

The following studies, report experimental results on the <u>effectiveness</u> of various remedial strategies. Swan (1983) concluded that error-specific remediation which involved an external validity check is more effective than Reteaching over a period of eight one-hour lessons. Unlike Swan (1983), Bunderson and Olsen (1983), suggest that pointing out an incorrect procedure is <u>not</u> more effective than merely reteaching under some circumstances:

> Thus in the area of subtraction at least, both error-specific and general attempts at remediation for students who had previously been taught produced excellent results. Interpreted in the light of error instability, it seems that any attention to correct a subtraction error by teaching the correcting procedure will enable students to call upon their procedure and use it correctly in a posttest given at the end of their remedial period. (Bunderson and Olsen, 1983, p.11)

The recent PIXIE studies (Kelly, Martinak, & Sleeman, 1987; Martinak, Schneider, & Sleeman, 1987) cast serious doubt on the need for a diagnostic specialist. In addition, Putnam (1987) and Kelly and Sleeman (1986) found that teachers generally do not adopt the role of a diagnostician, even when in a tutorial situation. Diagnosing errors tends not to be the teacher's primary goal. Putnam suggested that teachers in their remediation followed a "curriculum-based-script". However, Putnam did not compare the effectiveness of tutoring based on a curriculum script with, say, an MBR approach. Given the conflicting results reported above, more research is obviously needed to systematically compare error-specific remediation (or model-based remediation - MBR) with Reteaching.

The rest of the paper reports on the development of the RPIXIE system and on the detailed series of experiments carried out to address these issues.

I. Development of the Computer-based PIXIE

A. Student Protocols: PIXIE's Diagnostic System

Protocols of students working various types of mathematics tasks were analysed to determine the types of errors made, and these were used to create a knowledge-base for PIXIE. The domains investigated by the PIXIE Project include: linear algebra, fractions, precedence in arithmetic, and negative numbers.

II. Pilot Studies

A. Revisions to PIXIE (Diagnostic System):

In 1984 and early 1985, PIXIE (only a diagnostic system at this point) was field tested in several schools. Overall, the reaction to PIXIE was favourable; however, many students did not type complete algebraic equations at each step (a requirement of the first version of PIXIE). PIXIE was subsequently revised to include an alternative "interface" that requires only the final response, and not intermediate steps, be valid algebraic expressions. During these field trials, mal-rules that were not encountered in earlier work were noted and subsequently added to PIXIE's rule-base.

B. Teacher Protocols: PIXIE's Remedial subsystem.

With the intention of building a remedial sub-system in PIXIE, several studies of teachers performing tutoring were undertaken (Kelly & Sleeman, 1986). In the first study, four experienced teachers were asked to diagnose and suggest remediation for a series of equation-answer pairs, which were based on actual student protocols. In the second study, an experienced teacher remediated eight students, based on diagnostic information provided by the PIXIE system. Results from this series of studies indicated that:

a) Teachers frequently do NOT make a very detailed diagnosis.

b) Remediation is by and large very procedurally orientated - few teachers stress conceptual understanding.

c) Remediation frequently takes the form of reteaching.

C. Design of RPIXIE

On the basis of teacher protocols, an embryonic remedial system was implemented. The remedial system provided two basic types of remediation: Model Based Remediation (MBR) and Reteaching. MBR points out and labels as incorrect specific error(s) made by the student before presenting a correct method for solving the task. Reteaching merely presents a correct method for solving a task without pinpointing error(s). The following are examples of MBR and Reteaching:

An Example of a Student Interaction: MBR

(This protocol is slightly edited)

```
   Solve (f = Finished, q = Quit early)
   7x = 2 * 2
S: 7x = 4
S: f

   Simplify further 7x = 4
S: x = 4
S: f

   Thank you,
   but you didn't get that one right.
   We will now look at that task again.

   Here we are considering the task 7x = 2 * 2

   Did you work out the task something like this?:

   7x = 2 * 2
   7x = 4
   7x/7 = 4
   x = 4
   Please type YES or NO: y (Student replies)

   It seems that you changed 7x = 4 to something like 7x/7 = 4

   This is WRONG.
   Do NOT only divide ONE side by the coefficient.
   You should divide both sides of the equation by  the  coefficient  to
   change
   7x = 4 to 7x/7 = 4/7

   Now let me show you how you SHOULD have worked the whole task
   7x = 2 * 2          multiply 2 numbers
   7x = 4              divide both sides of the equation by  the  coeffi-
   cient
   7x/7 = 4/7          divide
   x = 4/7             Finished
```

An Example of a Student Interaction: Reteaching

(Same task; again the protocol has been edited slightly and the initial student interaction has been omitted.)

Thank you,
but you didn't get that one right.
We will now look at that task again.

Now let me show you how you SHOULD have worked the whole task

7x = 2 * 2	multiply 2 numbers
7x = 4	divide both sides of the equation by the coefficient
7x/7 = 4/7	divide
x = 4/7	Finished

D. Revisions to RPIXIE (the diagnostic and remedial systems) as a result of a pilot study:

RPIXIE (the remedial version of PIXIE) was taken to several schools to investigate whether students could use it productively. The initial version of RPIXIE required the student to work six tasks of the same format before receiving feedback about the correctness of any item. If at least 4 of the 6 tasks were worked consistently using the same mal-rule(s), the student received MBR on those tasks; for the other tasks in the set, the student received Reteaching. If tasks had been incorrectly worked, but with less consistency, then the student received Reteaching for all 6 items. However, if MBR was chosen, the students were required to recall how they worked each task; students often could not remember. On some occasions, the MBR trace was interpreted as the <u>correct</u> procedure by some students. Results from these field studies suggested the following revisions to RPIXIE:

a) Require the student to work fewer items of the same format, and provide a mode in which it is necessary to work only a single task before feedback and, if necessary, remediation is given.

b) Shorten the text students are required to read. Provide definitions for technical terms (e.g., coefficient).

c) Highlight the MBR trace so that it is more obvious to the student that it represents an incorrect method.

d) Provide two algebraic cancelling methods: one approach should use "move and change the sign", the other "cancel by doing the same thing to both sides".

e) Allow the student to select from a series of remedial models. Initially, RPIXIE remediated only when a single model fitted the student's response; the issue of which of several models to present to the student has subsequently been addressed (Moore & Sleeman, 1987).

These changes, with the exception of (c), were implemented.

FORMAL EXPERIMENTS

When a workable diagnostic and remedial system was in place, a series of studies to investigate its effectiveness were carried out. PIXIE was developed under the assumption that diagnosing a student's error(s) and specifically remediating the error(s) before showing the correct procedure (MBR) was educationally more beneficial than simply showing the student the correct procedure (Reteaching). The series of studies discussed below were carried out to test this assumption.

Six studies were conducted: three studies used the computer as a tutor and the other three studies used humans as tutors. All six studies were based on a pretest - intervention - posttest design. Some of the studies reported below have fewer students than we had hoped for because only a subset of the pretested students would need tutoring, or in the case of a later study, only a subset had stable errors. When confronted with the number of qualifying students, we had to decide, for example, whether to increase the number of students in each condition at the expense of the control group. This decision was guided by the major hypothesis being tested.

The studies are summarized in (approximate) chronological order so that the reader may gain a better understanding of the issues as they arose.

I. Initial Computer Based Remediation:

In this first study (California, spring 1986), the effectiveness of different forms of remediation was investigated by using an intelligent tutoring system (RPIXIE). The null hypothesis was that there would be no difference in performance among three conditions of tutoring: MBR, Reteaching, and Evaluation (simple knowledge of performance).

RPIXIE's Algebra Data Base

RPIXIE's algebra data base consists of 17 algebra task-sets that vary in difficulty from items of the form ax=b to items of the form ax=b*c(dx+e). In this study, three of RPIXIE's presentation modes (forms of remediation) were used:

(1) MBR: remediation in which students were shown what they had done wrong, and then presented with a correct solution-trace.

(2) Reteaching: remediation in which students were merely informed whether their responses were correct, and then presented with a correct solution-trace.

(3) Evaluation: remediation in which students were informed only whether items were answered correctly or incorrectly, with no

correct solution-trace shown.

Students

Two experiments were run, both with identical procedures. The first experiment involved 15 ninth- and tenth-grade algebra students from a high-achieving mathematics class (standardised mathematics achievement test scores were at or above tenth-grade level). In the second experiment, 24 students from a low-achieving algebra class were tested (test scores were below the eighth-grade level). The students came from a high school in the San Francisco area (School C).

Materials

Prior to the experiment, data were collected on the mathematics section of the Standardised Test of Educational Progress (STEP), and on the Mathematics Attitude Inventory (MAI) - a survey containing six subscales relating specifically to mathematics: attitude towards the teacher, motivation, anxiety, enjoyment, confidence, and value of mathematics.

Materials also included a 17-item pretest and 17-item posttest. Both tests contained algebra tasks that were similar to those worked individually with RPIXIE. Tasks on the pretest and posttest were matched for difficulty, by using the same templates (e.g., ax+b=c), to generate tasks for both tests. A four-item questionnaire reviewing RPIXIE was also developed.

Procedure

Students in the classes took a group pretest and the MAI during the regular class period. The week following the tests, students were randomly assigned to condition: MBR, Reteaching, or Evaluation. Very little researcher/student interaction occurred while the student individually worked with RPIXIE. If the student did not understand the instructions, they were explained by the researcher. All students began at the easiest level (Task Set 1) and worked until all 17 task-sets were completed or 45 minutes had elapsed - whichever was sooner. Students were then given 5 minutes to answer the four-item questionnaire. After all the students had interacted with RPIXIE, a group posttest was administered during the regular mathematics period.

Results

Effectiveness of feedback was measured by the posttest, by the number of items correctly answered on RPIXIE, and by the percentage of items attempted on RPIXIE that were answered correctly. See Table 1a and Table 1b for mean scores by condition for each group.

Table 1a. High Achieving Group

	Pretest*	Posttest*	Mean number of items worked on RPIXIE per student	% correct on RPIXIE
MBR	13.00 (5.35)	13.20 (6.63)	56.00 (3.00)	81
Reteaching	14.60 (1.50)	15.20 (1.30)	53.20 (2.82)	91
Evaluation	15.60 (1.34)	14.40 (2.30)	56.20 (7.82)	86

* Maximum = 17, standard deviations are shown in parentheses. N = 15.

Table 1b. Low Achieving Group

	Pretest*	Posttest*	Mean number of items worked on RPIXIE per student	% correct on RPIXIE
MBR	9.25 (4.03)	10.63 (3.30)	51.63 (4.31)	84
Reteach	10.29 (3.35)	12.13 (1.55)	51.75 (8.00)	86
Evaluation	8.14 (2.73)	10.00 (3.02)	49.75 (14.68)	67

* Maximum = 17, N = 24

Results by Condition

No significant differences among the conditions was found for the high-achieving group. For the low-achieving group the results were similar. We found no significant differences by condition (Hotelling's $t(14) = 1.77$, $p<.11$), although the students in the Evaluation condition appear to have worked fewer items correctly on the computer. This may be because of poor motivation due to not receiving remedial support for wrong answers. (Note: the apparent heterogeneity of variance for the mean number of items worked on RPIXIE is not of concern, because the differences between the means do not appear to be of practical consequence.)

RPIXIE's diagnostic power

The data were analysed to determine the percentage of errors that RPIXIE diagnosed. Specifically, we asked: 1) Did PIXIE have a model that the researchers believed explained the student's answer, and 2) Did the student agree that RPIXIE's model was acceptable? Under the first criterion, RPIXIE diagnosed approximately 33% of the errors made in the MBR condition; the students "approved" 91% of these, resulting in 30% of errors being followed by MBR.

Discussion

The finding of no significant differences between the MBR and Reteaching conditions may be due to RPIXIE's low diagnostic rate (70% of the students' errors in the MBR condition were followed with Reteaching). This rendered the two groups more similar than dissimilar because the default is Reteaching. This low rate of diagnosis arose because of the limited number of mal-rules used in RPIXIE, which in turn arose for a variety of technical reasons that have now been substantially overcome.

II. Human Remediation: A Study to Compare MBR and Reteaching

The above study led us to believe that the issues of diagnosis and remediation were much more subtle than initially suspected, we decided to replicate the study using human tutors. The first such experiment was carried out in the Autumn of 1986 at an Aberdeen school, School L. The null hypothesis was that there would be no difference in performance between two conditions of tutoring: MBR and Reteaching.

Subjects: Students from two 2nd- and two 3rd-year mathematics classes in School L participated in the study. Average ages of the students were 13 years 4 months and 14 years 6 months, respectively. On the basis of an analysis of each student's pretest, a subset of 44 students were selected for individual tutoring.

Materials: Materials included a 20-item pretest, a 20-item posttest, and scripts for remediation. As in the previous study the same templates were used to generate the items on both tests. See Appendix A for the pretest and posttest items.

Tutoring scripts based on RPIXIE's approach to remediation were developed. Separate scripts were written for MBR and Reteaching. (See Appendix B for sample scripts.) The MBR script directs the tutor to point out to the student each error made and to explain it before the correct procedure is retaught. The Reteaching script merely directs the tutor to reteach the procedure. Items to be worked during tutoring were written prior to the sessions so as to ensure uniformity across the several tutors, i.e., students making an error in a given item were given the same remedial tasks irrespective of the tutor. Each tutor was trained with these scripts. (See Appendix C for the items used at each level of tutoring.)

Procedure: All students in the study took a 40-minute group pretest during their regularly scheduled mathematics class. A subset of these students were randomly assigned to MBR or Reteaching. Each student was individually tutored for approximately 35 minutes. Tutoring occurred the week following the pretest, and all sessions were completed within a week. All tutoring sessions were audio-taped. Tutoring consisted of having the student first rework an item marked as incorrect on the pretest. If the item was again worked incorrectly, remediation appropriate to the condition was given and the student worked at most two practice items of the same type. This procedure was repeated for each item scored as incorrect on the pretest.

The week following tutoring, a group posttest was given to the students during a regular mathematics class. A delayed posttest, identical to the immediate posttest, was given approximately two months after the first posttest.

Results:

Analyses of the data are based on scores from 38 of these students (because 6 were absent from school for the tutoring or the posttest). Posttest scores were taken as a measurement of the effectiveness of tutoring. Table 2 presents the mean scores by condition; standard deviations are given in brackets.

Table 2. Mean Scores* by Condition

Condition	Pretest	Posttest 1	Posttest 2
MBR	12.53 (3.86)	14.32 (3.37)	12.24 (5.62)
Reteaching	11.63 (3.10)	14.42 (3.64)	12.76 (4.44)

*Maximum = 20; N = 3

Although there was a significant overall mean difference between the pre- and posttests for both groups $t_{27} = 4.20$, $p < .001$, there was no significant difference by condition: $t_{36} = .09$, $p > .92$. The overall mean scores for the delayed posttest (two months from the first posttest) were also not significant by condition $t_{32} = .30$, $p > .70$, and had reverted to pretests levels.

Further analyses of the data using only students who were poor on the pretest (i.e., those scoring 13 or less) showed no significant differences by condition, $t_{19} = 0.43$, $p > .66$. Errors were classified into several levels of severity, but again no correlation was found between the number of "severe" errors made and the tutoring condition.

Discussion

This experiment confirmed the previous computer study results and showed that MBR and Reteaching are very comparable, even with low-scoring students. We interpreted this to imply that the form of MBR used was not effectively communicating with the student. Study III attempted to make MBR more effective.

Subsequent studies do not attempt to categorise the severity of errors as we did here, because of the problems involved with the classification. Definitional difficulties were encountered because we did not have information pertaining to *why* students made particular errors. For example, is an inverted division (i.e., $ax = b \Rightarrow x = a/b$) a careless error, an algebric misunderstanding, or a severe misunderstanding of fractions?

III. Human Remediation: A study to compare variants of MBR and Reteaching

A study was run to investigate two factors that may have influenced the results from the above study: namely presumed lack of cognitive dissonance and lack of cognitive engagement on the part of the students. It was hypothesized that no differences between the conditions in the previous study had arisen because:

a) The students had no reason to accept the tutor's method of doing algebra as better than their own. Macnab and Cummine (1986) discuss the importance of demonstrating to the student the unsound nature of the pupil's incorrect method, i.e., create "cognitive dissonance" (CD). In the present study, we attempted to instil cognitive dissonance by having students check their answer by substituting it back into the equation to see if both sides of the equation balanced.

b) The students were not sufficiently involved with their learning (i.e., they were passive listeners to the tutor's instructions). By having students verbally repeat the correct procedure back to the tutor, we hoped to engage them more in their learning.

Procedure: Students from two 2nd-and one 3rd-year mathematics classes (average age 13 years 6 months and 14 years 8 months, respectively) from a different Aberdeen secondary school (School P) took a 20-item pretest in algebra equation solving. Anyone scoring 80% or better on the pretest was not seen for tutoring. Thus, a subset of 48 students were randomly assigned to one of four conditions:

a) MBR

b) Reteaching

c) MBR with cognitive dissonance (MBR + CD)

d) MBR with cognitive engagement (MBR + CE)

Given that there were a limited number of students to work with, the number of students per condition was increased at the expense of a pretest/posttest-only, Reteach + CD, and Reteach + CE conditions, in order to better test the differences among the above four conditions. (Note that the MBR and Reteaching conditions used in the previous study were designed into this study, so that the previous study could be replicated.) See Appendix D for samples of the scripts used in each of the four conditions.

Each student was individually tutored for approximately 35 minutes. All tutoring sessions were audiotaped. Tutoring consisted of having the student first rework an item missed on the pretest. If the item was again worked incorrectly, the student received remediation appropriate for the treatment condition and worked at most two more practice items of the same format. To discourage a student in the Reteaching condition from making comparisons between the incorrect and correct procedure, the student's workings of the incorrect procedures were taken away before the correct procedure was taught. After each of the students were tutored, group posttests were given to all students in the classes involved.

Results:

Both quantitative and qualitative analyses were performed on the data.

Results from an ANOVA indicated no significant differences among conditions on the pretest: $F_{3,44} = .626$, $p > .50$, (see Table 3 for means).

Table 3. Pretest, Posttest, and Delayed Posttest mean scores by condition

Condition	Mean Pretest Scores*	Mean Posttest Scores*	Delayed Posttest Scores*
Reteach	12.08 (1.98)	15.20 (2.92)	15.64 (3.07)
MBR	11.91 (3.33)	15.80 (3.68)	14.73 (2.05)
MBR+CD	12.60 (2.32)	14.00 (2.63)	14.80 (1.93)
MBR+CE	12.91 (2.81)	15.90 (3.33)	13.00 (1.80)

* Maximum score = 20. N = 48; standard deviations are given in brackets.

The overall mean for the posttest score was significantly higher than the overall mean for the pretest ($t_{43} = 5.83$, $p < .001$) showing a general pre- to posttest gain. An ANOVA on posttest scores showed no significant differences among the conditions: $F_{3,40} = .797$, $p > .50$. An ANOVA on the delayed posttest scores also showed no differences by condition: $F_{3,37} = 11.81$ $p > .10$.

Because the above analyses showed no differences among the groups, errors were reclassified as algebraic or non-algebraic (see Appendix E for examples). The mean number of algebraic errors for each condition on the (first) posttest were: Reteach: 2.00, MBR: 2.09, MBR+CD: 3, MBR+CE: 2.73, indicating no major differences among the conditions.

Following these results, we hypothesized that a significant number of student errors might be unstable; which we investigated by tracing a given error for a given task from pretest, to tutoring, to posttest. (In retrospect, we believe that this is a stringent stability criterion.) This analysis showed that approximately 80% of the errors from the pretest across conditions did not occur on the same items during tutoring. However, intermediate tutoring on items may have lowered this stability measure since the error on subsequent items may have been corrected as a result of tutoring on previous items. If stable errors are reclassified to include those tutored in an earlier part of the session, the average percentage of stable errors during tutoring increases to 26%.

An analogous retrospective analysis of the data from the study II at School L was carried out, which confirmed these results.

Discussion

Three analyses were applied to this study, each of which was logically driven by the earlier ones. Analyses 1 and 2 did not find differential effects among conditions, in spite of the modifications to the basic MBR treatment. (It should be noted, however, that the MBR + CD condition could not be properly tested with this sample, because the process of substitution and verification was new to these students.) Analysis 3 suggested that the phenomena of student errors is more complex than we had anticipated and pointed strongly to the instability of a high proportion of student errors. No further conclusions were drawn from this experiment as it was not designed to investigate stability; stability became the focus of study VI.

IV. RPIXIE-Based Remediation

The reader will recall that the main conclusion from study I (the previous computer-based study) was that MBR and Reteaching were comparable, and this in turn was attributed to the poor diagnostic capability of RPIXIE. Subsequently, additional mal-rules were added to the algebra knowledge base, and RPIXIE has been enhanced so that it can offer diagnoses in the cases in which multiple modes explain the student's error. Given that the above studies with human-tutors have shown that MBR and Reteaching are highly comparable, the foci of this study were:

a) Has the diagnostic capability of RPIXIE improved with the addition of mal-rules, and b) Would differences arise between conditions when the computer (namely RPIXIE) provided the remediation, which perhaps might not emerge with a human tutor?

Procedure: Forty-one students from 2nd-and-3rd-year classes (in School P) were randomly assigned to condition: MBR, Reteaching, or Control. Students in the control condition did not interact with RPIXIE. All students took a 20-item pretest during their regular maths class. The week following the pretest, students in the treatment groups individually interacted with RPIXIE. Very little researcher/student interaction occurred while the student worked with RPIXIE. If the student did not understand the instructions, they were explained by the researcher. All students began on RPIXIE at the easiest level (Task Set 1) and worked until all task-sets were completed or 45 minutes had elapsed, whichever was the sooner. Students were then given 5 minutes to answer a 4-item questionnaire pertaining to their experiences with the program. After all the students had interacted with RPIXIE, a group posttest was administered during the regular mathematics period.

Results:

The data were first analysed to determine the percentage of errors that RPIXIE diagnosed. As before, diagnosis was defined in two ways: a) Did PIXIE have a model that the researchers believed explained the student's answer, and b) Did the student agree that RPIXIE's model was acceptable?*

Under the first criterion, RPIXIE diagnosed approximately 60% of the errors made by students in the MBR condition; the student's "approved" 29% of these, resulting in 17% of errors being followed by MBR. Studies II and III reported here suggest that approximately 13% of the errors students make while solving algebra items are computational. When one takes this into account, RPIXIE's matching of non-arithmetic errors increases on the two measures given earlier to 69.0% (60 out of 87) and 19.5% (17 out of 87) respectively**.

---

* RPIXIE allows the student to disagree with the incorrect procedure produced by the inferred model. Allowing the student to disagree with the model prevents students from seeing remediation that does not reflect the particular incorrect procedure used. A student may disagree with RPIXIE's model because the trace contains steps the student did not do. For example, the model may show the following:

$3x = 9$
$3x/3 = 3/9$
$x = 3/9$
$x = 1/3$

The student may reject this trace because the step $3x/3 = 3/9$ was not typed by him.

** The analysis of study VI showed that unstable, undiagnosed and computational errors account for approximately 15% of errors. (See the discussion of study VI for the definition of stable errors and for further details.) However, that may not be the appropriate figure for computer-based studies; there are indications that the error rate may be higher, (Sleeman, 1982).

Table 4 - Pretest and Posttest Scores

Condition	Mean Pretest Score	Mean Posttest Score	Mean Number of items worked on RPIXIE	% correct
Model-based	14.43 (2.59)	16.13 (2.83)	39	90%
Reteach	15.00 (2.08)	15.71 (2.52)	40	87%
Control	14.38 (2.66)	14.75 (3.79)	N/A	N/A

The analyses in Table 4 showed no differences on the posttest by condition ($F_{2,36}$= .71, $p > .50$); standard deviations are given in brackets. A differential effect for tutoring may not have been found because students working with RPIXIE did not in fact receive much tutoring, essentially only 4 tasks (10% of 40) on average were worked incorrectly on both conditions.

Discussion

RPIXIE's diagnostic capability has been improved since study I. Additionally, a capability to handle multiple models was added; however, as Table 4 shows, the students incorrectly worked only a small number of tasks. Thus, there was relatively little opportunity for RPIXIE to provide either MBR or Reteaching to this sample, which probably helps explain the lack of difference among the three conditions on the posttest. Consequently, the hypothesis that MBR with RPIXIE might be better than MBR with humans, is left untested.

The notable change between study I and this study, is the reduction from 90% to 29% in the students' "approval" of the MBR traces. This can be explained by a difference in population, but more particularly by changes in the knowledge-base which resulted in highly redundant and longer traces. Moore and Sleeman (1987) discuss enhancements to address this latter concern.

V. A Third Study Using RPIXIE

An RPIXIE study, analogous to study IV, was run in School L and addressed the same issues as study IV: a) Had the diagnostic capability of RPIXIE improved since study I, and b) Would differences arise among conditions when the computer provided the remediation?

Procedure: Those students from study VI who were classified as having "unstable" errors or who only had one pretest score; 21 students in all were randomly assigned to three conditions. The procedure for this study was identical to that used in study IV except that students in this control group interacted with a very restricted RPIXIE, which provided no feedback.

Results and Discussion

Essentially, this study confirmed the results obtained with Study IV, with the exception that RPIXIE diagnosed a smaller percentage of errors. This latter result is not surprising given the nature of this sample (most of these students had decided not to study mathematics further.)

VI. Human-based Remediation: MBR and Reteaching in the context of stable errors

Given the apparent instability of errors suggested by the analyses of studies II and III, this study was carried out to investigate error stability in depth, and to compare the effects of MBR and Reteaching in the context of stable errors.

Subjects

Students from the 2nd- and 3rd- year mathematics classes at School L

V. A Third Study Using RPIXIE

An RPIXIE study, analogous to study IV, was run in School L and addressed the same issues as study IV: a) Had the diagnostic capability of RPIXIE improved since study I, and b) Would differences arise among conditions when the computer provided the remediation?

Procedure: Those students from study VI who were classified as having "unstable" errors or who only had one pretest score; 21 students in all were randomly assigned to three conditions. The procedure for this study was identical to that used in study IV except that students in this control group interacted with a very restricted RPIXIE, which provided no feedback.

Results and Discussion

Essentially, this study confirmed the results obtained with Study IV, with the exception that RPIXIE diagnosed a smaller percentage of errors. This latter result is not surprising given the nature of this sample (most of these students had decided not to study mathematics further.)

VI. Human-based Remediation: MBR and Reteaching in the context of stable errors

Given the apparent instability of errors suggested by the analyses of studies II and III, this study was carried out to investigate error stability in depth, and to compare the effects of MBR and Reteaching in the context of stable errors.

Subjects

Students from the 2nd- and 3rd- year mathematics classes at School L served as the subject pool.

Materials

A 51-item stability measure consisting of 17 sets of three algebra tasks was developed; each task within a set being generated by the same template [the template for the first set is aX=b; for set 17 it is aX=b*c(dX+e)]. Using these templates, a pretest and posttest were constructed, with the requirement that the first item of each set on both tests be identical. (See Appendix F.)

Procedure

Ninety-six students in the 2nd- and 3rd- year classes at School L were pretested; twenty-one of these were considered not to need tutoring because they had at least 88% of the items (45 of the 51) correct. A further 23 were not considered because they would be absent for a school function during the week of tutoring. A further 15 students were present for only one of the two pretests. Thirty-seven students remained. Of these, 28 had at least one stable error. In this study a stable error is defined as one that occurs at least twice on both pretests. These 28 students were then randomly assigned to condition (MBR, Reteaching or Control).

Students in the treatment conditions were seen individually for a 50 minute period. To put the student at ease, each student was asked to work the first six items from the pretest. After the first six items were worked, stable errors were tutored. An error, of course, was only tutored if the error occurred again when the student worked the item during the tutoring session. After all "stable" errors were tutored, any errors made on the other 51 items were tutored. Students in the served as the subject pool.

Materials

A 51-item stability measure consisting of 17 sets of three algebra tasks was developed; each task within a set being generated by the same template [the template for the first set is aX=b; for set 17 it is aX=b*c(dX+e)]. Using these templates, a pretest and posttest were constructed, with the requirement that the first item of each set on both tests be identical. (See Appendix F.)

Procedure

Ninety-six students in the 2nd- and 3rd- year classes at School L were pretested; twenty-one of these were considered not to need tutoring because they had at least 88% of the items (45 of the 51) correct. A further 23 were not considered because they would be absent for a school function during the week of tutoring. A further 15 students were present for only one of the two pretests. Thirty-seven students remained. Of these, 28 had at least one stable error. In this study a stable error is defined as one that occurs at least twice on both pretests. These 28 students were then randomly assigned to condition (MBR, Reteaching or Control).

Students in the treatment conditions were seen individually for a 50 minute period. To put the student at ease, each student was asked to work the first six items from the pretest. After the first six items were worked, stable errors were tutored. An error, of course, was only tutored if the error occurred again when the student worked the item during the tutoring session. After all "stable" errors were tutored, any errors made on the other 51 items were tutored. Students in the

Control group simply took the pretests and posttest. The week after tutoring, a posttest was given to the classes involved. A delayed posttest was also given a month after the initial posttest.

Results

Of the 28 students, 3 were absent from school - one from each condition - when the first posttest was given, resulting in the following cell sizes: MBR 9, Reteaching 8, Control 8. In addition, one student in MBR was subsequently found during a reliability check to have no stable errors. His data are included in the analyses because they are otherwise sound. The same three students and one other from the MBR condition, were absent from school on the day of the delayed posttest, leaving eight students per condition.

Table 5 - Mean Pretest and Posttest Scores* by Condition

Condition	Pretest 1	Pretest 2	Posttest 1	Delayed Posttest
MBR	27.50 (12.79)	29.70 (14.04)	41.22 (9.09)	45.38 (5.55)
Reteach	28.00 (14.49)	33.56 (11.06)	41.62 (4.47)	37.12 (8.69)
Control	23.44 (12.35)	25.44 (13.44)	26.00 (14.36)	28.38 (15.15)

*Maximum = 51

Table 5 gives the means and standard deviation by condition for the four testings. An ANOVA showed no significant differences among the conditions (N=28) on either pretest1: $\underline{F}$ (2,25)=.33, $\underline{p}$ >.72, or pretest2: $\underline{F}$ (2,25)=.88, $\underline{p}$ >.42, established that randomization had been successful.

Significant omnibus F ratios were found for both the posttest (N=25) and delayed posttests (N=24): $\underline{F}$ (2,22)=6.33, $\underline{p}$ <.008, and $\underline{F}$ (2,21)=5.78, $\underline{p}$ <.02, respectively. Post hoc analyses using the Scheffe test showed no differences ($\underline{p}$ >.05) between MBR and Reteaching, but both being better ($\underline{p}$ <.05) than the control condition on both posttests. Although, ANOVA is generally regarded to be robust with respect to violations of its assumptions, the reader should note the small cell sizes and the fact that heterogeneity of variance ($\underline{p}$ <.05) was demonstrated for both posttests using Bartlett's test.

A paired t-test comparing MBR and Reteaching scores on pretest1 with matched scores on the posttest showed a significant gain from a mean of 27.35 to one of 41.41, t(16)=-3.70, p<.001. However, the reader should note that the correlated variances (192.38 and 50.13, respectively) were found to be significantly different: t(15)=3.28, p<.01. A similar analysis for the control condition showed no significant differences between either the means on pretest1 and the posttest: $t(7)=-1.77$, $p<.12$, or the correlated variances (124.77 and 206.21, respectively): $t(6)=-0.5919$, $p>.1$.

Analyses of Stability

The total set of errors encountered in this experiment, were classified into 46 different types. Only 19 of these types appeared during tutoring; the other 27 types occurred infrequently on the pretests (i.e., were not stable) or only on the posttest.

Table 6: Number of times the 19 error types occurred

Condition	Pretest 1	Pretest 2	Posttest
MBR	173	167	93
Reteaching	235	181	79
Control	256	211	208

An inspection of Table 6 shows that for all three conditions there was a decrease in the number of errors from pretest1 to pretest2. The percentage decrease by condition from pretest1 to pretest2 and from pretest1 to posttest are shown in table 7.

Table 7: Percentage Decrease in the 19 Stable Errors by Condition.

	Pre1 to Pre2	Pre1 to Post
MBR	3.5	46.2
Reteaching	23.0	66.4
Control	17.6	18.75

[Note: There was an overall average decrease of 15.8% from pretest1 to pretest2.]

Table 7 shows that the percentage decrease for the control condition remains basically unchanged, whereas for the treatment conditions the percentage decreases are much more dramatic, mirroring Table 6. (This result is completely consistent with the result reported by Sleeman (1983) for a study involving 2 groups: essentially MBR and Control.)

Table 8: Analysis of the 19 stable errors.

Error Type		Number Students who had a particular (stable) error	Prevalence of stable error (as % of student population	Freq. of errors on Pretest 1 (as %)	Freq. of errors on Pretest 2 (as %)
1	Bracket	14	56	10	12
2	Precedence	13	52	10	8
3	Computational	7	28	9	7
4	Change side not sign of x-term	5	20	7	7
5	Subtract a coeff	4	16	16	18
6	Add x and constant	3	12	4	7
7	Add a negative sign	3	12	2	2
8	Uses a number twice	3	12	3	5
9	Drop an X	2	8	3	3
10	Inverted division	2	8	13	8
11	Minus before the wrong number	2	8	3	4
12	Subtract a multiplier	2	8	1	1
13	Subtract a multiplied x-term	2	8	2	3
14	ax = b => x = a - b	2	8	1	1
15	Drop a negative sign	1	4	3	2
16	Multiply across by a coeff	1	4	2	1
17	Ends task with ax = bx	1	4	1	1
18	ax = b => x = - (a + b	1	4	0	0
19	a * bx + cx = d => a * (-d) = -bx -cx	1	4	0	0

Table 8 shows that the percentage of students who made stable errors varies by type of error. Some stable errors (e.g., precedence errors) are more prevalent than others (e.g., "inverted division"). The relationship between the relative prevalence of a stable error and the relative frequency with which it occurs in the population of errors is not one-to-one. For example, precedence errors occur in 52% of the sample of students, yet account for only 10% of the total pretest1 errors; inverted division errors occur in only 8% of the student sample, but account for 13% of the pretest1 errors. (Note further that these frequency figures are NOT the same for both pretests.)

Noise

In addition to these stable errors, there were errors on the pretests that were either unstable (by the definition given above), or undiagnosed. These latter errors together with those labelled "computational" in Table 8 will be described collectively as "noise". "Noise" accounted for 16% of the pretest1 errors, and for 14% of the pretest2 errors. By way of comparison, Sleeman (1982) reported 29.6% noise for a student population when it interacted with a predecessor of RPIXIE. (Suggesting that students' responses in this domain are noisier when interacting with programs than when interacting with human tutors.

If one extended the definition of noise to include hose stable errors made by only a small percentage of the sample. The percentages of total errors described as noise would rise to 22% on pretest1 (18% on pretest2) if one included stable errors made by 4% or less of the student sample. The percentages would rise to 45% on pretest1 (38% on pretest2) if one included stable errors made by 8% or less of the student sample.

How many stable errors did the students have?

Table 9 shows the number of students who had at least one stable error, at least two stable errors, and so on up to eight stable errors. Clearly, most students have at least two stable errors, but the percentage drops off sharply.

Table 9: Number of students with N Stable errors

Number of stable errors	1	2	3	4	5	6	7	8
Number of students	24	22	9	7	4	2	1	0

Remediating Stable Errors by Conditions

If one lists by condition the number of stable errors which occurred for all 3 conditions during the pretests, and compares this with the number of errors that appeared on the posttest, one can express the remedial effectiveness of a condition as the percentage reduction of errors, see Table 10. (Note, not all stable errors occurred in each of the conditions, and so these groups are only partially "matched").

Table 10: Remediating Stable Errors which occur on all 3 conditions

Stable Errors

Condition	Number on Pretests	Still on posttest	% reduction
MBR	10	5	50
Reteaching	12	7	42
Control	12	12	0

Note: Of the six categories of stable errors in this table, only bracket (distributive law) errors and precedence errors appeared on the posttest for both conditions. We label these errors, "resistant".

An inspection of Table 10 reflects the overall findings of this study: both treatment conditions whilst outperforming the control condition are themselves very comparable.

Discussion

Effects of different forms of Remediation

Once again, study VI has shown that MBR and Reteaching are both better than no treatment, but we have been unable to distinguish between the two. So with this subject domain when taught procedurally with this type of student population, we concluded that even when one is remediating stable errors, MBR and Reteaching appear to be equally effective. This is still a surprising result to us. Below, we speculate further about this result:

1. Maybe the MBR and Reteaching treatments used in this study were still too similar. Perhaps, the scripts followed were not sufficiently different (in particular, we noted above that the MBR + Cognitive Dissonance treatment should be rerun, as the students in study III did not have an appropriate background). Another possible reason for lack of effect, is the duration of the exposure. Fifty minutes of tutorial seemed a substantial period; however, Swan (1983) reports effects after eight one-hour long lessons.

2. Even though the MBR and Reteaching groups were highly comparable on procedural tasks, had they been tested for conceptual understanding of algebra we might have found that the MBR students would have outperformed the Reteaching students.

3. As noted in the introduction, MBR essentially assumes that the student has a (stable) mental model, to which remedial comments are related. So an additional hypothesis we created, is that students taught algebra procedurally have a (weak) mental model of the

domain. Further, we have hypothesized that students taught algebra <u>conceptually</u> should have a much stronger mental model, and so one should observe with such students significant differences between the MBR and Reteaching treatments. Unfortunately, we were unable to find a high-school in Aberdeen where algebra was taught (largely) conceptually, so this hypothesis remains untested.*

4. A further speculation is that the MBR and Reteaching treatments are in fact <u>very</u> similar, because each student in the Reteaching treatment, essentially notes for himself the difference between his answer and that provided by the correct procedure, and so generates his own MBR. Although this is of little importance for instruction, we plan to run an experiment to probe this issue. (If this is the mechanism, one would expect to find differences between the mathematically able and less-able students.)

* Kelly and Sleeman (1986) found that most teachers teach algebra with a rule-based emphasis. Because the nature of the student's initial instruction became a concern for the PIXIE studies, we looked for a school that taught algebra conceptually. Unfortunately, no such school was available in Aberdeen. However, we did interview students to determine their understanding of algebra. The interview questionaire was a revised version of a questionaire developed by Sleeman, Steinburg & Ktorza (1985). We found that even students' conceptual understanding of algebra is inconsistent within a one-hour interview period. Indeed this experiment has led us to the not-too-surprising position that conceptual understanding is multifaceted (as is procedural competence) and <u>not</u> monolithic as is often assumed.

Stability of Errors

The assumption that student errors are completely stable has been clearly questioned by several studies, including VanLehn (1981); Bricken (1987) as well as Sleeman (1983)** The following issues on error stability are raised by the results of study VI:

1. Attentional nature of some errors. Forty-six types of errors were encountered during the study, but only 19 of these were "stable" during the tutorial sessions. This suggests that tutorial contact is effective in getting rid of some types of careless errors, such as dropping/adding signs. (The explanation for this appears to be the motivational effect of having a tutor work with a student.)

2. Prevalence and Frequency. We have introduced, and wish to stress, the two measures we have used to describe this phenomenon. Prevalence indicates the number of students in the population who have a particular error. Frequency indicates the proportion of the total number of errors that are explained by the specific error. Both measures are needed to discuss the phenomena of errors.

3. Use of this data for remediation: A class teacher will be essen-

**VanLehn (1981) investigated short-term and long-term stability of subtraction errors, and found that only 12% of the students who had errors on the first test had the same errors on the second (short term) test. He also found that the long term stability data are very similar to the short term stability data. VanLehn concluded that errors in general are not stable. Further, Bricken (1987) investigated the stability of algebra errors, and found that 50% (11 of 22 students) committed at least one error on the pretest which reoccurred during their interview held two weeks later. Although the studies mentioned above all confirm that errors generally are not stable, it is important to note that each investigation measures stability using different criteria. However, lack of stability seems to be an issue regardless of how it is measured.

tially concerned to know the most commonly occuring (i.e., the most <u>prevalent</u>) errors in the class. Whereas, a tutor will wish to have access to the actual error-profile for each student.

4. Taxonomy of errors: This study has confirmed and slightly extended the error taxonomy suggested by Sleeman (1983). In our view, it is feasible to talk about the following types of errors:

- Stable errors (both remediable and "resistant")

- Attentional errors (largely minor errors such as adding/dropping signs).

- Classes of mal-rules, used by the same student on different occasions with the same type of task.

- Mental slips, typing/transcription errors.

It is important for a tutor to categorize correctly the error as, for example, it may be counterproductive to tutor a student on an error which is a result of a slip; whereas it maybe important to address a stable error. (These judgements are both subtle and complex.)

Overall conclusions of the series of studies

* For introductory algebra when taught procedurally with this type of student, it appears that Reteaching is as effective as MBR. From which it follows that CAI would be as effective as an ITS It is vital that we investigate the range of subjects, instructional approaches and student age-ranges for which this result holds. (This information will be important for both educators and ITSs workers.)

* Despite the conclusions on the last paragraph, one should not conclude that reteaching by a classroom teacher will be as effective as reteaching in a one-to-one tutoring situation or by a computer system. Immediacy of feedback which is lacking from the normal classroom situation may well be a critical factor, Lewis & Anderson (1985).

* It is critical that ITSs receive extensive field-testing.

* The subfield of ITS should not conclude that the task of building an ITS is impossible, but it should conclude that the task is harder than we had initally thought*, c.f., Winograd & Flores (1986). It is possible that the more global analysis (Moore & Sleeman, 1987) which takes account of the student's performance at several levels might be needed before proper remediation can be undertaken. But before such a system is built it is suggested that experiments are run to see whether human tutors are more effective when they base their remediation on such a global approach.

* Cronbach and Snow (1977) warned that producing truly individualized instruction was a demanding task, but the ITS field chose to infer that their conclusions were unduly pessimistic given the complexity of analysing such a diverse set of experimental studies.

For a more extensive discussion of the conclusions of this experiment including a discussion of possible additional issues to be investigated, see Sleeman, Kelly, Martinak, Ward and Moore (1987).

References

Bricken, W. M. (1987). _Analyzing errors in elementary mathematics_. Doctoral Dissertation, School of Education, Stanford University.

Buckingham, G. E. (1933). _Diagnostic and remedial teaching in first-year algebra_. Northwestern University, IL, School of Education series, No. 11.

Bunderson, V. C. & Olsen, J. B. (1983). _Mental errors in arithmetic skills: their diagnosis in pre-college students_. WICAT Education Institute Provo, UT. Final project report, NSF SED 80-12500.

Cronbach, L. J. & Snow, R. E. (1977). _Aptitudes and Instructional Methods_. New York: Irvington.

Hartley, J. R. & Sleeman, D. H. (1973). Towards intelligent teaching systems. _International Journal of Man-Machine Studies_, _5_, 215-236.

Kelly, A. E., Martinak, R. & Sleeman, D. (1987). _How important is diagnosis for remediation? A concern for intelligent tutoring systems_. Proceedings of the Second European Conference for Research on Learning and Instruction, Tubingen, Federal Republic of Germany, September 19-27.

Kelly, A. E. & Sleeman, D. (1986). _A study of Diagnostic and Remedial Techniques used by Master Algebra Teachers_. Technical Report AUCS/TR8708, Department of Computing Science, University of Aberdeen.

Lewis, M. W. & Anderson, J. R. (1985). Discrimination of operator schemata in problem solving: Learning from examples. Cognitive Psychology, 17, pp 26-65.

Macnab, D. S. & Cummine, J. A. (1986). Teaching Mathematics 11-16: A Difficulty-Centred Approach. London: Basil Blackwell, England.

Martinak, R., Schneider, B. and Sleeman, D. (1987). A Comparative Analysis of Approaches for Correcting Algebra Errors via an Intelligent Tutoring System. Proceedings of AERA, Washington D.C.

Moore, J. L. & Sleeman, D. (1987). Enhancing PIXIE's tutoring capabilities. Technical Report AUCS/TR8709, Department of Computing Science, University of Aberdeen.

Putnam, R. T. (1987). Structuring and adjusting Content for students: A study of live and simulated tutoring of addition. American Educational Research Journal, Vol 24, No 1, pp. 13 - 48

Resnick, L. (1984). Beyond error analysis: the role of understanding in elementary school arithmetic. Learning Research and Development Centre, University of Pittsburgh, PA.

Resnick L. B. & Ford W. W., (1981). The psychology of mathematics for instruction. Hilsdale, NJ: Lawrence Erlbaum Associates.

Sleeman, D. (1982). Assessing aspects of competence in basic algebra. In Intelligent Tutoring Systems D. Sleeman & J. S Brown (Eds). London: Academic Press, pp 185-199.

Sleeman, D. (1983). Basic algebra revisited: a study with 14-year olds. HPP Report 83-9, Computer Science Dept, Stanford University. Republished in International Journal Man-Machine Studies, (1985), 22, pp 127-149.

Sleeman, D. (1987). PIXIE: A shell for developing Intelligent Tutoring Systems. In R.W. Lawler & M. Yazdani (Eds). Artificial Intelligence and Education: Volume 1, (pp 239-265). Norwood, N.J: Ablex.

Sleeman, D. & Brown, J. S. (1982). Intelligent Tutoring Systems, London: Academic Press.

Sleeman, D., Kelly, A. E., Martinak, R., Ward, R. D. & Moore, J. (1987) Diagnosis and Remediation in the context of Intelligent Tutoring Systems. Technical Report AUCS/TR8712, Department of Computing Science, University of Aberdeen.

Steinburg, R. M., Sleeman, D. and Ktorza, D. (1985) Transformations and Equivalence of Equations in High School Algebra. Technical Report, School of Education, Stanford.

Swan, M. B. (1983). Teaching decimal place value. A comparative study of conflict and positive only approaches. Research Report 31, Shell Centre for Mathematical Education, University of Nottingham.

VanLehn, K. (1981). Bugs are not enough: Empirical studies of bugs, impasses and repairs in procedural skills. XEROX, Palo Alto Research Centre, CA. Cognitive and Instructional Science series, CIS-11 (SSL-81-2).

Winograd, T. & Flores, F. F. (1986). Understanding computers and cognition: A new foundation for design. Norwood, N. J.: Ablex.

Appendix A

Pretest Items

1. $7x = 14$
2. $4x = -12$
3. $5x = 7$
4. $8x = 18$
5. $3x = 8 + 4$
6. $3x = 4 \ast 4$
7. $3x + 4x = 14$
8. $3x + 5 = 26$
9. $7 + 4x = 19$
10. $5x = 4x + 8$
11. $3x = 3(2 + 3)$
12. $12x = 2(3x + 3)$
13. $7x + 2 + 4 \ast 8$
14. $17x = 19x + 25$
15. $3 + 2x + 4x = 21$
16. $4 + 3x + 4x = 25$
17. $35x = 2 + 3(4x + 5)$

Posttest Items

1. $3x = 9$
2. $4x = -8$
3. $3x = 5$
4. $6x = 9$
5. $5x = 8 + 2$
6. $3x = 4 \ast 3$
7. $2x + 3x = 10$
8. $2x + 3 + 9$
9. $2 + 4x = 14$
10. $4x + 3x + 6$
11. $2x = 3(3 + 1)$
12. $24x = 3(2x + 3)$
13. $5x = 2 + 3 \ast 8$
14. $16x = 19x + 20$
15. $2 + 2x + 3x = 17$
16. $3 + 3x + 4x = 24$
17. $37x = 2 + 3(4x + 5)$

18. 3*2x + 3x = 19	18. 5*2x + 4x = 18
19. 3(2x + 4) = 12(9 + 2x)	19. 4(3x + 3) = 5(6 + 2x)
20. 21x = 3 * 2(2x + 3)	20. 24x = 3 * 2(2x + 4)

Control group simply took the pretests and posttest. The week after tutoring, a posttest was given to the classes involved. A delayed posttest was also given a month after the initial posttest.

Results

Of the 28 students, 3 were absent from school - one from each condition - when the first posttest was given, resulting in the following cell sizes: MBR 9, Reteaching 8, Control 8. In addition, one student in MBR was subsequently found during a reliability check to have no stable errors. His data are included in the analyses because they are otherwise sound. The same three students and one other from the MBR condition, were absent from school on the day of the delayed posttest, leaving eight students per condition.

Table 5 - Mean Pretest and Posttest Scores* by Condition

Condition	Pretest 1	Pretest 2	Posttest 1	Delayed Posttest
MBR	27.50 (12.79)	29.70 (14.04)	41.22 (9.09)	45.38 (5.55)
Reteach	28.00 (14.49)	33.56 (11.06)	41.62 (4.47)	37.12 (8.69)
Control	23.44 (12.35)	25.44 (13.44)	26.00 (14.36)	28.38 (15.15)

*Maximum = 51

Table 5 gives the means and standard deviation by condition for the four testings. An ANOVA showed no significant differences among the conditions (N=28) on either pretest1: F (2,25)=.33, p >.72, or pretest2: F (2,25)=.88, p >.42, established that randomization had been successful.

Significant omnibus F ratios were found for both the posttest (N=25) and delayed posttests (N=24): F (2,22)=6.33, p <.008, and F (2,21)=5.78, p <.02, respectively. Post hoc analyses using the Scheffe test showed no differences (p >.05) between MBR and Reteaching, but both being better (p <.05) than the control condition on both posttests. Although, ANOVA is generally regarded to be robust with respect to violations of its assumptions, the reader should note the small cell sizes and the fact that heterogeneity of variance (p <.05) was demonstrated for both posttests using Bartlett's test.

A paired t-test comparing MBR and Reteaching scores on pretest1 with matched scores on the posttest showed a significant gain from a mean of 27.35 to one of 41.41, t(16)=-3.70, p<.001. However, the reader should note that the correlated variances (192.38 and 50.13, respectively) were found to be significantly different: t(15)=3.28, p<.01. A similar analysis for the control condition showed no significant differences between either the means on pretest1 and the posttest: $t(7)=-1.77$, $p<.12$, or the correlated variances (124.77 and 206.21, respectively): $t(6)=-0.5919$, $p>.1$.

Analyses of Stability

The total set of errors encountered in this experiment, were classified into 46 different types. Only 19 of these types appeared during tutoring; the other 27 types occurred infrequently on the pretests (i.e., were not stable) or only on the posttest.

Table 6: Number of times the 19 error types occurred

Condition	Pretest 1	Pretest 2	Posttest
MBR	173	167	93
Reteaching	235	181	79
Control	256	211	208

An inspection of Table 6 shows that for all three conditions there was a decrease in the number of errors from pretest1 to pretest2. The percentage decrease by condition from pretest1 to pretest2 and from pretest1 to posttest are shown in table 7.

Table 7: Percentage Decrease in the 19 Stable Errors by Condition.

	Pre1 to Pre2	Pre1 to Post
MBR	3.5	46.2
Reteaching	23.0	66.4
Control	17.6	18.75

[Note: There was an overall average decrease of 15.8% from pretest1 to pretest2.]

Table 7 shows that the percentage decrease for the control condition remains basically unchanged, whereas for the treatment conditions the percentage decreases are much more dramatic, mirroring Table 6. (This result is completely consistent with the result reported by Sleeman (1983) for a study involving 2 groups: essentially MBR and Control.)

Table 8: Analysis of the 19 stable errors.

Error Type	Number Students who had a particular (stable) error	Prevalence of stable error (as % of student population	Freq. of errors on Pretest 1 (as %)	Freq. of errors on Pretest 2 (as %)
1 Bracket	14	56	10	12
2 Precedence	13	52	10	8
3 Computational	7	28	9	7
4 Change side not sign of x-term	5	20	7	7
5 Subtract a coeff	4	16	16	18
6 Add x and constant	3	12	4	7
7 Add a negative sign	3	12	2	2
8 Uses a number twice	3	12	3	5
9 Drop an X	2	8	3	3
10 Inverted division	2	8	13	8
11 Minus before the wrong number	2	8	3	4
12 Subtract a multiplier	2	8	1	1
13 Subtract a multiplied x-term	2	8	2	3
14 ax = b => x = a - b	2	8	1	1
15 Drop a negative sign	1	4	3	2
16 Multiply across by a coeff	1	4	2	1
17 Ends task with ax = bx	1	4	1	1
18 ax = b => x = - (a + b	1	4	0	0
19 a * bx + cx = d => a * (-d) = -bx -cx	1	4	0	0

Table 8 shows that the percentage of students who made stable errors varies by type of error. Some stable errors (e.g., precedence errors) are more prevalent than others (e.g., "inverted division"). The relationship between the relative prevalence of a stable error and the relative frequency with which it occurs in the population of errors is not one-to-one. For example, precedence errors occur in 52% of the sample of students, yet account for only 10% of the total pretestl errors; inverted division errors occur in only 8% of the student sample, but account for 13% of the pretestl errors. (Note further that these frequency figures are NOT the same for both pretests.)

Noise

In addition to these stable errors, there were errors on the pretests that were either unstable (by the definition given above), or undiagnosed. These latter errors together with those labelled "computational" in Table 8 will be described collectively as "noise". "Noise" accounted for 16% of the pretest1 errors, and for 14% of the pretest2 errors. By way of comparison, Sleeman (1982) reported 29.6% noise for a student population when it interacted with a predecessor of RPIXIE. (Suggesting that students' responses in this domain are noisier when interacting with programs than when interacting with human tutors.

If one extended the definition of noise to include those stable errors made by only a small percentage of the sample. The percentages of total errors described as noise would rise to 22% on pretest1 (18% on pretest2) if one included stable errors made by 4% or less of the student sample. The percentages would rise to 45% on pretest1 (38% on pretest2) if one included stable errors made by 8% or less of the student sample.

How many stable errors did the students have?

Table 9 shows the number of students who had at least one stable error, at least two stable errors, and so on up to eight stable errors. Clearly, most students have at least two stable errors, but the percentage drops off sharply.

Table 9: Number of students with N Stable errors

Number of stable errors	1	2	3	4	5	6	7	8
Number of students	24	22	9	7	4	2	1	0

Remediating Stable Errors by Conditions

If one lists by condition the number of stable errors which occurred for all 3 conditions during the pretests, and compares this with the number of errors that appeared on the posttest, one can express the remedial effectiveness of a condition as the percentage reduction of errors, see Table 10. (Note, not all stable errors occurred in each of the conditions, and so these groups are only partially "matched").

Table 10: Remediating Stable Errors which occur on all 3 conditions

Stable Errors

Condition	Number on Pretests	Still on posttest	% reduction
MBR	10	5	50
Reteaching	12	7	42
Control	12	12	0

Note: Of the six categories of stable errors in this table, only bracket (distributive law) errors and precedence errors appeared on the posttest for both conditions. We label these errors, "resistant".

An inspection of Table 10 reflects the overall findings of this study: both treatment conditions whilst outperforming the control condition are themselves very comparable.

Discussion

Effects of different forms of Remediation

Once again, study VI has shown that MBR and Reteaching are both better than no treatment, but we have been unable to distinguish between the two. So with this subject domain when taught procedurally with this type of student population, we concluded that even when one is remediating stable errors, MBR and Reteaching appear to be equally effective. This is still a surprising result to us. Below, we speculate further about this result:

1. Maybe the MBR and Reteaching treatments used in this study were still too similar. Perhaps, the scripts followed were not sufficiently different (in particular, we noted above that the MBR + Cognitive Dissonance treatment should be rerun, as the students in study III did not have an appropriate background). Another possible reason for lack of effect, is the duration of the exposure. Fifty minutes of tutorial seemed a substantial period; however, Swan (1983) reports effects after eight one-hour long lessons.

2. Even though the MBR and Reteaching groups were highly comparable on procedural tasks, had they been tested for conceptual understanding of algebra we might have found that the MBR students would have outperformed the Reteaching students.

3. As noted in the introduction, MBR essentially assumes that the student has a (stable) mental model, to which remedial comments are related. So an additional hypothesis we created, is that students taught algebra procedurally have a (weak) mental model of the

domain. Further, we have hypothesized that students taught algebra conceptually should have a much stronger mental model, and so one should observe with such students significant differences between the MBR and Reteaching treatments. Unfortunately, we were unable to find a high-school in Aberdeen where algebra was taught (largely) conceptually, so this hypothesis remains untested.*

4. A further speculation is that the MBR and Reteaching treatments are in fact very similar, because each student in the Reteaching treatment, essentially notes for himself the difference between his answer and that provided by the correct procedure, and so generates his own MBR. Although this is of little importance for instruction, we plan to run an experiment to probe this issue. (If this is the mechanism, one would expect to find differences between the mathematically able and less-able students.)

* Kelly and Sleeman (1986) found that most teachers teach algebra with a rule-based emphasis. Because the nature of the student's initial instruction became a concern for the PIXIE studies, we looked for a school that taught algebra conceptually. Unfortunately, no such school was available in Aberdeen. However, we did interview students to determine their understanding of algebra. The interview questionaire was a revised version of a questionaire developed by Sleeman, Steinburg & Ktorza (1985). We found that even students' conceptual understanding of algebra is inconsistent within a one-hour interview period. Indeed this experiment has led us to the not-too-surprising position that conceptual understanding is multifaceted (as is procedural competence) and not monolithic as is often assumed.

Stability of Errors

The assumption that student errors are completely stable has been clearly questioned by several studies, including VanLehn (1981); Bricken (1987) as well as Sleeman (1983)** The following issues on error stability are raised by the results of study VI:

1. Attentional nature of some errors. Forty-six types of errors were encountered during the study, but only 19 of these were "stable" during the tutorial sessions. This suggests that tutorial contact is effective in getting rid of some types of careless errors, such as dropping/adding signs. (The explanation for this appears to be the motivational effect of having a tutor work with a student.)

2. Prevalence and Frequency. We have introduced, and wish to stress, the two measures we have used to describe this phenomenon. Prevalence indicates the number of students in the population who have a particular error. Frequency indicates the proportion of the total number of errors that are explained by the specific error. Both measures are needed to discuss the phenomena of errors.

3. Use of this data for remediation: A class teacher will be essen-

---

**VanLehn (1981) investigated short-term and long-term stability of subtraction errors, and found that only 12% of the students who had errors on the first test had the same errors on the second (short term) test. He also found that the long term stability data are very similar to the short term stability data. VanLehn concluded that errors in general are not stable. Further, Bricken (1987) investigated the stability of algebra errors, and found that 50% (11 of 22 students) committed at least one error on the pretest which reoccurred during their interview held two weeks later. Although the studies mentioned above all confirm that errors generally are not stable, it is important to note that each investigation measures stability using different criteria. However, lack of stability seems to be an issue regardless of how it is measured.

tially concerned to know the most commonly occuring (i.e., the most prevalent) errors in the class. Whereas, a tutor will wish to have access to the actual error-profile for each student.

4. Taxonomy of errors: This study has confirmed and slightly extended the error taxonomy suggested by Sleeman (1983). In our view, it is feasible to talk about the following types of errors:

- Stable errors (both remediable and "resistant")
- Attentional errors (largely minor errors such as adding/dropping signs).
- Classes of mal-rules, used by the same student on different occasions with the same type of task.
- Mental slips, typing/transcription errors.

It is important for a tutor to categorize correctly the error as, for example, it may be counterproductive to tutor a student on an error which is a result of a slip; whereas it maybe important to address a stable error. (These judgements are both subtle and complex.)

Overall conclusions of the series of studies

* For introductory algebra when taught procedurally with this type of student, it appears that Reteaching is as effective as MBR. From which it follows that CAI would be as effective as an ITS It is vital that we investigate the range of subjects, instructional approaches and student age-ranges for which this result holds. (This information will be important for both educators and ITSs workers.)

* Despite the conclusions on the last paragraph, one should not conclude that reteaching by a classroom teacher will be as effective as reteaching in a one-to-one tutoring situation or by a computer system. Immediacy of feedback which is lacking from the normal classroom situation may well be a critical factor, Lewis & Anderson (1985).

* It is critical that ITSs receive extensive field-testing.

* The subfield of ITS should not conclude that the task of building an ITS is impossible, but it should conclude that the task is harder than we had initally thought*, c.f., Winograd & Flores (1986). It is possible that the more global analysis (Moore & Sleeman, 1987) which takes account of the student's performance at several levels might be needed before proper remediation can be undertaken. But before such a system is built it is suggested that experiments are run to see whether human tutors are more effective when they base their remediation on such a global approach.

* Cronbach and Snow (1977) warned that producing truly individualized instruction was a demanding task, but the ITS field chose to infer that their conclusions were unduly pessimistic given the complexity of analysing such a diverse set of experimental studies.

For a more extensive discussion of the conclusions of this experiment including a discussion of possible additional issues to be investigated, see Sleeman, Kelly, Martinak, Ward and Moore (1987).

References

Bricken, W. M. (1987). *Analyzing errors in elementary mathematics*. Doctoral Dissertation, School of Education, Stanford University.

Buckingham, G. E. (1933). *Diagnostic and remedial teaching in first-year algebra*. Northwestern University, IL, School of Education series, No. 11.

Bunderson, V. C. & Olsen, J. B. (1983). *Mental errors in arithmetic skills: their diagnosis in pre-college students*. WICAT Education Institute Provo, UT. Final project report, NSF SED 80-12500.

Cronbach, L. J. & Snow, R. E. (1977). *Aptitudes and Instructional Methods*. New York: Irvington.

Hartley, J. R. & Sleeman, D. H. (1973). Towards intelligent teaching systems. *International Journal of Man-Machine Studies*, *5*, 215-236.

Kelly, A. E., Martinak, R. & Sleeman, D. (1987). *How important is diagnosis for remediation? A concern for intelligent tutoring systems*. Proceedings of the Second European Conference for Research on Learning and Instruction, Tubingen, Federal Republic of Germany, September 19-27.

Kelly, A. E. & Sleeman, D. (1986). *A study of Diagnostic and Remedial Techniques used by Master Algebra Teachers*. Technical Report AUCS/TR8708, Department of Computing Science, University of Aberdeen.

Lewis, M. W. & Anderson, J. R. (1985). Discrimination of operator schemata in problem solving: Learning from examples. Cognitive Psychology, 17, pp 26-65.

Macnab, D. S. & Cummine, J. A. (1986). Teaching Mathematics 11-16: A Difficulty-Centred Approach. London: Basil Blackwell, England.

Martinak, R., Schneider, B. and Sleeman, D. (1987). A Comparative Analysis of Approaches for Correcting Algebra Errors via an Intelligent Tutoring System. Proceedings of AERA, Washington D.C.

Moore, J. L. & Sleeman, D. (1987). Enhancing PIXIE's tutoring capabilities. Technical Report AUCS/TR8709, Department of Computing Science, University of Aberdeen.

Putnam, R. T. (1987). Structuring and adjusting Content for students: A study of live and simulated tutoring of addition. American Educational Research Journal, Vol 24, No 1, pp. 13 - 48

Resnick, L. (1984). Beyond error analysis: the role of understanding in elementary school arithmetic. Learning Research and Development Centre, University of Pittsburgh, PA.

Resnick L. B. & Ford W. W., (1981). The psychology of mathematics for instruction. Hilsdale, NJ: Lawrence Erlbaum Associates.

Sleeman, D. (1982). Assessing aspects of competence in basic algebra. In Intelligent Tutoring Systems D. Sleeman & J. S Brown (Eds). London: Academic Press, pp 185-199.

Sleeman, D. (1983). Basic algebra revisited: a study with 14-year olds. HPP Report 83-9, Computer Science Dept, Stanford University. Republished in International Journal Man-Machine Studies, (1985), 22, pp 127-149.

Sleeman, D. (1987). PIXIE: A shell for developing Intelligent Tutoring Systems. In R.W. Lawler & M. Yazdani (Eds). Artificial Intelligence and Education: Volume 1, (pp 239-265). Norwood, N.J: Ablex.

Sleeman, D. & Brown, J. S. (1982). Intelligent Tutoring Systems, London: Academic Press.

Sleeman, D., Kelly, A. E., Martinak, R., Ward, R. D. & Moore, J. (1987) Diagnosis and Remediation in the context of Intelligent Tutoring Systems. Technical Report AUCS/TR8712, Department of Computing Science, University of Aberdeen.

Steinburg, R. M., Sleeman, D. and Ktorza, D. (1985) Transformations and Equivalence of Equations in High School Algebra. Technical Report, School of Education, Stanford.

Swan, M. B. (1983). Teaching decimal place value. A comparative study of conflict and positive only approaches. Research Report 31, Shell Centre for Mathematical Education, University of Nottingham.

VanLehn, K. (1981). Bugs are not enough: Empirical studies of bugs, impasses and repairs in procedural skills. XEROX, Palo Alto Research Centre, CA. Cognitive and Instructional Science series, CIS-11 (SSL-81-2).

Winograd, T. & Flores, F. F. (1986). Understanding computers and cognition: A new foundation for design. Norwood, N. J.: Ablex.

Appendix A

Pretest Items

1. $7x = 14$
2. $4x = -12$
3. $5x = 7$
4. $8x = 18$
5. $3x = 8 + 4$
6. $3x = 4 * 4$
7. $3x + 4x = 14$
8. $3x + 5 = 26$
9. $7 + 4x = 19$
10. $5x = 4x + 8$
11. $3x = 3(2 + 3)$
12. $12x = 2(3x + 3)$
13. $7x + 2 + 4 * 8$
14. $17x = 19x + 25$
15. $3 + 2x + 4x = 21$
16. $4 + 3x + 4x = 25$
17. $35x = 2 + 3(4x + 5)$

Posttest Items

1. $3x = 9$
2. $4x = -8$
3. $3x = 5$
4. $6x = 9$
5. $5x = 8 + 2$
6. $3x = 4 * 3$
7. $2x + 3x = 10$
8. $2x + 3 + 9$
9. $2 + 4x = 14$
10. $4x + 3x + 6$
11. $2x = 3(3 + 1)$
12. $24x = 3(2x + 3)$
13. $5x = 2 + 3 * 8$
14. $16x = 19x + 20$
15. $2 + 2x + 3x = 17$
16. $3 + 3x + 4x = 24$
17. $37x = 2 + 3(4x + 5)$

18. 3*2x + 3x = 19

19. 3(2x + 4) = 12(9 + 2x)

20. 21x = 3 * 2(2x + 3)

18. 5*2x + 4x = 18

19. 4(3x + 3) = 5(6 + 2x)

20. 24x = 3 * 2(2x + 4)

Appendix B (Sample Scripts)

Reteach

1. [FRESH PAPER]

2. Have student work the task aloud

3. If wrong, say "THIS IS WRONG".

4. [FRESH PAPER]

5. Say, "LET ME SHOW YOU HOW TO DO IT AND WHY" - (using the four rules).

6. GIVE PRACTICE TASKS

FOUR RULES:

1. Precedence - "My Dear Aunt Sally" - Multiply or Divide before Adding or Subtracting

2. "Get all the Xs to one side, all the numbers to the other"

3. "To undo added things, you subtract, to undo multiplied things, you divide."

4. "Whatever you do to one side, you must do the same thing to the other side."

<u>MBR</u>

1. [FRESH PAPER]

2. Have student work the task aloud

3. After the student has completed the task, go back to <u>EACH</u> error, say:

   "IT LOOKS LIKE YOU (DID) .....THIS IS WRONG BECAUSE ..."

   (Address the Four Rules)

4. Say, "LET ME SHOW YOU <u>HOW</u> TO DO IT AND <u>WHY</u>" - (Using the four rules).

5. GIVE PRACTICE TASKS

<u>FOUR RULES</u>

1. Precedence - "My Dear Aunt Sally" - Multiply or Divide before Adding or Subtracting

2. "Get all the Xs to one side, all the numbers to the other"

3. "To undo added things, you subtract,
   to undo multiplied things, you divide."

4. "Whatever you do to one side, *you must do the same thing to the* other side."

MBR

1. [FRESH PAPER]

2. Have student work the task aloud

3. After the student has completed the task, go back to EACH error, say:

   "IT LOOKS LIKE YOU (DID) .....THIS IS WRONG BECAUSE ..."

   (Address the Four Rules)

4. Say, "LET ME SHOW YOU HOW TO DO IT AND WHY" - (Using the four rules).

5. GIVE PRACTICE TASKS

FOUR RULES

1. Precedence - "My Dear Aunt Sally" - Multiply or Divide before Adding or Subtracting

2. "Get all the Xs to one side, all the numbers to the other"

3. "To undo added things, you subtract,
   to undo multiplied things, you divide."

4. "Whatever you do to one side, you must do the same thing to the other side."

Appendix C

ITEMS TO USE FOR TUTORING (the first item is from the pretest)

1. $7x = 14$	2. $4x = -12$
$8x = 16$	$6x = -24$
$3x = 6$	$5x = -15$
3. $5x = 7$	4. $8x = 18$
$6x = 13$	$9x = 15$
$4x = 9$	$6x = 14$
5. $3x = 8 + 4$	6. $3x = 4 * 4$
$5x = 19 + 6$	$2x = 3 * 3$
$2x = 7 + 5$	$8x = 5 * 2$
7. $3x + 4x = 1$	8. $3x + 5 = 26$
$6x + 3x = 36$	$3x + 4 = 19$
$2x + 3x = 20$	$6x + 3 = 21$
9. $7 + 4x = 19$	10. $5x = 4x + 8$
$6 + 4x = 26$	$7x = 5x + 18$
$3 + 5x = 11$	$6x = 3x + 2$
11. $3x = 3(2 + 3)$	12. $12x = 2(3x + 3)$
$4x = 3(4 + 5)$	$17x = 2(4x + 3)$
$5x = 4(1 + 4)$	$14x = 3(3x + 7)$
13. $7x = 2 + 4 * 8$	14. $17x = 19x + 25$
$2x = 2 + 4 * 6$	$15x = 18x + 17$
$5x = 8 + 3 * 4$	$21x = 24x + 13$
15. $3 + 2x + 4x = 21$	16. $4 + 3x + 4x = 25$

Appendix C

ITEMS TO USE FOR TUTORING (the first item is from the pretest)

1. $7x = 14$
   $8x = 16$
   $3x = 6$

2. $4x = -12$
   $6x = -24$
   $5x = -15$

3. $5x = 7$
   $6x = 13$
   $4x = 9$

4. $8x = 18$
   $9x = 15$
   $6x = 14$

5. $3x = 8 + 4$
   $5x = 19 + 6$
   $2x = 7 + 5$

6. $3x = 4 * 4$
   $2x = 3 * 3$
   $8x = 5 * 2$

7. $3x + 4x = 1$
   $6x + 3x = 36$
   $2x + 3x = 20$

8. $3x + 5 = 26$
   $3x + 4 = 19$
   $6x + 3 = 21$

9. $7 + 4x = 19$
   $6 + 4x = 26$
   $3 + 5x = 11$

10. $5x = 4x + 8$
    $7x = 5x + 18$
    $6x = 3x + 2$

11. $3x = 3(2 + 3)$
    $4x = 3(4 + 5)$
    $5x = 4(1 + 4)$

12. $12x = 2(3x + 3)$
    $17x = 2(4x + 3)$
    $14x = 3(3x + 7)$

13. $7x = 2 + 4 * 8$
    $2x = 2 + 4 * 6$
    $5x = 8 + 3 * 4$

14. $17x = 19x + 25$
    $15x = 18x + 17$
    $21x = 24x + 13$

15. $3 + 2x + 4x = 21$

16. $4 + 3x + 4x = 25$

$2 + 3x + 4x = 16$

$4 + 3x + 7x = 9$

$7 + 2x + 3x = 19$

$5 + 5x + 6x = 33$

17. $35 = 2 + 3(4x + 5)$

$29x = 3 + 4(3x + 7)$

$19x = 5 + 2(2x + 3)$

18. $3 * 2x + 3x = 19$

$4 * 3x + 2x = 25$

$2 * 6x + 3x = 40$

19. $3(2x + 4) = 12(9 + 2x)$

$4(5x + 2) = 6(8 + 5x)$

$7(2x + 2) = 5(3 + 6x)$

20. $21x = 3 * 2(2x + 3)$

$14x = 2 * 3(2x + 5)$

$36x = 2 * 3(4x + 5)$

$2 + 3x + 4x = 16$

$4 + 3x + 7x = 9$

$7 + 2x + 3x = 19$

$5 + 5x + 6x = 33$

17. $35x = 2 + 3(4x + 5)$

    $29x = 3 + 4(3x + 7)$

    $19x = 5 + 2(2x + 3)$

18. $3 * 2x + 3x = 19$

    $4 * 3x + 2x = 25$

    $2 * 6x + 3x = 40$

19. $3(2x + 4) = 12(9 + 2x)$

    $4(5x + 2) = 6(8 + 5x)$

    $7(2x + 2) = 5(3 + 6x)$

20. $21x = 3 * 2(2x + 3)$

    $14x = 2 * 3(2x + 5)$

    $36x = 2 * 3(4x + 5)$

Appendix D (Sample Scripts)

Reteach

1. [FRESH PAPER]

2. Have student work the task aloud

3. If wrong, say "THIS IS WRONG".

4. [FRESH PAPER]

5. Say, "LET ME SHOW YOU HOW TO DO IT AND WHY" - (using the four rules).

6. GIVE PRACTICE TASKS

FOUR RULES:

1. Precedence - "My Dear Aunt Sally" - Multiply or Divide before Adding or Subtracting

2. "Get all the Xs to one side, all the numbers to the other"

3. "To undo added things, you subtract, to undo multiplied things, you divide."

4. "Whatever you do to one side, you must do the same thing to the other side."

Appendix D (Sample Scripts)

Reteach

1. [FRESH PAPER]

2. Have student work the task aloud

3. If wrong, say "THIS IS WRONG".

4. [FRESH PAPER]

5. Say, "LET ME SHOW YOU HOW TO DO IT AND WHY" - (using the four rules).

6. GIVE PRACTICE TASKS

FOUR RULES:

1. Precedence - "My Dear Aunt Sally" - Multiply or Divide before Adding or Subtracting

2. "Get all the Xs to one side, all the numbers to the other"

3. "To undo added things, you subtract,
 to undo multiplied things, you divide."

4. "Whatever you do to one side, you must do the same thing to the other side."

MBR

1. [FRESH PAPER]

2. Have student work the task aloud

3. After the student has completed the task, go back to EACH error, say:

   "IT LOOKS LIKE YOU (DID) .....THIS IS WRONG BECAUSE ..."

   (Address the Four Rules)

4. Say, "LET ME SHOW YOU HOW TO DO IT AND WHY" - (Using the four rules).

5. GIVE PRACTICE TASKS

FOUR RULES

1. Precedence - "My Dear Aunt Sally" - Multiply or Divide before Adding or Subtracting

2. "Get all the Xs to one side, all the numbers to the other"

3. "To undo added things, you subtract,
   to undo multiplied things, you divide."

4. "Whatever you do to one side, you must do the same thing to the other side."

MBR

1. [FRESH PAPER]

2. Have student work the task aloud

3. After the student has completed the task, go back to EACH error, say:

   "IT LOOKS LIKE YOU (DID) .....THIS IS WRONG BECAUSE ..."

   (Address the Four Rules)

4. Say, "LET ME SHOW YOU HOW TO DO IT AND WHY" - (Using the four rules).

5. GIVE PRACTICE TASKS

FOUR RULES

1. Precedence - "My Dear Aunt Sally" - Multiply or Divide before Adding or Subtracting

2. "Get all the Xs to one side, all the numbers to the other"

3. "To undo added things, you subtract, to undo multiplied things, you divide."

4. "Whatever you do to one side, you must do the same thing to the other side."

MBR + "cognitive engagement"

1. [FRESH PAPER]

2. Have the student work the task aloud

3. "TELL ME HOW (AND WHY) YOU DID (the errors)" No need to review correct steps. [Student "targets" own errors]

4. For each error say, "THIS IS WRONG."

5. Say, "LET ME SHOW YOU HOW TO DO IT AND WHY" - (Using the four rules).

6. [FRESH PAPER]

7. Re-present task.

   Say, "NOW YOU TELL ME HOW/WHY TO DO THIS TASK"

   You need a how and why for each step, if possible. If (s)he makes an error, correct it on the spot; (s)he does not have to repeat this step.

8. GIVE PRACTICE TASKS

MBR + "cognitive engagement"

1. [FRESH PAPER]

2. Have the student work the task aloud

3. "TELL ME HOW (AND WHY) YOU DID (the errors)" No need to review correct steps. [Student "targets" own errors]

4. For each error say, "THIS IS WRONG."

5. Say, "LET ME SHOW YOU HOW TO DO IT AND WHY" - (Using the four rules).

6. [FRESH PAPER]

7. Re-present task.

   Say, "NOW YOU TELL ME HOW/WHY TO DO THIS TASK"

   You need a how and why for each step, if possible. If (s)he makes an error, correct it on the spot; (s)he does not have to repeat this step.

8. GIVE PRACTICE TASKS

FOUR RULES

1. Precedence - "My Dear Aunt Sally" - Multiply or Divide before Adding or Su .racting

2. "Get all the Xs to one side, all the numbers to the other"

3. "To undo added things, you subtract,
to undo multiplied things, you divide."

4. "Whatever you do to one side, you must do the same thing to the other side."

FOUR RULES

1. Precedence - "My Dear Aunt Sally" - Multiply or Divide before Adding or Subtracting

2. "Get all the Xs to one side, all the numbers to the other"

3. "To undo added things, you subtract, to undo multiplied things, you divide."

4. "Whatever you do to one side, you must do the same thing to the other side."

MBR + "cognitive dissonance"

1 [FRESH PAPER]

2. Have the student work the task aloud

3. Even if the answer is correct, say,

"PLEASE CHECK YOUR ANSWER"

Have the student substitute their answer for X. If they don't remember substitution, remind them. Substitute in their wrong answer, and have them agree that the two sides do not balance. Ask, "HOW DO WE KNOW THAT IT IS WRONG?" — because the two sides don't balance. If they seem TOTALLY lost, you should give a very obvious example (e.g., 5X=15).

If their wrong answer DOES balance the sides, say, "EVEN THOUGH THIS VALUE FOR X IS RIGHT, YOU GOT IT FOR THE WRONG REASON.....(and explain)...

4. Say, "THIS IS WRONG"/"THIS METHOD IS WRONG."

5. Say, "LET ME SHOW YOU HOW TO DO IT" (And WHY - Using the four rules)" Do entire task.

6. GIVE PRACTICE TASKS

MBR + "cognitive dissonance"

1 [FRESH PAPER]

2. Have the student work the task aloud

3. Even if the answer is correct, say,

   "PLEASE CHECK YOUR ANSWER"

   Have the student substitute their answer for X. If they don't remember substitution, remind them. Substitute in their wrong answer, and have them agree that the two sides do not balance. Ask, "HOW DO WE KNOW THAT IT IS WRONG?" — because the two sides don't balance. If they seem TOTALLY lost, you should give a very obvious example (e.g., 5X=15).

   If their wrong answer DOES balance the sides, say, "EVEN THOUGH THIS VALUE FOR X IS RIGHT, YOU GOT IT FOR THE WRONG REASON.....(and explain)...

4. Say, "THIS IS WRONG"/"THIS METHOD IS WRONG."

5. Say, "LET ME SHOW YOU HOW TO DO IT" (And WHY - Using the four rules)" Do entire task.

6. GIVE PRACTICE TASKS

FOUR RULES

1. Precedence - "My Dear Aunt Sally" - Multiply or Divide before Adding or Subtracting

2. "Get all the Xs to one side, all the numbers to the other"

3. "To undo added things, you subtract, to undo multiplied things, you divide."

4. "Whatever you do to one side, you must do the same thing to the other side."

FOUR RULES

1. Precedence - "My Dear Aunt Sally" - Multiply or Divide before Adding or Subtracting

2. "Get all the Xs to one side, all the numbers to the other"

3. "To undo added things, you subtract,
to undo multiplied things, you divide."

4. "Whatever you do to one side, you must do the same thing to the other side."

Appendix E (Classification of errors)

Algebraic Errors

1. Changed the side but not sign of an x-term

2. Changed the side of the larger positive x-term but placed the minus sign before the smaller x-term.

3. Added negative sign(s), e.g., $3x+5=9+6 \Rightarrow 3x=-9-6-5$

4. Added an x-term to a constant, e.g., $4x+3 \Rightarrow 7x$

5. Ended with the step $ax=bx$.

6. Inverted division. $ax=b \Rightarrow x=a/b$

7. Subtracted the coefficient.

8. Changed the side of a multiplier, e.g., $5 * 3x + 7 = 9 \Rightarrow 3x + 7 = 9 * 5$

9. Dropped an x, e.g., $3x+3=6x \Rightarrow 3x+3=6$

10. Other low frequency algebraic errors.

Appendix E (Classification of errors)

Algebraic Errors

1. Changed the side but not sign of an x-term

2. Changed the side of the larger positive x-term but placed the minus sign before the smaller x-term.

3. Added negative sign(s), e.g., 3x+5=9+6 => 3x=-9-6-5

4. Added an x-term to a constant, e.g., 4x+3 => 7x

5. Ended with the step ax=bx.

6. Inverted division. ax=b => x=a/b

7. Subtracted the coefficient.

8. Changed the side of a multiplier, e.g., 5 * 3x + 7 = 9 => 3x + 7 = 9 * 5

9. Dropped an x, e.g., 3x+3=6x => 3x+3=6

10. Other low frequency algebraic errors.

Non algebraic errors

1. Precedence errors
2. Distributive property errors
3. Misreading a sign/number
4. Incomplete working
5. Item not attempted
6. Arithmetic errors
7. Dropping a negative sign
8. Other non-algebraic errors.

Non algebraic errors

1. Precedence errors

2. Distributive property errors

3. Misreading a sign/number

4. Incomplete working

5. Item not attempted

6. Arithmetic errors

7. Dropping a negative sign

8. Other non-algebraic errors.

Appendix F

PRETEST

NAME: ______________________ Date of Birth: __________

Teacher: ___________________ Today's Date: __________

INSTRUCTIONS: Please solve for x. Please show all of the steps you take for each task. Use the back of the paper if you need more space.

1. $3x = 9$	2. $2x = 10$	3. $9x = 18$
4. $3x = -9$	5. $2x = -10$	6. $9x = -18$
7. $5x = 3$	8. $7x = 2$	9. $8x = 5$
10. $5x = -3$	11. $7x = -2$	12. $8x = -5$
13. $3x = 5$	14. $7x = 9$	15. $5x = 7$
16. $3x = -5$	17. $7x = -9$	18. $5x = -7$
19. $4 + 3x = 19$	20. $5 + 2x = 20$	21. $6 + 4x = 21$
22. $9x + 5x = 70$	23. $6x + 4x = 20$	24. $8x + 5x = 26$
25. $3x + 4 = 19$	26. $2x + 8 = 21$	27. $7x + 9 = 13$
28. $7x = 5x + 17$	29. $12x = 9x + 21$	30. $10x = 6x + 19$
31. $2x = 3(3 + 1)$	32. $4x = 3(2 + 5)$	33. $6x = 3(4 + 2)$
34. $7x = 2(5x + 6)$	35. $8x = 4(3x + 4)$	
36. $3x = 3(5x + 6)$	37. $2x = 2 + 4 * 6$	
38. $4x = 12 + 8 * 2$	39. $5x = 6 + 2 * 3$	

Appendix F

PRETEST

NAME: ____________________________ Date of Birth: _____________

Teacher: _______________________ Today's Date: _____________

INSTRUCTIONS: Please solve for x. Please show all of the steps you take for each task. Use the back of the paper if you need more space.

1. $3x = 9$	2. $2x = 10$	3. $9x = 18$
4. $3x = -9$	5. $2x = -10$	6. $9x = -18$
7. $5x = 3$	8. $7x = 2$	9. $8x \quad 5$
10. $5x = -3$	11. $7x = -2$	12. $8x = -5$
13. $3x = 5$	14. $7x = 9$	15. $5x = 7$
16. $3x = -5$	17. $7x = -9$	18. $5x = -7$
19. $4 + 3x = 19$	20. $5 + 2x = 20$	21. $6 + 4x = 21$
22. $9x + 5x = 70$	23. $6x + 4x = 20$	24. $8x + 5x = 26$
25. $3x + 4 = 19$	26. $2x + 8 = 21$	27. $7x + 9 = 13$
28. $7x = 5x + 17$	29. $12x = 9x + 21$	30. $10x = 6x + 19$
31. $2x = 3(3 + 1)$	32. $4x = 3(2 + 5)$	33. $6x = 3(4 + 2)$
34. $7x = 2(5x + 6)$	35. $8x = 4(3x + 4)$	
36. $3x = 3(5x + 6)$	37. $2x = 2 + 4 * 6$	
38. $4x = 12 + 8 * 2$	39. $5x = 6 + 2 * 3$	

40. $2 + 3x + 4x = 16$

41. $4 + 5x + 2x = 25$

42. $7 + 9x + 4x = 24$

43. $36x = 2 + 3(4x + 5)$

44. $30x = 2 + 7(2x + 6)$

45. $31x = 4 + 2(4x + 6)$

46. $2 * 3x + 2x = 12$

47. $4 * 3x + 4x = 7$

48. $7 * 2x + 6x = 13$

49. $20x = 3 * 3(2x + 4)$

50. $22x = 5 * 2(2x + 3)$

51. $31x = 4 * 2(2x + 3)$

40. $2 + 3x + 4x = 16$

41. $4 + 5x + 2x = 25$

42. $7 + 9x + 4x = 24$

43. $36x = 2 + 3(4x + 5)$

44. $30x = 2 + 7(2x + 6)$

45. $31x = 4 + 2(4x + 6)$

46. $2 * 3x + 2x = 12$

47. $4 * 3x + 4x = 7$

48. $7 * 2x + 6x = 13$

49. $20x = 3 * 3(2x + 4)$

50. $22x = 5 * 2(2x + 3)$

51. $31x = 4 * 2(2x + 3)$

POSTTEST

NAME: ______________________ Date of Birth: __________

Teacher: __________________ Today's Date: __________

<u>INSTRUCTIONS</u>: <u>Please solve for x</u>. <u>Please show all of the steps you</u> take for each task. Use the back of the paper if you need more space.

1. $3x = 9$ 2. $7x = 14$ 3. $8x = 16$

4. $3x = -9$ 5. $7x = -14$ 6. $8x = -16$

7. $5x = 3$ 8. $7x = 5$ 9. $3x = 2$

10. $5x = -3$ 11. $7x = -5$ 12. $3x = -2$

13. $3x = 5$ 14. $8x = 11$ 15. $4x = 7$

16. $3x = -5$ 17. $8x = -11$ 18. $4x = -7$

19. $4 + 3x = 19$ 20. $5 + 3x = 21$ 21. $6 + 7x = 24$

22. $9x + 5x = 70$ 23. $2x + 8x = 40$ 24. $3x + 7x = 60$

25. $3x + 4 = 19$ 26. $6x + 4 = 15$ 27. $7x + 2 = 6$

28. $7x = 5x + 17$ 29. $9x = 6x + 12$ 30. $12x = 6x + 23$

31. $2x = 3(3 + 1)$ 32. $3x = 2(3 + 4)$ 33. $7x = 5(4 + 3)$

34. $7x = 2(5x + 6)$ 35. $2x = 4(3x + 5)$

36. $4x = 6(5x + 8)$ 37. $2x = 2 + 4 * 6$

38. $4x = 12 + 3 * 4$ 39. $6x = 3 + 3 * 2$

40. $2 + 3x + 4x = 16$ 41. $3 + 6x + 2x = 19$

POSTTEST

NAME: ______________________ Date of Birth: __________

Teacher: ____________________ Today's Date: __________

INSTRUCTIONS: Please solve for x. Please show all of the steps you take for each task. Use the back of the paper if you need more space.

1. $3x = 9$	2. $7x = 14$	3. $8x = 16$
4. $3x = -9$	5. $7x = -14$	6. $8x = -16$
7. $5x = 3$	8. $7x = 5$	9. $3x = 2$
10. $5x = -3$	11. $7x = -5$	12. $3x = -2$
13. $3x = 5$	14. $8x = 11$	15. $4x = 7$
16. $3x = -5$	17. $8x = -11$	18. $4x = -7$
19. $4 + 3x = 19$	20. $5 + 3x = 21$	21. $6 + 7x = 24$
22. $9x + 5x = 70$	23. $2x + 8x = 40$	24. $3x + 7x = 60$
25. $3x + 4 = 19$	26. $6x + 4 = 15$	27. $7x + 2 = 6$
28. $7x = 5x + 17$	29. $9x = 6x + 12$	30. $12x = 6x + 23$
31. $2x = 3(3 + 1)$	32. $3x = 2(3 + 4)$	33. $7x = 5(4 + 3)$
34. $7x = 2(5x + 6)$	35. $2x = 4(3x + 5)$	
36. $4x = 6(5x + 8)$	37. $2x = 2 + 4 * 6$	
38. $4x = 12 + 3 * 4$	39. $6x = 3 + 3 * 2$	
40. $2 + 3x + 4x = 16$	41. $3 + 6x + 2x = 19$	

42. $4 + 2x + 7x = 24$

43. $36x = 2 + 3(4x + 5)$

44. $38x = 3 + 6(3x + 5)$

45. $31x = 3 + 5(2x + 3)$

46. $2 * 3x + 2x = 12$

47. $2 * 7x + 2x = 7$

48. $5 * 3x + 4x = 11$

49. $20x = 3 * 3(2x + 4)$

50. $23x = 3 * 3(2x + 5)$

51. $35x = 2 * 5(2x + 4)$