1984

Identification And Remediation Of Children's Errors In Subtraction

Larry Anthony Cebulski

Identification and Remediation of Children's Errors in Subtraction

by

Larry A. Cebulski

Department of Psychology

Submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Faculty of Graduate Studies

The University of Western Ontario

London, Ontario

November, 1983

Abstract

Three studies were conducted in order to determine the source and frequency of children's difficulties in subtraction and to examine different approaches to remediation. In Study 1, 56 third grade children were asked to solve subtraction problems and were observed and questioned regarding their solution processes. Analysis of verbal reports and written solutions suggested that the main source of difficulty involved borrowing procedures. Children who had difficulty either attempted to borrow incorrectly or made inversion errors, that is, they ignored the location of the digits and subtracted the smaller number from the larger. Study 2 examined two minimally intrusive methods of remediation. Eighty third grade children were given either instructions to borrow, promised rewards for accurate performance or no intervention and were asked to solve a series of subtraction problems requiring borrowing. Neither experimental condition resulted in a significant increase in the number of problems solved correctly. Those children who initially failed to solve any problems correctly responded to instructions with a decrease in inversion errors and an increase in borrowing errors. These results suggested that more intensive instruction was required. In Study 3, 67 third and fourth grade children were assigned to one of three conditions: Component skills Training, Criterion Training or a regular classroom control condition. The Component Skills Training condition attempted to teach the skills required for borrowing in a step by step fashion with feedback, while the Criterion Training condition simply provided feedback in the form of correctly worked solutions. Children who solved fewer than 60 percent of the problems correctly on a pretest significantly increased the

number of problems solved correctly on posttests conducted 1-2 days and two weeks following training. These increases were reflected by a reduction of errors involving borrowing. Those children who did not complete the training programs successfully showed patterns of performance across days which were suggestive of fatigue, boredom, or the presence of some other interfering factor. In all, the results suggested that intensive programming through careful attention to contingencies, as well as an appreciation of how program variables interact with learner characteristics and development may be necessary for effective remedial intervention in subtraction.

Acknowledgements

I wish to express my gratitude to my advisor, Dr. Bradley Bucher, for his invaluable guidance and support throughout this project. Thanks also to Drs. Marvin Simner, Stephen Lupker, Harry Murray, David Pederson and Charles Brainerd for their comments and suggestions on earlier drafts of this paper. Finally I would like to express my appreciation to the principals, teachers, and especially, to the children of C.C. Carrothers, Evelyn Harrison, Victoria, Kensal Park, J.P. Robarts, Lord Nelson and Sir Winston Churchill public schools, whose cooperation and participation made this project possible.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

Introduction

Studies which have examined children's arithmetic performance reveal that subtraction presents more difficulty for elementary school children than does addition, at least in the primary grades (Cox, 1974; Ginsburg, 1977; Reiss, 1943). It also appears that this problem becomes more pronounced when borrowing is required (Cox, 1974, Ellis, 1972; Graeber and Wallace 1977).

Although it seems that subtraction is a problem area, reports of why this is the case conflict. Most research designed to find effective remedial procedures (Harvey and Kyte, 1965; Post and Reys, 1979; Brueckner and Bond, 1955; Fuson, 1979; Mick and Brazier, 1979; Carpenter, 1980; Parson, 1973 and others) have not based these procedures on hypotheses concerning the causes of errors in any consistent fashion. Attempts to systematically evaluate cost effectiveness are also absent. Both of these considerations are important if effective remedial procedures are to leave the laboratory and enter the classroom.

The purpose of this thesis was twofold. First, a normative investigation of the sources of children's errors on subtraction problems involving borrowing was conducted. Second, an attempt was made to construct and implement a remedial program for children who had difficulty with subtraction. My goal was to construct an effective remedial package which could be easy to administer, which would be appropriate for use by a classroom teacher in a regular class, and which would result in rapid improvements in performance which could then be maintained by contingencies existing within the classroom. This required an evaluation of four different approaches to remediation:

modification of antecedent conditions, motivation enhancement, instruction in component skills and use of feedback.

Although an attempt was made to draw on findings from the cognitive and learning literature when designing these programs, the motivating force behind this investigation was empirical. That is, a procedure which is effective in reducing the errors of children selected from a local school system was sought.

Models of Children's Solutions

The logical place to begin a study of children's errors is to examine the literature dealing with how arithmetic problems are solved. Cumulative evidence from a number of studies of children's and adults' mental arithmetic suggests a developmental progression from counting to fact retrieval for most simple addition and subtraction problems. The evidence comes from two types of research. The first type is largely qualitative in nature and examines children's solution strategies for problems involving the addition and subtraction of small numbers of objects (Gelman and Tucker, 1975; Gelman and Gallistel, 1978) and for arithmetic algorithms (Ginsburg, 1977). Gelman and Gallistel (1978) report that the 3, 4 and 5 year olds in their experiments requiring numerosity estimations engaged in counting whenever the opportunity presented itself. Further, much of the child's development in the numerical sphere from about age 3 involves the improvement of particular counting skills rather than the acquisition of many new ones. Ginsburg (1977), following intensive observations of individual children, reports the frequent use of counting procedures when children attempt to solve addition and subtraction problems, and describes some of the more popular counting strategies.

The largest body of evidence that preschoolers use counting strategies to answer arithmetic questions comes from studies examining chronometric models of addition and subtraction (Groen and Parkman, 1972; Groen and Poll, 1973; Woods, Resnick and Groen, 1975). Chronometric analysis, based on Sternberg's (1966 and 1969) early research into psychological models of cognitive processing, involves the partitioning of reaction time into separate, additive components. This type of analysis requires that a process such as addition be broken down into a number of identical steps, where the response latency forms a linear function of the number of times a step must be repeated for the solution of a problem. The counting models employed by Groen and his associates assumed two operations: 1) setting a counter to a value from which to begin counting and, 2) incrementing or decrementing by one. It was assumed that the setting time is independent of the value to which the counter is set and that the incrementing or decrementing time is constant. Also, it was assumed that there exits a mechanism for keeping track of the number of times the counter has been incremented or decremented.

These studies examined children age 6 to 9. Only numerals smaller than 10 were used. Children were required to press one out of 10 buttons (0 to 9) in response to an arithmetic algorithm presented on an opaque screen.

For problems of the format a-b=_, children appeared to use a process which involves incrementing or decrementing, whichever is faster. That is, the counter is set to "a" and decremented "b" times, or the counter is set to "b" and incremented "_" times until "a" is reached. This latter procedure was observed by Ginsburg (1977) who referred to it as "counting on". For problems of the type a+b=_, the model which best fit the data involved setting the counter to the maximum of "a" or "b" and then adding the minimum of "a" or "b" by increments of one. For problems of the format _+b=c, the data were not fit by any of the models examined, although the error rates and reaction times were not substantially different from those seen in the other problems.

When problems involved "ties", such as _+2=4, or 2+2=_, or 4-2=_, the data were not fit by any of the models, and the short response latencies suggested that the subjects retrieved the answers from memory.

For problems of the type 4-_=2 (Woods et al., 1975), older children (4th graders) responded with shorter latencies than younger ones (2nd graders), perhaps due to more experience with these types of problems. This last study also revealed that some second graders used a model which consistently involved decrementing only. No fourth grade subject responded in this manner, lending some credence to the suggestion that some children move from a consistent rule bound strategy to a more heuristic procedure which more accurately reflects specific problem demands.

Around the third or fourth grade, it appears that children begin to rely less on counting and more on number fact retrieval in order to solve simple addition and subtraction algorithms. Ashcraft and Fireman (1982) examined the performance of subjects from grades 1, 3, 4, 5, 6 and college on algorithms arranged in a true-false format. They found that chronometric models of counting increasingly failed to account for the reaction times of subjects from grade 4 upward. Fact retrieval appeared to be the process most relied on by adults, except when errors occurred (Ashcraft and Battaglia, 1978; Ashcraft and Stazyk, 1981).

The above studies suggest a developmental progression from a reliance on procedural knowledge of counting to the use of a network of stored number facts in order to solve simple arithmetic problems. This progression refers to children's mental arithmetic involving single digit algorithms only. Also, this progression addresses only the processing or software aspects of developing arithmetical competency. Brainerd (1981), using small visual arrays of objects, presents evidence suggesting that changes in system hardware, namely short term memory, accounted for a greater amount of age related improvement than processing developments. This was most noticeably the case when the encoding demands were larger and where processing demands were minimal. However, when encoding demands were decreased (by encouraging positional versus numerosity encoding of linear arrays) and when processing demands were increased (subtraction versus addition), processing failure accounted for a greater proportion of children's errors.

In an attempt to integrate the above findings, it follows that for the types of problems studied by Groen et al. and Ashcraft et al., short term memory improvements and processing improvements should go hand in hand. That is, as short term memory capacity improves, information regarding counting procedures and heuristics should be handled much more efficiently (perhaps in the encoding process) in order to facilitate processing. This would result in the ability to form a network representation of stored number facts and procedures from which to draw in order to solve arithmetic problems. The shift from a use of procedures by rote to a reliance on recall of information from memory begins to occur at around grade 3 or 4.

However, at this grade level, children also learn to solve more complicated arithmetic problems presented in algorithm format. The need to "borrow" and to "carry" in these problems would be expected to increase encoding and short term memory demands as well as processing requirements. With these types of problems it appears that errors could stem from two sources. In the first, information about borrowing procedures could be encoded correctly, organized effectively and stored as a set of rules and heuristics. Errors, in this case, would most likely be the result of failure to retrieve the necessary information or of problems in the execution of these heuristics, that is, problems in processing. Errors which result from this process would likely be erratic, with correct solutions interspersed by errors with different topographies. In the second source of errors, short term memory capacity might be overloaded by the sheer amount and complexity of the information about procedures for borrowing. This could result in failure to store these heuristics, and would result in errors which reflect gaps in the child's knowledge about borrowing. Failure to develop a stored representation of heuristics (or of number facts for that matter) could also be a result of things like failure to organize the incoming information to facilitate storage. (In all this, it is assumed that the short term memory component includes the organization of information for storage while in working memory). To further complicate matters, it is also recognized that difficulties in processing could lead to failure to develop a stored representation of number facts or heuristics as well. This would be the case where behavioral styles, attention span, etc. contribute to incorrect encoding of incoming information, or where no attempt is made to rehearse or organize information while in short term memory or to retrieve the information from long term store. As we shall see, when subtraction problems involve borrowing, it appears that errors mainly reflect gaps in knowledge and further suggests that children who have difficulty do not have a complete network of stored heuristics about borrowing, or else do not retrieve them. As a result, the errors they make appear to be the result of the use of faulty procedures.

Hypotheses Concerning Children's Errors

Although they are usually not interpreted in these terms, use of faulty procedures appears to be most frequently implicated when children make subtraction errors. Ginsburg (1977) suggests that errors result from an incorrect application of strategies and rules. For example, one skill required to solve the problem $\begin{array}{r} 47 \\ -19 \\ \hline \end{array}$ demands that the child locate the largest number in the ones' column. If it is located in the minuend (top), the child must then compute the correct answer. If it is located in the subtrahend (bottom), the child must apply a set of borrowing skills, that is, he or she must borrow ten from the 4 in the tens' place, add that ten to the 7 in the ones' place, then subtract 9 from 17, as in the example,

$$\begin{array}{r} \overset{3}{\not{4}}\,{}_{1}7 \\ -1\;\;9 \\ \hline 2\;\;8 \end{array} .$$

In the case where the larger number is located in the subtrahend, the child does not have to apply these borrowing skills to arrive at an answer. Instead, he or she may simply subtract the smaller number from the larger, regardless of its location in the problem. Application of this erroneous strategy results in failure to produce the correct answer, as in the following example,

$$\begin{array}{r} 47 \\ -19 \\ \hline 32 \end{array} .$$

This error of always subtracting the smaller number from the larger has been referred to as an inversion error (Blankenship, 1976; Cox, 1974; Graeber and Wallace, 1977; Smith, 1968).

Another example of children's errors in applying the correct procedures is taken from the many examples provided by Graeber and Wallace (1977). In the problem

$$\begin{array}{r} 47{}_{1}4 \\ -145 \\ \hline 339 \end{array} ,$$

the child appeared to have applied some of the borrowing rules correctly. It looks as if the child recognized that borrowing was required and he/she therefore added ten to the 4 in the minuend of the ones' column in order to make it 14. It also appears that he/she forgot to reduce 7 in the tens' column to make it 6. All computations were done correctly. Brown and Burton (1978) and Young and O'Shea (1981) also present evidence that errors children make on subtraction problems requiring borrowing are the result of the use of faulty procedures or the result of omissions of procedures.

<u>Frequency of Errors</u>

Studies investigating the frequency with which various types of errors are committed differ with respect to the types of problems used, the conditions under which children solve the problems, the age and characteristics of the subjects, the methods of assessing which types of errors have been committed, and the error categories used. Given these differences, it is surprising that there is a high level of agreement concerning the most frequent source of error when children solve subtraction problems which require borrowing.

In one of the most frequently cited of these studies, Cox (1974) administered addition, subtraction, multiplication and division problems to children in grades 2 through 6 and to children in special education classes. Only the results for subtraction will be presented here. In her study, six types of subtraction problems were presented. These were:

1) two digit minuend and one digit subtrahend with no borrowing, as in the example $\begin{array}{r} 37 \\ -\ 4 \\ \hline \end{array}$,

2) two digit minuend and one digit subtrahend with borrowing, for example $\begin{array}{r} 37 \\ -\ 9 \\ \hline \end{array}$,

3) two digit minuend and subtrahend with no borrowing,

4) two digit minuend and subtrahend with borrowing,

5) three digit minuend and two digit subtrahend with borrowing in the ones' column, for example $\begin{array}{r} 493 \\ -\ 45 \\ \hline \end{array}$,

6) four digit minuend and three digit subtrahend with borrowing in more than one column.

Classroom teachers administered problem sheets to the children on a weekly basis, one per week. This process was continued until all 6 problem types (as well as problems from the other four operations) had been completed. The children included in this study all had received instruction in solving the particular subtraction problems used and were expected to know the number facts to 18.

Judgements concerning the types of errors that were committed were based on the written solutions produced by each child. No interscorer reliability for these judgements was presented. A problem sheet was considered to contain a systematic error if 3 out of the 5 problems contained errors of the same type. Random errors were said to have occurred if 3 of the 5 problems were solved incorrectly but contained different types of errors. Careless errors referred to situations where only one or two out of 5 problems were solved incorrectly. The statistic of interest to Cox was the number of problem sheets containing systematic errors, random errors, careless errors and no errors. In addition, the types of systematic errors committed and the number of problem sheets containing each error type were also listed.

Of interest here is the frequency of each type of subtraction error. These data were reported for each type of problem but not for separate grades. Grades 2 through 6 were treated as one group and special education classes were treated as another group. For problem types 2, 4 and 5, that is, 2 and 3 digit problems requiring borrowing in one or two columns, inversion was the most frequently occurring type of systematic error. When problems became more difficult, that is, when they contained more than 3 digits and required borrowing in 3 columns (type 6), multiple borrowing errors were found most frequently (a more elaborate definition of this type of error was not provided). In this last case, inversion was the next most frequent category of error.

Graeber and Wallace (1977) conducted a similar study with grades 1 through 4. The children were tested in order to determine their knowledge of subtraction skills so that they could begin individual programmed instruction at the appropriate level (no post-instruction data were provided). Five types of subtraction problems were presented, with 5 to 10 problems adminstered for each type. The types of problems were:

1) two digit minuend and one digit subtrahend without borrowing,

2) two digit minuend and one digit subtrahend with borrowing,

3) two digit minuend and two digit subtrahend with borrowing,

4) three or four digit minuend and subtrahend with borrowing in the ones' column,

5) three or four digit minuend and subtrahend with borrowing in all locations.

In an attempt to be consistent with Cox (1974), Graeber and Wallace attempted to classify errors the same way. When three or more errors of the same type occurred, the error was considered systematic. Data were again reported for all grades taken as a group. As in the Cox study, the most frequently occurring systematic error was inversion.

These studies used relatively similar procedures, overlapping age ranges and problem types, and they obtained similar results. They show that borrowing problems cause difficulties and inversion is the most frequent source of error on these problems. Although the results are fairly clear, the definition of systematic errors poses some interpretational problems. The definition fails to take into consideration the actual number of times the error was committed. Neither does it consider the number of opportunities when that error could have been committed. The classification of an error as systematic in the above manner would not yield much descriptive or diagnostic information. For example, when a borrowing error is committed on 3 out of 5 problems and the other 2 problems are solved correctly, a systematic error in borrowing is said to have occurred. If a borrowing error is committed on all 5 problems, again a systematic error is said to have occurred. In the former case, the fact that some problems were solved correctly suggests that the child has some of the necessary skills to solve the problems, while failure on all 5 problems, suggests that the second child does not. To consider the two sets of responses as equal, is misleading because the source of the two errors is likely different.

Smith (1968) and Ellis (1972) also found inversion errors to be the most frequently occurring errors in subtraction with borrowing. Smith (1968) investigated the relationship between the type and frequency of subtraction errors and children's knowledge of place value. Only the error frequency results are relevant here. A 50 item multiple choice test was administered to third and fourth grade students. Each item consisted of a subtraction problem containing 2, 3 or 4 digits in the minuend and subtrahend (with or without borrowing) and 4 possible answers. Errors on the subtraction test were classified into eleven categories based on which of the responses were selected by the subject. No detailed description of how these error selections were made was provided. Neither was information provided concerning the reliability of these selections.

Results showed that 70 percent of the grade 4 and 80 percent of the grade 3 pupils committed one or more systematic errors in subtraction (systematic errors were again defined as any error occurring three or more times on a particular test). Inversion was found to be the most frequently occurring systematic error, committed by 58.5 percent of all students.

This study, like that of Cox (1974) and Graeber and Wallace (1977) used systematic errors as the dependent measure and therefore poses similar problems of interpretation. In the absence of criteria for making error classifications, it is hard to evaluate whether errors were reliably classified. For example, if children did not actually attempt to compute the correct answer for a subtraction problem, but instead simply wrote down an answer, no information about how the child arrived at this answer would be available. A subsequent classification of the child's choice as a particular type of error would seem, at least to this author, to leave a wide margin for error.

Ellis (1972) presented grade 6 students with screening tests consisting of 8 different types of subtraction problems. The most simple type of problem contained a two digit minuend and a one digit subtrahend and required no borrowing, while the most difficult type contained 5 digits in both minuend and subtrahend and required borrowing in three columns. On the basis of performance on this screening test, those students who "... demonstrated the ability to perform some of the tasks but had difficulties in some areas" were selected for further study. Those students who made errors on every problem were excluded from further study.

Subjects were given diagnostic tests containing 3 problems of each of the 8 types used on the screening test. Twenty-four problems were presented in all. Errors were, as in the previous studies, obtained from written records and no inter-scorer reliability was reported. No a priori error categories were defined.

The results showed that, in general, problems that did not require borrowing posed few difficulties for these children. However, borrowing problems did cause difficulty. Across all types of problems, inversion accounted for 26 percent of the total number of errors committed. Twenty-seven percent of the errors involved 5 different types of incorrect borrowing procedures. Ten percent of the errors were due to failure to recall the correct number fact.

The four studies reported here have been described in some detail because they are among the few error frequency studies to provide more than a bare minimum of procedural information. Although procedures differ between studies, the basic methodology seems to be to present children with different types of subtraction problems and to analyze the written solutions in order to determine the type of error committed. Despite differences in error categories used, the types of problems used, etc., these studies all agree that inversion errors are the most frequently committed of all errors for primary grade children when borrowing is required.

Other investigators have suggested that failure to recall the correct number facts is responsible for most errors in subtraction (Engelhardt, 1977; Morton, 1925; Williams and Whitaker, 1937). These studies are sketchy in their reporting of procedures and results. Conclusions must, therefore, be limited.

Engelhardt (1977) presented grades 3 and 6 pupils with 84 addition, subtraction, multiplication and division problems selected from the Stanford Diagnostic Arithmetic Test and analyzed the solutions for errors. Error frequency information was not presented for the individual operations and inter-scorer reliability was not reported.

Williams and Whitaker (1937) administered the Buswell-John Diagnostic Chart for fundamental processes in arithmetic to 516 fourth to eighth grade children. The subjects were required to speak out loud when solving the problems and the investigator wrote down the types of errors that were committed. No information about how errors were classified was provided. Data were reported as the percentage of pupils having difficulty with the various subskills, although the criteria used to evaluate "having difficulty" were not presented.

Morton (1925) administered 32 verbal arithmetic problems to fifth and sixth grade pupils and divided their solutions into 8 error categories (subjects were required to write down and solve the problems). Rules for assigning responses to error categories were not presented in detail, and no information about the types of problems or the operations involved (addition, subtraction, etc.) was reported. Morton found that "procedure wholly wrong or entirely inadequate" was responsible for more than half of the errors committed, while errors in computations were the second most frequent errors observed.

The most complete evidence suggests that inversion is the most frequent source of errors for children when solving subtraction problems which require borrowing. There is little information available about the persistence of these errors over time. After a one year interval, Cox (1974) retested her subjects who had initially made systematic errors. She concluded that children do commit systematic errors at a later point in time, although a close inspection of her data indicates that only 14 percent of the group she tested (those who initially made systematic errors comprised 13 percent of her original sample) made the same systematic error upon retesting. A considerable amount of learning probably occurred throughout the year, suggested by the finding that 44 percent of those who originally made systematic errors made no error one year later. However, it is interesting that many children did not benefit from classroom instruction over the course of a year and committed the same type of systematic error. It might be more informative to examine the actual number of occurrences of errors in order to obtain a more sensitive index of how persistent subtraction errors are.

None of the studies reviewed above addressed the question of inter-scorer agreement when error judgements were used. Given that it is sometimes difficult to determine which type of error was committed from examining the solution alone, information concerning the ability of two raters to agree on the likely source of an error would be helpful. It should be noted, however, that the presence of adequate inter-scorer reliability does not mean that the observer may be sure exactly what the child did to reach his/her answer. High reliability simply reduces the amount of error variance attributable to observer variables.

After reviewing research on children's arithmetic errors, Burrows (1976) observed that many authors recommend the analysis of written solutions be accompanied by a diagnostic interview. With this procedure, children are questioned about how they solved the problems, and their answers are compared with the written solutions in order to determine the source of the error (Brueckner and Bond, 1955). This technique should yield more detailed and reliable information than the "solutions only" approach. However, the interview technique is extremely time consuming and tiring for the child, requiring extensive questioning on an individual basis. And there is always the possibility that the behavior of the children is different in this situation than under more normal classroom conditions.

The decision about which procedure to use, that is, whether a "solutions only" versus a combined method is in order, must reflect the intent of the investigator. When it is important to obtain descriptive information which is as reliable as possible and when testing time is unlimited, the combined method might be more appropriate. In any case, the investigator should maximize the reliability of the method chosen by using precisely defined and discrete error categories and by making the necessary reliability checks.

Inversion appears to be the most frequent source of errors in childen's subtraction when borrowing is required. It is not clear, however, how persistent these errors are, either at a particular point in time or over a given period of time. Whether children's errors have the consistency which would reflect gaps in knowledge or whether they are more erratic, characterizing problems with processing is, therefore, not clear. Blankenship (1976) and Smith et al. (1973) point out that two types of children exist. These children have different profiles which suggest different causal mechanisms for errors and different methods of remediation. The first group make errors on all problems and require direct instruction in order to eliminate these errors. This suggests that these children have gaps in their knowledge about correct procedures. The second group makes errors on some problems, but also solves some problems correctly. Contingency management is the recommended remedial procedure here, suggesting that faulty processing is the source of errors and errors may be reduced by facilitating attention to correct procedures.

The purpose of Study 1 was to determine which type of subtraction error occurs most frequently on problems that require borrowing for a local sample of children. Children were placed into two groups on the basis of the number of correct solutions, as suggested by Blankenship (1976) and Smith et al. (1973). Group 1 children were those who had made errors on every problem while Group 2 children were those who had made errors but solved some problems correctly. The type and frequency of errors for these groups were examined in order to address the question of differential mechanisms for errors, as described above.

Study 1

Method

Subjects

Fifty-six third grade children (28 male and 28 female) were selected from two elementary schools in lower middle class areas of London, Ontario. Children were selected on the basis of parental consent. Grade three was used because it is the first grade in which children are given instruction in borrowing on subtraction problems as part of the regular mathematics curriculum. The mean age of the subjects was 8.9 years. All children were exposed to and were expected to have learned the subtraction facts to 18, and were experienced in trying to solve problems of the type described below.

Procedure

The subtraction problems consisted of three digit minuends and three digit subtrahends and required borrowing from both the tens' and hundreds' columns, as in the example

$$\begin{array}{r} 5\ 4\ 7 \\ \underline{-2\ 5\ 9} \end{array}\ .$$

Problems were composed of the digits 1 to 9 inclusive, and those used in each problem were selected at random. Two test forms (A and B) were constructed, each containing the same 20 problems arranged in a different random order.

For the initial testing, half of the subjects were randomly assigned form A and the other half form B. Each subject was individually tested by the examiner, and each testing session lasted from 20 to 40 minutes.

Subjects were told that the experimenter wished to learn how children solve subtraction problems. Each child was seated at a table and was presented with one 4x6 inch slip of white paper on which was written one subtraction problem. The child was given a pencil and was asked to solve the problem and to try to get the correct answer. When the child had solved the problem, the paper was turned face down on the table and the child was presented with a second slip of paper, again containing a subtraction problem. Testing continued in this manner until all 20 problems had been administered.

During administration of the test, the examiner was seated beside the subject on the opposite side of his/her writing hand in order to observe the written responses. After the child solved the problem, he/she was asked how each digit of the answer was obtained, moving from the ones' to the tens' to the hundreds' column. All questions were presented in a nonspecific manner, such as "How did you get the answer here?".

The above procedure was repeated with each child following a three to four week interval. This was done primarily to determine whether or not the same errors were repeated after a short time interval. All subjects were tested using an alternate form.

Error Categories

In order to categorize errors, verbal reports were compared with the written protocols for each subtraction problem. If the examiner's judgement of error type based on the written solution confirmed the child's verbal report, the error was recorded as such. When the error could not be determined on the basis of the written answer, the child's description of his/her procedure formed the basis for the error categorization. In cases where a discrepancy existed between the child's report and the examiner's judgement, the latter was used to determine the type of error committed. The error categories used are listed below along with their descriptions.

1) Counting error: the child indicated that counting was used to get the answer and the solution indicated that the answer was computed incorrectly, as in the example,

$$\begin{array}{rccc} & & 3 & \\ & 2 & \not{4}_1 & 5 \\ - & 1 & 2 & 9 \\ \hline & 1 & 1 & 5 \end{array} .$$

2) Number fact error: the child indicated that he/she retrieved the answer from memory with a statement such as "I just knew it", and the written answer indicated that the difference was again computed incorrectly.

3) Inversion error: the child indicated that the smaller number was subtracted from the larger, and the written solution verified this, as in the example,

$$\begin{array}{rccc} & 2 & 4 & 5 \\ - & 1 & 2 & 9 \\ \hline & 1 & 2 & 4 \end{array} .$$

4) Borrowing error: this error was recorded if the child indicated that he/she applied a sequence of steps to the solution of the problem which was not explicitly taught in the classroom. The error occurred somewhere in the borrowing process, for example,

$$\begin{array}{rccc} & & 5 & \\ & 2 & \not{4}_1 & 5 \\ - & 1 & 2 & 9 \\ \hline & 1 & 3 & 6 \end{array} \qquad\qquad \begin{array}{rccc} & & 13 & \\ & 2 & \not{4}_1 & 5 \\ - & 1 & 2 & 9 \\ \hline & 1 & 1 & 16 \end{array} .$$

5) Other: this category included errors which, as described by subjects, or as shown by written work, did not fit any of the above categories, as in the example,

$$\begin{array}{r} 245 \\ -129 \\ \hline 100 \end{array} \qquad \begin{array}{r} 245 \\ -129 \\ \hline 374 \end{array} .$$

In some situations, it appeared that more than one error had been committed on a particular problem. In these instances, all errors committed were recorded. For example, in the problem

$$\begin{array}{r} 245 \\ -129 \\ \hline 125 \end{array} ,$$

the child stated that she subtracted the smaller number from the larger in the ones' column. The answer "5" suggests that the child must also have committed a computational error here. Because she indicated that she counted on her fingers to get the answer, a counting error was also recorded in addition to the inversion error.

Dependent Measure

There were three opportunities for number fact, counting, borrowing and other errors to occur in each problem, one in each of the ones', tens', and hundreds' columns. However, inversion errors could only occur in the ones' and tens' columns. In order to take into account this unequal opportunity for occurrence, the dependent measure for each type of error was the number of errors divided by the number of opportunities for it to occur.

Inter-Observer Agreement

A second observer was trained by the experimenter in the categorization of each type of error. This observer was provided with written descriptions of the errors and was then presented with examples of each type. The observer then practiced making error judgements which were based on written and verbal protocols of subjects (obtained during pilot testing). During the experiment, this observer was seated across from the subject and recorded explanations given by the subject for each answer. Occurrence agreement between observers for each type of error was calculated using the formula:

$$\frac{\text{\# of agreements of error type}}{\text{\# of agreements plus disagreements}} \times 100.$$

This information was obtained for 20 subjects.

Results

Inter-Observer Agreement

Inter-observer agreement for each type of error was as follows: inversion errors, 100%; number fact errors, 100%; counting errors, 97.8%; borrowing errors, 93.8%; other errors, 100%.

Error Frequency

The number of problems solved correctly and the frequency of each error type (expressed as a percentage of the number of opportunities for occurrence) for the 20 subtraction problems are presented in Table 1. Inspection of Table 1 shows that, at initial testing (Time 1), 11 subjects did not solve a single problem correctly. Also, 27 subjects made more than one error but solved at least one problem correctly.

Insert Table 1 about here

The absence of correct responses for some subjects suggests that these children may be different in some way from those who solve problems correctly as indicated in the Introduction. If so, pooling these two subgroups in an analysis of errors might mask potentially useful information. It was decided, therefore, to investigate possible differences between subjects who solved no problems correctly and those who made errors but also solved at least one problem correctly.

The number of errors as a proportion of the total number of possible occurrences for each error category are presented in Table 2. Group 1 refers to those subjects who erred on every problem.

Insert Table 2 about here

Group 2 refers to those who made more than one error but solved at least one problem correctly. For Group 1, inversion errors occurred in over

Table 1: Number of Problems Solved Correctly and the Proportion* of Errors Committed for Individual Subjects

Error	# Correct		Number Fact		Counting		Inversion		Borrow		Other	
Time	1	2	1	2	1	2	1	2	1	2	1	2
Subject												
1	10	0	0	0	12	0	5	100	7	0	0	0
2	17	17	2	0	0	0	3	5	0	0	2	2
3	5	12	3	0	17	10	5	3	0	3	8	2
4	20	18	0	2	0	0	0	0	0	0	0	2
5	10	13	0	0	12	10	0	0	3	0	5	5
6	9	18	5	3	2	0	0	0	13	0	2	0
7	17	19	3	0	0	0	0	0	0	0	0	3
8	2	4	0	5	30	25	20	23	0	0	0	7
9	19	18	0	0	0	0	3	5	0	0	0	0
10	15	19	0	2	10	0	0	0	0	0	0	0
11	0	0	7	5	2	2	100	100	0	2	0	0
12	19	20	0	0	0	0	0	0	0	0	2	0
13	20	19	0	0	0	0	0	0	0	0	0	2
14	20	20	0	0	0	0	0	0	0	0	0	0
15	20	20	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	100	100	0	0	0	0
17	0	0	0	5	5	3	100	100	0	0	0	0
18	19	20	0	0	2	0	0	0	0	0	0	0
19	0	0	3	5	15	15	100	100	0	0	0	0
20	18	18	2	3	0	0	5	0	0	0	2	0
21	0	0	0	0	0	0	100	100	0	0	0	0
22	0	0	0	0	0	0	100	100	0	0	0	0
23	0	0	0	0	0	2	100	98	0	0	0	0
24	0	0	0	0	0	0	100	100	0	0	0	0
25	20	20	0	0	0	0	0	0	0	0	0	0
26	20	20	0	0	0	0	0	0	0	0	0	0
27	15	16	0	0	12	8	0	0	0	0	2	5
28	14	19	0	2	7	0	0	0	0	0	5	0
29	12	16	18	3	2	2	0	0	0	0	0	2
30	12	10	8	3	3	8	0	0	5	10	0	3
31	17	17	2	2	2	2	0	0	0	0	2	2
32	17	19	5	0	0	0	0	0	0	0	3	2
33	17	19	2	0	2	0	0	0	2	2	0	0
34	19	19	0	2	0	0	0	0	0	0	2	0
35	19	19	0	2	0	0	3	0	0	0	0	0
36	20	20	0	0	0	0	0	0	0	0	0	0
37	19	20	0	0	0	0	0	0	0	0	2	0
38	20	13	0	0	0	12	0	0	0	0	0	0
39	10	0	18	0	2	0	0	100	0	0	0	0
40	19	0	0	2	5	0	0	0	0	0	0	35
41	0	0	3	8	8	25	100	100	0	0	0	2
42	14	16	0	0	10	3	0	3	0	0	0	3
43	19	19	0	0	0	0	0	0	2	2	0	0
44	11	4	5	17	12	17	0	0	0	3	5	3
45	12	6	10	5	3	7	0	23	0	0	0	2

Table 1 (continued)

Error	# Correct		Number Fact		Counting		Inversion		Borrow		Other	
Time	1	2	1	2	1	2	1	2	1	2	1	2
Subject												
46	1	0	0	0	45	8	3	50	2	0	13	3
47	4	14	15	7	15	2	3	0	0	0	3	3
48	12	0	2	3	10	2	0	43	0	7	3	0
49	0	0	3	8	0	2	50	50	5	3	2	5
50	7	17	0	0	27	3	5	3	0	0	0	0
51	18	15	0	7	0	0	0	0	0	0	3	2
52	19	0	2	0	0	2	0	50	0	0	0	0
53	20	13	0	3	0	3	0	5	0	2	0	2
54	0	0	5	0	18	13	50	50	0	0	32	35
55	14	0	3	0	10	0	0	50	0	0	0	3
56	15	0	3	0	7	2	0	100	0	0	0	0

*Proportions of errors were calculated using the formula: number of errors/number of opportunities for occurrence. For number fact, counting, borrowing and other errors, the number of errors committed was divided by 60 (3 opportunities for occurrence in each of 20 problems). For inversion errors, the number of errors committed was divided by 40. All proportions were rounded to the nearest whole number.

Table 2: Mean Number of Problems Solved Correctly and Mean Proportion of Errors Committed for Group 1 and Group 2

Time 1

	# Correct	Number Fact	Counting	Inversion	Borrow	Other
Group 1	0.00	1.96	4.50	90.90	.45	3.04
Group 2	12.04	3.95	7.40	2.13	1.17	2.10

Time 2

	# Correct	Number Fact	Counting	Inversion	Borrow	Other
Group 1	0.00	2.85	3.56	90.72	.45	3.79
Group 2	11.41	2.29	4.01	18.50	.93	1.91

90 percent of all possible opportunities. A dependent t test indicated that the proportion of inversion errors was greater than that of all other error types combined, $t(10)=16.05$, $p<.001$. It appears that inversion was by far the most frequent source of error for those subjects who made errors on every problem. Table 1 shows that Group 1 subjects usually made inversion errors to the exclusion of almost all other errors.

For Group 2, more non-inversion errors (counting, number fact, borrowing and other) were committed than inversion errors, $t(26)=7.44$, $p<.001$. Inspection of Table 2 shows that counting was the most frequent source of error. There was no significant difference between the proportion of counting and number fact errors, $t(26)=1.7$, although counting errors occurred more frequently than inversion errors, $t(26)=4.23$, $p<.001$, than borrowing errors, $t(26)=3.69$, $p<.001$, and than other errors, $t(26)=3.49$, $p<.01$.

Error Consistency

Subjects were retested after a period of 3 to 4 weeks in order to examine changes in subtraction performance. During initial testing (Time 1), all Group 1 children made inversion errors on every problem (In Table 1, a 50% frequency of inversion errors indicates that an inversion error was committed in either the ones' or the tens' column of every problem, but not in both columns). When retested following a three to four week interval, all 11 children again erred on every problem. Ten children made inversion errors on every problem while one child made inversion errors on 19 problems and a counting error on the remaining problem. The probability that a Group 1 child will make inversion errors on every problem one month later was .91.

Of the 27 children who were assigned to Group 2, 16 fell into this category at Time 2. Five subjects improved their performance to 19 out of 20 or 20 out of 20 and were no longer considered members of Group 2. Six subjects made errors on all problems, and qualified as members of Group 1. The probability that a child initially categorized as a member of Group 2 will remain so after one month was .59. The retest performance of children in both Group 1 and 2 is shown in Table 2.

Sex

Of the 11 subjects who initially fell into Group 1, 64 percent were male. Of the 27 subjects composing Group 2, 30 percent were male. And, of the 18 who were excluded from the study on the basis of 19 or 20 out of 20 problems solved correctly, 72 percent were male. It appears that more males than females experienced extreme difficulty with subtraction and also demonstrated mastery of subtraction at time 1.

Subjects who solved no problems correctly at Time 1 committed inversion errors on every problem. Almost all of these subjects repeated this pattern a month later. Computing the correct answer-number fact and counting errors - gave Group 2 more trouble than any of the other skills. These children were more likely to change their number of correct responses after one month.

Discussion

The finding that Group 1 children committed more inversion errors than any other type may be consistent with findings of Cox (1974), Graeber and Wallace (1977) and Smith (1968). However, whether or not subjects in these studies were equivalent to the Group 1 subjects examined here is not known. In the absence of specific selection criteria, it is likely that the earlier samples included a mixture of Group 1 and Group 2 children. In that case, it may be that Group 1 subjects were responsible for the high frequency of inversion errors found in these studies.

Group 2 children's difficulty with computations may support findings of Engelhardt (1977), Morton (1925) and Williams and Whitaker (1937) who reported that number fact errors are the main problem in children's subtraction. Engelhardt and Morton did not use interviews in order to obtain error frequency information. In those studies, counting errors could not be distinguished from number fact errors. Williams and Whitaker, although using an interview technique, did not report how errors were categorized. It seems likely that counting and number fact errors were grouped together and reported simply as number fact errors in all of these studies. Whether the subjects used in these studies could be categorized as Group 1 or Group 2 is again unclear, however. To say that the findings of the present study are consistent with these earlier investigations may not be accurate for this reason. Ellis (1972), on the other hand, did use only Group 2 subjects and found that inversion was the most frequent error, a finding not supported in this study.

It appears that the separate analysis of errors for Group 1 and

and Group 2 subjects may help to explain some of the inconsistencies in the error frequency literature. Group 1 clearly demonstrated a different pattern of behavior than Group 2. The errors of Group 1 children were much more consistent both at one point in time and again after a month compared with those of Group 2. This consistency over time is especially disturbing given that subtraction with borrowing had been reviewed in class between initial testing and follow up.

Besides consistency, another feature that differentiated Group 1 from Group 2 was the type of error that was most common. Group 1 children tended to make procedural errors on every problem. That is, they subtracted the smaller number from the larger to get the answer when the appropriate procedure involved borrowing. Group 2 children made more computational errors - number fact and counting - than any other type. They made relatively few errors involving incorrect solution procedures (inversion or borrowing errors).

The consistent pattern of inversion errors observed in Group 1 suggest three possible causes. The first two reflect gaps in their knowledge about borrowing. The children may have made no attempt to borrow because they did not have the necessary skills in their repertoire. Instead, they may have used an available alternative which was to subtract the smaller number from the larger. Alternatively, these children may not have recognized the conditions where borrowing was appropriate. For example, borrowing may have been under the control of teachers' instructions to borrow, or under the control of contextual cues (such as written instructions or the type of lesson presented in class that day) rather than the actual characteristics of the problems. If this is the case, in situations where these extra cues are unavailable, Group 1 children may simply subtract the smaller number from the larger.

Of course, another possible explanation for inversion errors is that Group 1 children were not motivated to apply their borrowing skills. This situation could occur because the contingency structure in the regular classroom usually reinforces problem completion or time on task rather than accuracy. Feedback regarding the accuracy of solutions is rarely immediate, while escape from the unpleasant or boring task of solving problems, as well as teacher attention, often follows problem completion. As a result, children may have applied an easier strategy of subtracting the smaller number from the larger because it helps them finish faster.

This motivation hypothesis may also explain the erratic performance of Group 2 children. Both the stimulus control and absence of borrowing skills explanations used for Group 1 do not explain the errors committed by Group 2 because they do apply the correct procedures some of the time. They do this even though additional cues regarding the requirements of the problems are absent.

Group 1 children may commit errors because correct borrowing occurs only in the presence of stimuli not used in this study. If so, providing these children with instructions to borrow might result in a decrease in the number of inversion errors committed and an increase in the number of problems solved correctly. However, if borrowing skills are not in the child's repertoire, presentation of instructions to borrow might have the effect of changing the type of errors committed. Instead of making inversion errors, these children might attempt to borrow but make errors doing so.

Children (Group 1 and Group 2) may have not been motivated to solve problems correctly because reinforcement was not contingent upon correct solutions. In this case, the provision of contingencies for accuracy might increase the number of problems solved correctly. This latter approach has been examined by several investigators. Marholin and Steinman (1977) compared the effect of reinforcement for on task behavior versus academic rate and accuracy. Eight special class children were given teacher attention and points contingent upon periods of on task behavior and on increases in academic performance. These points could be exchanged for free time. They found that the number of addition, subtraction, multiplication, division and word problems completed and the number of correct solutions were highest when reinforcement was contingent on rate and accuracy. These findings applied to situations where the teacher was in the room as well as outside.

In an earlier study, Ferritor et al. (1972) examined the effect of contingencies for on task behavior (attending), correct solutions and for both. Subjects were third graders who were disruptive in the classroom and who demonstrated varying levels of arithmetic ability. The children worked on addition, subtraction, multiplication and division problems for 20 minutes each day and received tokens for target behavior. These tokens could be exchanged for tangible backup reinforcers such as candy, gum, small toys, etc. Results showed that, as in the Marholin and Steinman study, contingencies for attending had little effect on the accuracy of solutions, while contingencies for accuracy resulted in an increase in the number of problems solved correctly.

In a study by Copeland et al. (1974), the school principal praised three grade 5 pupils for criterion performance in reading and arithmetic. Increases in the number of words read correctly and in the number of addition problems solved correctly were found relative to a condition where the students were given drill in reading and arithmetic. The authors then repeated this procedure with two grade three classes. Students were given a daily problem sheet containing 100 addition problems. Each day they also received feedback concerning the number of problems solved correctly the previous day. The names of those children who improved their scores from the previous day or who were among the 5 students with the highest scores were placed on a list. Twice weekly for three weeks, the principal entered the classroom and asked the listed students to stand. The principal then praised these students for their academic performance. As expected, this procedure resulted in an increase in the number of problems solved correctly by both classes.

Broughton and Lahey (1968) examined the effects of response cost, positive reinforcement, and a combination of the two on the academic performance of fourth and fifth grade remedial mathematics students. In the baseline condition, the teacher walked around the room and monitored each child's work. Check marks were placed under all correctly worked problems and Xs were placed under all incorrect responses. Positive comments were made for correctly worked solutions only. During the reinforcement condition, the teacher behaved exactly as during baseline. In addition, children were given one point for each correct response. These points could be traded later for free time activities. In the response cost condition, the children began with 20 points and lost one point for each problem solved incorrectly. In the mixed condition, subjects began with 20 points and lost points for problems solved incorrectly as well as gained points for correct solutions. Results showed that all three treatment groups solved a greater percentage of problems correctly than the control group. Gains were maintained in response cost and reinforcement conditions during a two week follow up.

Harris and Sherman (1973) found that when contingencies were provided for 90 percent accuracy, the power of an arithmetic tutorial procedure was enhanced for fourth grade children. Lovitt and Esveldt (1970) used a procedure where successively more rapid response rates (number of problems solved correctly per minute) were reinforced by correspondingly greater payoffs. Compared with a fixed ratio reinforcement schedule, this procedure produced an increase in the rate of arithmetic problems solved correctly. Chadwick and Day (1971), in an 11 week study of minority children age 8 through 12, examined the effect of systematic reinforcement for reading and arithmetic accuracy. A condition where social and tangible reinforcers were awarded for reading, spelling and arithmetic accuracy was compared with a condition where reinforcers were awarded noncontingently. Contingent reinforcement resulted in increases in time on task, accuracy on these tasks, and in an increase in the rate of task items completed. Accuracy and rate increases were maintained by teacher mediated social reinforcement alone.

These studies which attempted to improve arithmetic performance through the provision of positive reinforcement utilized fixed ratio or variable ratio reinforcement schedules. In these cases, the number of correct solutions or the rate of correct solutions were tabulated following the completion of an assigned number of problems. Reinforcement was then provided. These procedures were continued over a number of sessions with reinforcement always contingent on a specific number or rate of correct answers. Visual displays provided in some of these reports, however, suggest that subjects' performances improved during the first session, before reinforcement was actually delivered. In this case, enhanced motivation resulting from the promise of reward for correct answers may have been responsible for some of these "reinforcement" effects.

This motivation-reinforcement distinction is an important one. When reinforcement is provided on fixed or variable ratio schedules, desired behavior is shaped and undesired behavior eliminated through joint processes of feedback, increased motivation and extinction. When performance improves following the promise of reward alone, the skills and knowledge required for correct responding are already present. In this case, the promise of reward probably serves to increase attention to the selection and execution of correct procedures.

The purpose of Study 2 was to investigate the effects of instructions to borrow and the promise of reward for correct solutions on the number and types of subtraction errors committed by Group 1 and Group 2 children. A second purpose was to replicate the findings of Study 1.

In Study 2, the large number of subjects required and the limited amount of time these children could be absent from the classroom precluded the use of diagnostic interviews in determining the type of errors committed. Therefore, the error categories used in Study 1 were modified to help keep reliability as high as possible. It is usually impossible to differentiate counting and number fact errors on the basis of written solutions alone. Therefore, these two categories were joined to form a new category called "computational errors". Inversion errors are easily identified on the basis of written solutions alone, so this category was retained. Borrowing errors can also be identified by examining a child's written solutions, although, as is the case with all of these error categories, exclusion of interview information sacrifices some descriptive precision.

Study 2

Method

Subjects

On the basis of parental consent, 134 grade three children were selected from 6 elementary schools in lower middle class areas of London, Ontario. Of this original sample, 80 subjects were selected for further study using the criteria described below. This final sample included 30 males and 50 females.

As in Study 1, all children had been exposed to and were expected to have learned the subtraction facts to 18 as well as how to solve the type of subtraction problems used in this study.

Items

The subtraction problems consisted of 3 digit minuends and subtrahends and were of two types. The first type required borrowing from the tens' column, as in the example,

$$\begin{array}{r} 5\ 4\ 6 \\ -2\ 3\ 9 \\ \hline \end{array} .$$

The second type required borrowing from the hundreds' column, as in the example,

$$\begin{array}{r} 3\ 4\ 8 \\ -1\ 6\ 3 \\ \hline \end{array} .$$

All problems consisted of a random selection of digits from 1 to 9 inclusive, with the restriction that the digits selected meet the specifications of the item types described above.

Two test forms, A and B, were constructed. Each form consisted of a random ordering of the 10 subtraction problems, 5 of each type. Each form contained a different set of problems.

Design

All subjects were randomly assigned to complete either form A or form B. On the basis of pretest performance on these problems, subjects were classified as belonging to either Group 1 or Group 2. Group 1 subjects were those who erred on every problem. Group 2 subjects were those who made more than one error but who solved at least one problem correctly. Following the pretest, subjects were randomly assigned to either a Reinforcement, Instructions, or Control condition. Following the experimental manipulation, subjects were administered an alternate form (either A or B) of the subtraction test as in the pretest. This constituted the posttest. The design, then, was a 3 (experimental condition) by 2 (group) factorial, where subjects were nested in experimental condition and in Group.

Procedure

Pretest

Subjects were selected from their classrooms by their teacher and sent to an experimental room in groups of 4 to 6. When they were seated, the experimenter presented them with the pretest, along with instruction to solve the problems. Following the pretest, subjects were presented with the alternate test form and the instructions for the selected experimental condition.

Motivation Control

Subjects in the Motivation Control condition were shown a table containing various small toys, candies, pens, pads, toy jewellery, etc. Beside the toys was a chart listing the number of correct solutions out of 10 required to earn a particular toy. The examiner told the children that these toys were to help them to do their best to answer all the questions correctly. The examiner then read each item on the reinforcement menu and pointed out the number of correct solutions required in order to obtain each toy. Following this, the children were seated and presented with a test form and reminded that they should try to solve as many problems correctly as they could in order to earn a prize. In order to avoid the disappointment which would result if some children earned prizes of their choice and some did not, subjects were not given their actual scores following the posttest. Instead, they were told that they did quite well and were each allowed to pick a toy of their choice. They were also asked not to tell their classmates about their experiences during the testing session. Subsequent questioning of subjects indicated that the children were indeed naive regarding the procedures used in this study.

Instructions

The children in this condition were given the following instructions: I am now going to give you more problems. They are borrowing (regrouping) problems, and you will have to borrow (regroup) in order to get the right answer. But each problem is different, and you will have to decide whether or not you should borrow (regroup) in the ones' or in the tens' place. I want you to look at each problem very carefully and to decide what you should do to get the right answer. Look at each one and decide whether you need to borrow (regroup) in the ones' or in the tens' place. Do you understand? (E answers questions if necessary). Here are the problems.

At the completion of this session, each child was allowed to pick a toy of his/her choice.

Control

Subjects in the Control condition were given the second set of problems under the same conditions as in the pretest. At the end of the session they were also provided with a toy of their choice.

Error Analysis

Errors committed by each child were assigned to one of four categories based on the child's written performance.

1) Inversion error: this error was recorded if the answer indicated that the child subtracted the smaller number from the larger when the smaller number was located in the minuend, as in the example,

$$\begin{array}{r} 6\;2\;5 \\ -3\;4\;3 \\ \hline 3\;2\;2 \end{array} .$$

Sometimes borrowing procedures were applied correctly but the answer appeared to have been obtained by ignoring some or all of the borrowing steps and subtracting the smaller number from the larger, as in the examples,

$$\begin{array}{l} 5 \\ \not{6}_1 2\;5 \\ -3\;4\;3 \\ \hline 3\;2\;2 \end{array} \qquad \begin{array}{l} 5 \\ \not{6}_1 2\;5 \\ -3\;4\;3 \\ \hline 2\;2\;2 \end{array} .$$

In cases such as these, the error was assigned to the inversion category.

2) Computational error: this error was recorded if the child's written work indicated that he/she performed all of the borrowing steps correctly but still failed to produce the correct answer, as in the example,

$$\begin{array}{r} 5 \\ \not{6}{}_1 2\;5 \\ -3\;4\;3 \\ \hline 2\;7\;2 \end{array} .$$

Computational errors were usually characterized by answers that were only slightly too large or too small. For a few subjects who demonstrated no borrowing throughout the problem sheet (suggesting that the child borrowed "in his/her head") and who did not commit inversion errors, errors were assigned to the computational category, as in the example

$$\begin{array}{r} 6\;2\;5 \\ -3\;4\;3 \\ \hline 2\;9\;2 \end{array} .$$

For a few subjects, inversion errors had been committed, no use of borrowing was observed on the problem sheet, but some answers did not immediately appear to fall into the inversion category. In these cases, it was assumed that the child committed an inversion error and also performed the required computations incorrectly. The errors were assigned to both inversion and computational categories. The following is an example:

$$\begin{array}{r} 6\;2\;5 \\ -3\;4\;3 \\ \hline 3\;1\;2 \end{array} .$$

This pattern occurred rarely in the sample of children examined here.

3) Borrowing error: this error was recorded if the child's written solution demonstrated an incorrect application of borrowing procedures, as in the example,

$$\begin{array}{r} 5\;11 \\ 6\;\not{2}{}_1 5 \\ -3\;4\;3 \\ \hline 2\;712 \end{array} .$$

If the computations appeared to have been performed incorrectly, one error was assigned to the borrowing category and one error was assigned to the computational category, as in the example,

$$\begin{array}{r} 5\ 11 \\ 6\ \not{2}_{1}5 \\ -3\ \ 4\ \ 4 \\ \hline 2\ \ 812 \end{array} .$$

If the child made an error in applying one or more of the borrowing procedures and if the answer fit the inversion category only, the error was categorized as an inversion error, as in the example,

$$\begin{array}{r} 5\ 11 \\ 6\ \not{2}_{1}5 \\ -3\ \ 4\ \ 3 \\ \hline 2\ \ 2\ \ 2 \end{array} .$$

In this case, it appeared that the child subtracted correctly in the ones' place, although he/she incorrectly added a digit to the 5, making it 15. He/she appeared to have inverted in the tens' column, ignoring the fact that the 2 was changed to 11. And the child subtracted correctly in the hundreds' place, taking into account that a hundred had been borrowed from the 6.

4) Other: this error denoted errors which could not be placed in any of the above categories, as in the examples,

$$\begin{array}{r} 6\ 2\ 5 \\ -3\ 4\ 3 \\ \hline 3\ 0\ 2 \end{array} \qquad \begin{array}{r} 6\ 4\ 7 \\ -2\ 7\ 4 \\ \hline 4\ 2\ 1 \end{array} \qquad \begin{array}{r} 3\ 4\ 3 \\ -1\ 8\ 2 \\ \hline 1\ 8\ 2 \end{array} .$$

The above method of error analysis was intended as a compromise between the laborious but accurate observation/interview technique used in Study 1, and the expedient but oversimplified right/wrong approach used in some classrooms. It was felt that this type of analysis would yield data pertinent to the hypotheses of interest in this study while simplifying the procedure to allow group testing with minimal disruption in classroom routine.

Inter-Observer Agreement

A second scorer was provided with a written description of the error categories and was given practice with feedback in categorizing errors, as in Study 1. One test form for each of 20 children was selected randomly and scored by this second scorer. Inter-observer agreement was computed using the formula: (# of agreements of error type/# of agreements plus disagreements) X 100.

Analysis

In this experiment, a pre-post design was used where subjects were nested in experimental conditions and in groups. Because the number of problems solved correctly was equated prior to random assignment to treatment groups, analysis of variance using posttest scores as the dependent measure was used to examine treatment effects for this measure. Analysis of covariance, with pretest scores as the covariate, was used to examine the effects of treatments on the frequency of errors in situations where parametric statistics were considered appropriate. The rationale for this selection is presented in Appendix 1. For analysis of variance and analysis of covariance, comparisons between adjusted cell means were made using the Bonferroni t test. The effective error used in comparisons of adjusted cell means (ANCOVA) was calculated using the formula:

$$\frac{2MS\ error\ (adj)}{n}\left[\frac{1 + (Txx/k-1)}{Exx}\right], \text{ where}$$

Txx = treatment sum of squares for the covariate,

k = the number of treatments, and

Exx = sum of squares error for the covariate (Winer, 1971).

Where large differences existed in cell variances, homogeneity of variance was examined using Hartley's Fmax statistic (Winer, 1971). In cases where the Fmax value exceeded the 95 percent level of confidence, the Kruskal-Wallace H test was used. This test is an analogue of the one-way analysis of variance and is used when more than two groups are to be compared (Horowitz, 1974). H is distributed as X^2 with k-1 degrees of freedom, where k is the number of treatment conditions.

Results

Inter-Observer Agreement

Inter-observer agreement for each of the error categories was as follows: inversion errors, 100 percent; computational errors, 89 percent; borrowing errors, 95 percent; other errors, 96 percent.

Frequency of Error Type

Thirty-one subjects were assigned to Group 1 on the basis of pretest performance (14 males and 17 females). Forty-nine subjects were assigned to Group 2 (16 males and 23 females). Fifty-four subjects who solved 9/10 or 10/10 problems correctly were eliminated from the analyses.

In order to simplify interpretation, each type of error was counted only once per problem. That is, the dependent measure was the number of problems where each type of error occurred. In order to examine the frequency of errors uncontaminated by the experimental manipulations,

Insert Table 3 about here

the comparisons reported below were made using pretest data only. The distribution of errors is presented in Table 3. As Table 3 shows, Group 1 subjects committed a greater proportion of inversion errors relative to total errors than Group 2 subjects, $t(78)=6.32$, $p<.003$ (p. values were controlled for alpha slippage using the formula: $\alpha' = \alpha$ for single comparison/# of comparisons, where α refers to the desired level of significance for each comparison). Group 2 subjects committed a greater proportion of computational errors relative to total errors than Group 1 subjects, $t(78)=4.83$, $p<.003$. The proportion of borrowing errors committed did not differ between groups, $t(78)=1.97$, $p<.15$.

Table 3: Mean Number of Subtraction Errors per Subject and Mean Proportion* of Subtraction Errors Committed by Group 1 and Group 2

	Inversion	Computational	Borrowing	Other
	Mean Number of Errors per Subject			
Group 1	7.32	1.19	1.90	.55
Group 2	1.22	2.16	1.31	.14
	Mean Proportion of Errors per Subject			
Group 1	.71	.11	.13	.05
Group 2	.19	.48	.28	.03

* Proportions were calculated for each subject using the following formula: # of errors of each type/# of total errors

Although the error classification scheme differs from that employed in Study 1, these results basically replicate those of Study 1. Group 1 subjects committed many more inversion errors than any other type of error, while Group 2 subjects committed more computational (counting and number fact) errors.

Number Correct

The number of problems solved correctly on the posttest was examined using a one way analysis of variance with experimental condition as the between subjects factor. The analysis was repeated for Group 1 and Group 2. The results indicated that the treatments had no significant effects on the number of problems solved correctly for Group 1 or Group 2. Planned contrasts between cell means obtained for Group 1 Instructions versus controls, and between Motivation Control and controls failed to yield any significant differences. This suggests that the Instructions and Motivation Control conditions as implemented here are not powerful enough to improve performance on this variable. These results are presented in Table 4.

Insert Table 4 about here

Inversion Errors

Group 1 and Group 2 were not equal in the number of errors committed across experimental conditions. Therefore, for Group 1, the effect of these conditions on the number of problems containing inversion errors was examined using a one way analysis of covariance with experimental condition as the between subjects factor and with the number of inversion errors committed on the pretest as the covariate.

For Group 2, the presence of zero variance suggests that the scores

Table 4: Mean Number of Problems Solved Correctly per Subject for Group 1 and Group 2

Group 1

		Pretest	Posttest
Motivation Control	mean	0.00	.80
	sd	0.00	2.20
Instructions	mean	0.00	2.45
	sd	0.00	3.78
Control	mean	0.00	1.00
	sd	0.00	2.54

Group 2

		Pretest	Posttest
Motivation Control	mean	4.72	6.39
	sd	2.99	3.90
Instructions	mean	5.47	6.60
	sd	2.45	3.91
Control	mean	6.31	7.38
	sd	2.72	2.87

for this dependent measure were not normally distributed. The Kruskal-Wallace H test was therefore performed on these data. The dependent measure was the number of errors committed on the pretest minus the number of errors committed on the posttest.

For Group 1, a significant main effect was found for treatment, $F(2, 27)=4.18$, $p=.0262$. Bonferroni t comparisons showed that Group 1 subjects in the Instructions condition committed fewer inversion errors than controls, $t(27)=2.5$, $p<.025$, one tailed. Those in the Instructions condition also committed fewer inversion errors than those in the Motivation Control condition, $t(27)=2.45$, $p<.025$, one tailed. Kruskal-Wallace H for Group 2 yielded no significant effect. These means and standard deviations are presented in Table 5.

Insert Table 5 about here

Borrowing Errors

For Group 1, the same one way ANCOVA with the number of problems containing borrowing errors as the dependent variable yielded a significant main effect of experimental condition, $F(2, 27)=3.33$, $p=.05$. Bonferroni t comparisons between adjusted cell means indicated that Group 1 subjects made more borrowing errors in the Instructions condition than in the Control condition, $t(27)=2.5$, $p<.025$, one tailed. Adjusted means did not differ significantly between Instructions and Motivation Control conditions, $t(27)=1.76$. ANCOVA yielded no significant effect for Group 2. Treatment means are presented in Table 6.

Insert Table 6 about here

Table 5: Mean Number of Problems Containing Inversion Errors per Subject for Group 1 and Group 2

Group 1

		Pretest	Posttest
Motivation Control	mean	7.90	7.20
	sd	3.96	4.40
Instructions	mean	7.36	3.64
	sd	4.50	4.37
Control	mean	6.70	6.30
	sd	4.30	4.64

Group 2

		Pretest	Posttest
Motivation Control	mean	1.83	.72
	sd	3.17	2.14
Instructions	mean	.80	.33
	sd	1.70	1.29
Control	mean	.94	0.00
	sd	2.32	0.00

Table 6: Mean Number of Problems Containing Borrowing Errors per Subject for Group 1 and Group 2

Group 1

		Pretest	Posttest
Motivation Control	mean	2.50	2.60
	sd	4.10	4.20
Instructions	mean	2.00	4.91
	sd	4.00	4.53
Control	mean	1.20	.80
	sd	3.16	2.53

Group 2

		Pretest	Posttest
Motivation Control	mean	1.39	1.06
	sd	2.43	2.46
Instructions	mean	1.80	2.20
	sd	2.60	3.55
Control	mean	.75	.56
	sd	.93	1.75

Computational Errors

Comparison of cell variances using Hartley's Fmax test for Group 1 yielded significant differences, Fmax(10)=17.19, p<.01. A Kruskal-Wallace H test was performed on these data. Analysis of covariance was performed on the data for Group 2. No significant main effects were found in either analysis. Means and standard deviations are presented in Table 7.

Other Errors

For Group 1, zero variance in the Motivation Control Condition suggested the scores for this error were not normally distributed. For Group 2, comparison of cell variances yielded significant differences, Fmax (18)=132.21, p<.01. Kruskal-Wallace H tests were performed for Group 1 and Group 2 and yielded no significant differences. Means and standard deviations are presented in Table 8.

Insert Tables 7 and 8 about here

Sex

Of the total sample of 80 children examined in Study 2, 38 percent was male. In Group 1, 45 percent was male, while 33 percent of Group 2 was male. It appears that the sex distribution within Group 1 and Group 2 reflects the characteristics of the sample.

Table 7: Mean Number of Problems Containing Computational Errors per Subject for Group 1 and Group 2

Group 1

		Pretest	Posttest
Motivation Control	mean	.70	.60
	sd	.95	.84
Instructions	mean	1.45	1.18
	sd	2.90	3.00
Control	mean	1.40	2.00
	sd	2.80	3.50

Group 2

		Pretest	Posttest
Motivation Control	mean	2.39	2.17
	sd	2.62	2.68
Instructions	mean	2.33	1.73
	sd	2.35	2.52
Control	mean	1.40	1.50
	sd	1.77	1.59

Table 8: Mean Number of Problems Containing Other Errors per Subject for Group 1 and Group 2

Group 1

		Pretest	Posttest
Motivation Control	mean	.10	0.00
	sd	.32	0.00
Instructions	mean	.09	1.18
	sd	.30	2.27
Control	mean	1.50	1.00
	sd	3.14	2.49

Group 2

		Pretest	Posttest
Motivation Control	mean	.17	.60
	sd	.38	.24
Instructions	mean	.20	.07
	sd	.56	.26
Control	mean	.06	1.00
	sd	.25	2.70

Discussion

Study 1 showed that children who committed errors on all subtraction problems did so because they failed to apply borrowing procedures or applied them incorrectly. Those children who utilized correct procedures at least some of the time committed errors that were primarily computational in nature. That is, they made number fact or counting errors.

Study 2 showed that two simple manipulations, one which altered the stimulus conditions and one which offered rewards for correct solutions did not produce a significant increase in the number of problems solved correctly by either Group 1 or Group 2. Instructions to borrow, rather than improving the performance of Group 1, produced a shift in the most frequent type of error committed. In the absence of special instruction, Group 1 children did not attempt to borrow. When given instructions to do so, they attempted to borrow but their behavior suggested that they did not have the necessary skills to solve the problems correctly. That is, they committed other procedural errors.

Group 1 and Group 2 behaved in a similar fashion in some ways. First, no significant changes in the number of errors committed were observed as a result of the offer of reinforcement or as a result of instructions to borrow. From the standpoint of remediation, this is an important finding, because it suggests that errors for both groups are not easily eliminated, at least not by the simple techniques employed here. Second, promise of reward did not significantly alter the types of errors committed by either group, contrary to results of studies cited earlier. Again, this suggests that simply attempting to increase a child's level of motivation is insufficient to alter subtraction performance. Third, instructions to borrow did not alter the frequency of computational or other types of errors.

Group 1 and Group 2 differed in one respect. Group 1 responded to instructions to borrow with an increase in borrowing errors and a reduction of inversion errors. Group 2 did not show any significant changes in the frequency of these errors. This finding may be further understood by examining the data for individual subjects in Group 2 and in the Instructions condition. These data are reported in Table 9.

Insert Table 9 about here

In Group 2, three subjects made inversion errors on the pretest. All 3 responded to instructions with a reduction in inversion errors, these errors being completely eliminated in 2 cases. For two subjects, this reduction in inversion errors was accompanied by an increase in the number of borrowing errors. In short, this is essentially the same pattern found for subjects in Group 1.

From Table 9, it is also apparent that quite a few subjects made borrowing errors on the pretest. If the presence of either inversion or borrowing errors reflects difficulty with the procedures required for borrowing, an examination of the performance of Group 2 shows that 35 out of 49, or 71 percent of these subjects made procedural errors. It seems that most children in Group 2 also had some difficulty with borrowing.

Further, many Group 2 children in Study 1 also demonstrated that they had difficulty with borrowing. Forty-four percent of these children made procedural errors. In addition, 8 children who had initially solved some problems correctly (two subjects had solved 19 out of 20 problems correctly) erred on every problem when retested one month

Table 9: Mean Proportions of Inversion and Borrowing Errors Relative to Total Errors per Subject in Group 2 Instructions Condition

	Pretest		Posttest	
Subject	Inversion	Borrowing	Inversion	Borrowing
1	.44	.11	0.00	.54
2	.60	0.00	.38	.54
3	0.00	.50	0.00	0.00
4	0.00	0.00	0.00	.64
5	0.00	.23	0.00	.10
6	0.00	.33	0.00	0.00
7	0.00	1.00	0.00	0.00
8	0.00	0.00	0.00	0.00
9	0.00	.86	0.00	0.00
10	0.00	1.00	0.00	1.00
11	0.00	0.00	0.00	0.00
12	0.00	.33	0.00	1.00
13	.83	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00
15	0.00	.38	0.00	0.00

later. Seven out of these 8 subjects made procedural errors on every problem.

Group 1 children were those who had the greatest difficulty with subtraction in the sense that they never solved borrowing problems correctly. Inversion errors appeared to characterize this group. That is, they did not attempt to borrow unless they were specfically directed to do so. They also did not appear to have the borrowing skills available when asked to apply them. If borrowing is viewed as a set of skills, one of which involves recognizing when it is necessary to borrow, Group 1 children appear to have mastered few of these. But the high percentage of children who made procedural errors in Group 2 suggests that children in this group had not mastered all of the borrowing skills either (assuming that mastery is defined as the correct execution of all procedures on the majority of problems presented at various points in time). Their apparent mastery was transitory, disappearing after one month, or their execution of borrowing procedures was erratic from problem to problem during one sitting.

These findings suggest that both Group 1 and Group 2 have varying degrees of difficulty with the borrowing process and often behave the same way when rewards are promised or instructions to borrow are presented. In fact, it appears that these children fall at different locations on a mastery continuum, but are not necessarily qualitatively different.

In the third study reported in this paper, it was determined to be more useful to divide children on the basis of a commonly accepted standard of satisfactory academic performance, the 60 percent criterion. This was because there did not appear to be any strong empirical support for maintaining the Group 1, Group 2 dichotomy. As children who achieve grades lower than 60 percent are of concern to educators, it was decided to focus attention on this group.

Instructional Programs in Arithmetic

The results of Study 2 suggest that simply trying to increase children's level of motiviation, or making them aware of what procedures are required are inadequate remedial techniques for children who experience difficulty with subtraction. Since difficulties appear to reflect failure to master one or more of the skills required for borrowing, or failure to use them consistently, a more intrusive approach to remediation seems necessary. A review of the remedial arithmetic literature suggests that instruction in specific skills and feedback are two viable remedial techniques.

Reports of remedial approaches to subtraction involving instruction in component skills, however, often fail to provide details of the procedures advocated or tested (Harvey and Kyte, 1965; Post and Reys, 1979; Brueckner and Bond, 1955; Fuson, 1979). In cases where the remedial programs were evaluated empirically, one must speculate about their relative effectiveness. In addition, the programs are generally outlined in a haphazard or non-specific manner. For example, Mick and Brazier (1979) noted the importance of varying the redundant stimuli while maintaining the stability of critical stimulus features when programming for generalization. They suggested that stimulus attributes should be varied and more and more instances experienced until the new boundary (of a concept) is clear. Unfortunately, they did not explain in detail how to do this. Post and Reys (1979) wrote that instruction which is based on a variety of experiences encourages the abstraction of essential ideas that are common to several activities. For students experiencing difficulty, "encouragement" using different experiences may not be enough to ensure recognition of the essential components of critical stimulus features.

When programming is carried out in a nonsystematic manner, such as attempting to produce generalization simply by providing many examples, teacher control over the child's behavior is not ensured. In fact, many authors (Brueckner and Bond, 1955; Usiskin, 1974; Post and Reys, 1979; Mick and Brazier, 1979; Mierkiewicz, 1979) advocate the inclusion of various components of a teaching program on the assumption that certain behaviors, conceptual 'leaps' or understanding will result on the part of the child. No direct attempt is usually made to ensure control of specific behaviors through attention to stimulus features, reinforcement contingencies or both. As an illustration, Usiskin (1974) advocated that a teacher should encourage different ways of visualizing an operation, primarily through the incorporation of concrete objects into teaching. This will, he claimed, show the child the important relationship between arithmetic and the real world. These guidelines, however, are too general in that they do not tell the teacher exactly what to do. Also, there is little evidence to suggest that children automatically make the connection between the examples used and the 'real world'.

Fuson (1979) stressed the importance of making sure the desired behavior is the behavior actually reinforced. However, she suggested that a child be allowed to use the procedure of his/her choice when solving a problem. Following this, the teacher should discuss the advantages of the alternate (correct) procedure with the child. There are two problems here. First, Fuson assumed that the child will realize that the alternate procedure is superior and adopt it. However, the reinforcement likely received through the completion of the problem by the child's own method could serve to strengthen the undesired behavior when no specific method to control the child's behavior is used. Second, as in the previous example, very little in the way of concrete suggestions were provided for the teacher.

Consequences of failure to maintain control over the target behavior are illustrated in studies done by Carpenter (1980) and Parsons (1973). Carpenter's subjects were trained to write open number sentences (addition and subtraction) when solving verbal problems. In spite of repeated directions to write the number sentence before solving each problem, one quarter of the subjects solved the problems before they wrote the number sentences. Often, once the subjects had written the number sentence, they ignored it and used the verbal problem to decide on a solution strategy. Parsons found that simply having children circle the operation sign before solving an addition or subtraction problem did not ensure that the child performed the correct operation. Having the child respond differentially to the signs, however (by verbalizing the operation), did ensure that the correct operation was performed.

Remedial procedures which assume children will learn certain things when target behaviors are not specifically programmed ultimately lead to explanations of failure that invoke deficits in the child. Mick and Brazier (1979) stated that the failure of a child to make a final 'leap' to a larger class of concepts (to generalize) may mean the child had not reached the Piagetian stage of formal operations. Such "explanations" of failure often result in a "wait until the child is ready" approach to remediation, curbing the pursuit of academic gains through more effective remedial programming.

Harvey and Kyte (1965), Logan (1976), Parsons (1973) and Harris and Sherman (1973) all reported improvements in mathematics performance as a result of remedial procedures involving instruction. Unfortunately, Harvey and Kyte did not include a control group or a description of the procedures used in their study, so conclusions from that investigation are tentative. Logan (1976) provided modelling in the form of vignettes where the subject read about a child who solved problems correctly, incorrectly or both. This procedure resulted in reduced errors for subjects who initially made inversion errors on all problems attempted, and produced a moderate degree of maintenance and generalization to other problem types. No controls were employed. Parsons (1973), although not examining the issue of generalization, appeared to have obtained good short term results using a carefully programmed behavioral procedure to teach developmentally handicapped students to solve story problems.

Harris and Sherman (1973) conducted a study to evaluate the effects of peer tutoring with and without consequences for accuracy on elementary mathematics performance. Fourth and fifth grade students who scored one year below grade level in math on the Iowa Test of Basic Skills were selected as subjects. Performance on problems selected from the students' math workbooks was evaluated under three conditons. In the first condition, immediate feedback consisting of the number of problems solved correctly was provided. No instruction was given. In the second condition, subjects were arranged in groups of 2 to 3 and asked to simply help each other solve the problems. Helping behavior was praised by the teacher. Later, early recess was used as a reinforcer for 90 percent accuracy in this condition. The third condition was similar to the second, except that subjects were tutored on related but not identical problems.

Results showed that the tutorial procedure resulted in an 11 to 16 percent increase in accuracy and in a 50 percent increase in the rate of performance relative to feedback alone. When the effects of tutoring were compared with those of independent study (keeping time engaged in each activity constant), the tutorial procedure was still superior. Tutoring on related problems also produced increases in accuracy, but these increases took longer to take effect than when the identical problems were used. It is important to note that the tutorial procedure used in this study was unstructured. Children were told only to help each other, and the nature of the help may have differed from child to child. Larger gains in performance might have been obtained if a more structured peer tutoring procedure had been employed. In this study again, it was not clear what tutoring actually consisted of and what elements could be responsible for changes in performance.

Smith, Lovitt and Kidder (1973) suggested that reinforcement is an appropriate remedial technique for children who solve arithmetic problems correctly some of the time, while teaching is a more effective remedial procedure for those who never respond correctly. They performed two studies. The first used withdrawal of positive reinforcement with an eleven year old girl who demonstrated erratic performance in subtraction (equivalent to a Group 2 subject). The design was a multiple baseline across three different types of subtraction problems, where the subject lost one minute of recess time for each incorrect answer. No instruction or feedback was provided. Accuracy improved and was maintained when the contingency was removed for a particular problem type, but generalizaion to other problem types (some of which the girl had previously solved correctly) did not occur.

In the second study, the experimenter used instructions, paper clips, an abacus, and Cuisinaire Rods in a withdrawal design with a ten year old boy. Remediation was aimed at helping the child solve subtraction problems of the type 15-6=__. Although teaching methods were not well defined, the authors reported no attempt to fade out the instructional aids used; they were simply withdrawn. It is not surprising, then, that when the teaching procedure was discontinued, performance on that particular problem deteriorated.

In summary, it appears that performance in different areas of elementary mathematics can be improved through the use of instructional techniques. The optimal types of instruction, however, are unclear. Indications are that the crucial elements include systematic programming and good control over the behavior of the student. Such instructional material, developed through task analysis and including the shaping of successive approximations, practice, and advancement following achievement of mastery criteria have been advocated by Skinner (1968), Holland (1965) and Gagné (Gagné and Brown, 1961).

Performance Feedback

A number of studies have examined the effects of performance feedback on math performance. Conlon, Hall and Hanley (1973) used a peer to provide performance feedback to 2 students who performed erratically on arithmetic problems. This procedure resulted in an increase in the number of problems solved correctly relative to baseline. Sagotsky, Patterson and Lepper (1978) conducted a self monitoring versus goal attainment study with fifth and sixth graders. These subjects were unselected relative to any specific performance criteria. In the self monitoring condition, subjects were required to periodically note whether or not they were engaging in on task behavior, that is, working on math problems. In the goal setting condition, subjects were required to set daily performance goals and to record the number of problems completed. Self monitoring alone resulted in improvement in the rate of mathematics problems solved correctly. Short term maintenance was also found when this condition was terminated.

Kirby and Shields (1972) investigated the effects of immediate praise and immediate correctness feedback on the number of problems solved correctly per minute by a thirteen year old grade seven student. In the baseline phase of an ABAB design, a daily worksheet containing 20 multiplication problems was presented. The subject's solutions were corrected and the papers were returned. In the treatment phase, after all 20 problems were completed, reinforcement in the form of verbal praise was provided by the teacher for every 2 problems solved correctly. The number of correct solutions required for praise was gradually increased in this condition. Results indicated an increase in the number of problems solved correctly per minute during treatment relative to baseline. The amount of on task behavior also increased. These effects carried over into the reversal phase of the study.

Fink and Carnine (1975) performed a study with first graders. In the first condition, subjects were provided with feedback in the form of the number of problems solved correctly. This number was written at the top of a daily arithmetic worksheet. In the second condition, subjects received feedback as above, and were required to graph their daily performance on charts located on their desks. The authors reported that feedback plus graphing resulted in an increase in the number of problems solved correctly relative to baseline. However, the report does not clarify whether feedback alone constituted the baseline condition, so it is difficult to evaluate the effect of feedback alone.

Baxter (1973) examined the effect of feedback given for class and homework assignments on the performance of sixth graders. The study was conducted over an 8 week period. One half of the subjects received feedback in the first four weeks. Each subject in both conditions received one classwork and one homework assignment per week. At the beginning of each week, subjects in the feedback condition were given feedback at the top of their assignment page based on last week's performance. The feedback consisted of the number of each type of error committed. Assignments contained four addition, subtraction, multiplication and division problems, 16 problems in all.

Errors committed were categorized by one investigator and were based on the child's written solutions. No inter-scorer reliability was reported. Errors were classified as: 1) number fact; 2) renaming (borrowing); 3) algorithm; 4) blunders (careless). Descriptions of each error type were provided.

Due to pre-treatment inequality in both the number of problems solved correctly and the number of certain types of errors committed between treatment groups, the ANOVA performed by the authors was not appropriate. Inspection of the data presented in graphic form, therefore, affords the best analysis. There was some indication that feedback resulted in an increase in number fact errors. No other effects appeared related to treatments. These results should be interpreted cautiously given the small number of errors committed by these subjects.

Blankenship (1976) examined the effects of demonstration and feedback on the subtraction performance of children who committed a high frequency of inversion errors. Nine learning disabled children from grades three through five were selected. These children committed inversion errors on 90 percent of the subtraction problems presented on a screening test. An ABA treatment design followed. In baseline, students were presented with one problem sheet per day for six days. The problem sheets contained nine different types of subtraction problems. These problems ranged from those with a two digit minuend and one digit subtrahend with borrowing, to those with three digits in both minuend and subtrahend and with borrowing in two columns. There were nine rows of problems with five problems in each row. Each row consisted of one type of problem.

During the treatment phase, students were again given one problem sheet a day for six days. Training was given for problems contained in row one. Before the day's assignment, students received a written demonstration accompanied by a verbal explanation of how to solve a problem of the type $\begin{array}{r}37\\ \underline{-9}\end{array}$. They were then asked to solve a problem and were given feedback and additional instruction until the problem was solved correctly. Instruction on the other problem types was not provided. During the final or reversal phase, students were given a test sheet 15 days and 30 days following treatment.

Results showed that in baseline, subjects solved no problems correctly and inverted on 91.7 percent of all opportunities. The percentage of problems solved correctly increased to 95 percent on treated rows following instruction, and this percentage gradually declined through rows 2 to 9 (10 percent correct in row 9) as problems became less similar to training problems. Improvements were noted on all problem types relative to baseline, but no statistical analyses were done. Gains in the number of problems solved correctly were accompanied by a reduction of inversion errors (to 5 percent on treated problems). A slight drop in accuracy and slight increase in inversion errors were noted during reversal. Three students demonstrated little or no generalization to non-training problems.

In some studies where feedback consisted of [illegible] number of problems solved correctly, additional reinforcement was also available to the subjects. Other studies, which did not provide reinforcement along with feedback suffered from methodological or reporting problems. However, it appears that feedback has promise as a way of improving children's performance in arithmetic, especially when additional reinforcement is also included.

The purpose of Study 3 was to directly compare the effectiveness of two approaches to the remediation of subtraction difficulties. The first remedial technique was a step-by-step presentation of the procedures necessary for solving three digit subtraction problems requiring borrowing. This program attempted to teach directly the component skills required to solve these problems. Some skills were extracted from components of computer programs designed to simulate subtraction performance (Young and O'Shea, 1981; Brown and Burton, 1978). Others were obtained through task analysis by the author.

The second remedial technique presented children with subtraction problems followed by performance feedback in the form of correctly worked solutions. All the information necessary to solve the problems correctly was included in this feedback. This procedure, if found to be effective, would represent an improvement over the component skills training approach in terms of ease of preparation, presentation and perhaps remediation time.

As stated earlier, a major aim of this research was to examine the feasibility of remedial procedures which could be used by teachers within the regular classroom. To be useful in this way, these techniques would have to work quickly, be easy to administer, and be non-disruptive to other students in the class. For these reasons, as well as to facilitate consistency of presentation, both programs were presented in a programmed instruction format.

Most of the studies of remedial arithmetic instruction cited above used small samples of children. While having advantages, small samples do not provide much information about the number of children who benefit from a particular remedial approach and whether or not some children have difficulty. In addition to the evaluation of treatment effects, it was decided, therefore, to examine the progress of children as they worked through the two training programs. Lastly, the effects of these two programs were assessed for subjects who demonstrated substandard academic performance (less than 60 percent correct on a pretest) as well as for those whose academic performance was satisfactory (above 60 percent).

Study 3

Method

Subjects

Subjects were from third and fourth grade classrooms in seven schools located in lower middle class areas of London, Ontario. One hundred and seventy-seven subjects were selected on the basis of parental consent. These children were given a pretest, which is described in detail below. Only those children who solved less than 80 percent of the problems correctly on the pretest were included in the study. This selection procedure resulted in a sample of 67 subjects who were eligible for inclusion. Three subjects were excluded due to misclassifications. Of the final 64 subjects, 24 were male and 40 were female. The average age was 9 years and 3 months. All children had been exposed to and were expected (by their teachers) to have learned the subtraction facts to 18 as well as how to solve the types of subtraction problems used in this study.

Items and Test Forms

The training and test items consisted of 2 and 3 digit subtraction problems. Some of these required no borrowing, and some required borrowing from the tens', hundreds' or both tens' and hundreds' columns.

Three tests were constructed, a pretest, a posttest #1 and a posttest #2. Each test consisted of 24 different problems. The pretest and the posttest #1 were composed of 7 problems which did not require borrowing, 5 problems which required borrowing from the tens' column, 6 problems which required borrowing from the hundreds' column, and 6 problems which required borrowing from both the tens' and hundreds' columns (The unequal distribution of problem types was the result of an error which was not detected until the study was in progress). The posttest #2 was composed of 6 problems of each type. Two forms of each test were also constructed, each form consisting of the same items but arranged in a different order.

<u>Procedure</u>

Children were pretested while in the regular classroom. Half of the children received one form and half received the other. The teacher asked the children to solve the problems, and they were given the time to complete all problems. The experimenter was not in the room during this phase of the study. Two to three days following the experimental conditions, posttest #1 was administered. Posttest #2 was administered in the same manner two weeks later.

Those children who solved less than 60 percent of the problems correctly on the pretest were placed in the Unsatisfactory group. Those who solved 60 to 80 percent of the problems correctly were classified as Satisfactory. Of the 64 children who composed the final sample of children, 32 were placed in the Unsatisfactory group and 32 were placed in the Satisfactory group.

<u>Component Skills Training</u>

Subjects in this condition were presented with a series of 5 programmed instruction booklets (designed by the experimenter) over the course of 3 daily 30 to 40 minute sessions. Booklets 1, 2 and 3 were presented on day 1, booklet 4 was presented on day 2 and booklet 5 was presented on day 3. The program was designed to re-instruct each child in the component skills required for subtraction (operation sign recognition, column recognition, minuend and subtrahend recognition, recognition of the relative size of numerals, etc.), and to teach the steps involved in borrowing. In the program, one item of information was presented on each page, followed by a question based on that item. An item consisted of a sentence containing some information about subtraction. As in the following example, a question was asked and a space was provided for the child to enter his/her response.

If the bigger number is on the bottom in the ones' place or in the tens' place you must borrow to subtact.

$$\begin{array}{r} 4\ 6\ 8 \\ -2\ 7\ 4 \\ \hline \end{array}$$

Should you borrow on this problem? ________ (yes or no)

The child was instructed to correct his/her answer on the item by referring to the answer provided on the following page. Finally, he/she recorded the score obtained on that item. If the item was completed perfectly the first time, the child was instructed to place a coloured "star" beside his or her answer.

The individual components included in this training package were selected on the basis of a task analysis of the solution of a subtraction problem requiring borrowing performed by the author. In addition, some components were obtained from those included in computer programs designed to simulate children's solutions to subtraction problems (Young and O'Shea, 1981; Brown and Burton, 1978). These latter components were adapted to fit the requirements of the booklet form of presentation used in this study. A listing of the program components included in the Component Skills Training program is provided in Figure 1.

<u>Insert Figure 1 about here</u>

Figure 1: Program Components for Component Skills Training

Booklet #1

1. Judgements of relative magnitude of numerals (<,>).
2. Sign recognition (+,-).
3. Column identification (ones', tens', hundreds').
4. Identification of the number of ones', tens', hundreds' in each column.
5. Review.

Booklet #2

1. Order of operations, i.e. which column to start, where to go.
2. Location of largest and smallest number within a column (minuend, subtrahend).
3. Subtract number in subtrahend from number in minuend (always subtract <u>down</u>).
4. Identify when largest number is on bottom (subtrahend) in ones' column.
5. Review.

Booklet #3

1. Identify when largest number is on bottom (subtrahend) in ones' column.
2. Same as above for tens' column.
3. Identify when a problem requires borrowing (presence of 1 and/or 2).
4. Review.

Booklet #4

1. Review identification of borrowing problems.
2. Reduce numeral in minuend of tens' column.
3. Add ten to numeral in minuend of ones' column.
4. Identify problems requiring borrowing from tens' column, and apply 2 and 3.
5. Review.
6. Heuristics for identification of borrowing problems and execution of borrowing procedures in tens' column.
7. A mnemonic aid for above heuristics.
8. Review.

Booklet #5

1. Review above heuristics.
2. Identification of problems requiring borrowing from hundreds' column.
3. Application of heuristics to hundreds' column.
4. Review.
5. Identification of problems requiring borrowing from tens' and hundreds' columns.
6. Application of heuristics to 5.
7. Review and practice.
8. Criterion.

The demands of group testing and the technical aspects of providing this type of instructional package to young children made it impossible to ensure that each child attained a criterion level of performance on each item before progressing to the next item. Instead, pilot work with children experiencing difficulties in subtraction was conducted in an attempt to provide the item types and sequencing that would pose the least amount of difficulty for these children. Each of the last 2 items in the program consisted of 5 subtraction problems (10 in all) which were designed to serve as an index of how successfully the target skill was learned or, alternately, how well the program fared in teaching the skill.

Criterion Training

The format for this condition was similar to that of the Component Skills package. Subjects were presented with four booklets over the course of 3 daily sessions. The booklets consisted of 5 pages each, containing 1 to 10 problems on 4 of the 5 pages. Table 10 shows the distribution of problems throughout the program

Insert Table 10 about here

An item in this program consisted of one or more subtraction problems which the child was required to solve. The following is an example:

Solve these problems:

```
  2 7 4        7 4 6
 -1 4 2       -1 8 2
 ------       ------
```

The child was then instructed to turn the page, correct his or her answers, record the obtained score and reward perfect performance as in

Table 10: Distribution of Problems in the Criterion Training Program by Page Number and Booklet

page number	1	2	3	4	total
booklet					
1	1	2	3	4	10
2	2	4	6	8	20
3	4	6	10	10	30
4	10	10	10	10	40

the Component Skills program. Booklets 1 and 2 were administered on day 1, booklet 3 was administered on day 2, and booklet 4 was administered on day 3. This format resulted in the presentation of 30 problems per day for each child. The same types of problems were presented in each session. The problems consisted of 2 and 3 digit subtraction problems of the types contained in the pretest and posttests.

As in the Component Skills program, the last 10 problems of booklet 4 were used as a criterion index of how well the child had acquired the target skills.

Control Condition

Subjects in this condition received regular classroom instruction. This condition may alternately be viewed as an uncontrolled instruction condition in the sense that some children received additional classroom instruction in solving subtraction problems. In light of the results of Study I where children's subtraction performance remained fairly stable after one month despite additional instruction, the regular classroom control seemed an appropriate comparison group for the treatments examined here.

Design

Members of the Unsatisfactory and Satisfactory groups were randomly assigned to Component Skills Training, Criterion Training or Control conditions (note that subjects were not assigned to conditions on the basis of class membership). Following the pretest, groups of 4 to 8 subjects in the two experimental conditions were presented with the programmed instruction booklets in 3 daily sessions conducted in an empty classroom in the schools. Instructions on how to use the programs were provided by means of an introductory booklet which explained how the program worked and gave examples of items for the children to complete and score. Throughout the remedial sessions, the experimenter was present to answer questions about specific aspects of the programs. No information about how to complete any of the items was provided. Answers were limited to clarifying instructions.

Within 2 days following the remedial phase of the study, the subjects were presented with posttest #1 in the same manner as the pretest. Posttest #2 was presented 2 weeks later.

Error Analysis

As in Study 2, errors committed by each child on the pretest and posttests were assigned to one of the following four categories: 1) inversion errors; 2) computational errors; 3) borrowing errors; 4) other errors.

Inter-Observer Agreement

As in Studies 1 and 2, one test form for each of 20 children was selected randomly and scored by a second scorer who was trained in these error analysis procedures. Inter-observer agreement was calculated using the following formula:

(# of agreements of error type/# of agreements plus disagreements) X 100.

Analysis

This experiment employed a pre-post design where subjects were nested in experimental conditions. Analysis of covariance was selected as the analysis of choice where between group differences existed in pretest scores. The pretest scores were used as covariates. Comparisons between adjusted means were performed using the Bonferroni t test, as in Study 2.

Where large differences existed in cell variances, the Kruskal-Wallace H statistic was used, again as in Study 2. Multiple comparisons between cells were made using the Protected Rank Sum Test. A large experimentwise type 1 error is avoided by using this test only in the case of a significant H (Welkowicz et al., 1976).

Results

Inter-Observer Agreement

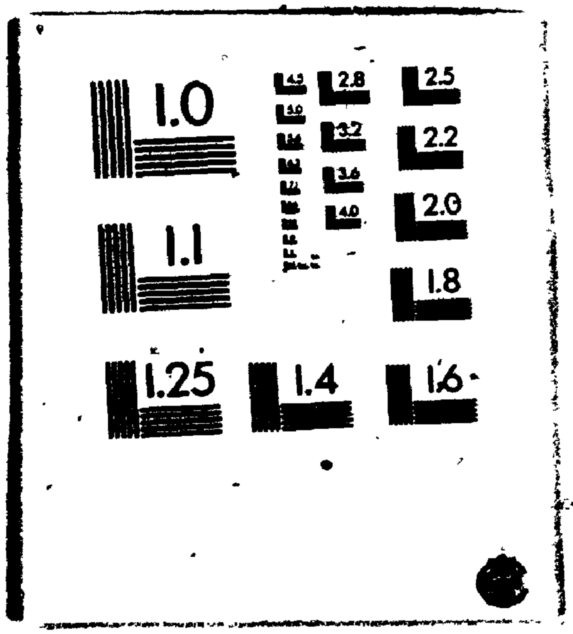
Inter-observer agreement for each type of error was as follows: inversion errors, 98 percent; computational errors, 81 percent; borrowing errors, 88 percent and other errors, 92 percent.

Problems which did not require borrowing were included in pretest and posttest measures. This was because one goal of training was to teach children to recognize when it was appropriate to borrow and when it was not. Because these problems did not require borrowing, the most likely type of error was computational. The remedial programs used in this study, however, were focused on teaching the procedures required for borrowing and did not attempt to promote computational accuracy. The errors which occurred on non-borrowing problems, therefore, could be interpreted as "noise" and would serve to introduce an unwanted source of variance into the analyses. In order not to obscure improvements in subjects' ability to apply borrowing procedures correctly, it was decided to eliminate non-borrowing problems from the analyses.

Number Correct

Comparison of cell variances for Unsatisfactory and Satisfactory groups yielded significant differences. Hartley's Fmax statistic computed for Unsatisfactory subjects was 12.08, $p<.05$. For the Satisfactory group, $F_{max}(10)=32.10$, $p<.01$. Due to the violation of the homogeniety of variance assumption, a non-parametric analogue of the analysis of variance, the Kruskal-Wallace H test, was used to examine the null hypothesis that the locations of the experimental populations were identical. Within Unsatisfactoy and Satisfactory conditions, Component Skills Training, Criterion Training and Control groups were compared using the number of borrowing problems solved correctly as the

2 OF/DE 2



For Unsatisfactory subjects at posttest #1, $H(2)=6.55$, $p<.05$. Between cell comparisons using the Protected Rank Sum test indicated that subjects in both Component Skills Training and Criterion Training conditions solved more borrowing problems correctly than controls, $Z=2.32$, $p=.0204$ and $Z=1.97$, $p=.0488$ respectively. At posttest #2, no significant differences were found between the two treatment conditions and controls. For Satisfactory Subjects, no significant effects of treatment condition were found at posttest #1 or at posttest #2. Means and standard deviations are presented in Table 11.

Insert Table 11 about here

Inversion Errors

For Unsatisfactory subjects, treatment effects were examined using a two way repeated measures analysis of covariance. Treatment condition was the between subject factor and trials (posttest #1 and posttest #2) was the within subject factor. The number of problems containing inversion errors on the pretest was the covariate. No main effects or interactions were found. Examination of Table 12 indicates that both treatment conditions resulted in a statistically nonsignificant reduction in inversion errors relative to controls.

Insert Table 12 about here

Comparison of cell variances for inversion errors for Satisfactory subjects yielded an $F_{max}(10)=197.00$, $p<.01$. The dependent measure used for the Kruskal-Wallace H test was the number of problems containing inversion errors on the pretest minus the number of problems containing inversion errors on the posttest. This was because subjects were matched on the basis of the number of problems solved correctly on the pretest before random assignment to treatment conditions. Consequently,

Table 11: Mean Number of Borrowing Problems Solved Correctly

		prestest	posttest #1	posttest #2
		Unsatisfactory		
GST	mean	1.27	8.27	11.27
	sd	2.97	5.92	7.20
	n	11		
CRT	mean	.91	7.64	6.64
	sd	2.07	6.58	7.42
	n	11		
Control	mean	.80	2.20	7.10
	sd	2.53	4.54	7.72
	n	10		
		Satisfactory		
CST	mean	11.80	14.80	15.80
	sd	1.14	1.75	2.35
	n	10		
CRT	mean	10.55	11.81	13.18
	sd	3.59	6.40	5.10
	n	11		
Control	mean	9.55	10.55	11.55
	sd	3.91	4.52	5.97
	n	11		

Table 12: Mean Number of Borrowing Problems Containing Inversion Errors

		pretest	posttest #1	posttest #2
		Unsatisfactory		
CST	mean	13.00	6.09	4.55
	sd	6.02	6.55	6.86
	n	11		
CRT	mean	8.27	3.82	4.27
	sd	7.79	6.55	6.71
	n	11		
Control	mean	6.10	7.20	7.90
	sd	5.67	7.00	7.61
	n	10		
		Satisfactory		
CST	mean	1.20	.20	.40
	sd	1.62	.42	.52
	n	10		
CRT	mean	1.64	1.27	.73
	sd	3.88	3.91	1.68
	n	11		
Control	mean	1.73	2.18	3.45
	sd	2.05	3.92	5.89
	n	11		

between group differences in the number of errors committed on the pretest could not be avoided.

The Kruskal-Wallace H test yielded no significant treatment effects at posttest #1 or at posttest #2 for Satisfactory subjects. Means and standard deviations are presented in Table 12.

Computational, Borrowing and Other Errors

Comparisons of cell variances for computational, borrowing and other errors yielded the following results for Unsatisfactory subjects: $F_{max}(10)=21.32$, $p<.01$; $F_{max}(10)=20.86$, $p<.01$; $F_{max}(10)=54.38$, $p<.01$ respectively. For Satisfactory subjects, $F_{max}(10)=11.28$, $p<.05$; $F_{max}(10)=125.41$, $p<.01$; and $F_{max}(10)=35.71$, $p<.01$ respectively. The dependent measure for these error types was the number of problems containing the error on the pretest minus the number of problems containing the error on the posttest.

Kruskal-Wallace H tests yielded no significant treatment effects for Unsatisfactory subjects at posttest #1 and posttest #2 for any of the above dependent measures. As well, no significant treatment effects were found for Satisfactory subjects at posttest #1 or posttest #2 for any of the dependent measures. Means and standard deviations are presented in Tables 13, 14 and 15.

Insert Tables 13, 14 and 15 about here

The results of these analyses indicate that both Component Skills Training and Criterion Training were effective in increasing the number of borrowing problems solved correctly, but only for those subjects who fell into the Unsatisfactory category. A reduction of inversion errors appears to have contributed most to this result, although small reductions in borrowing and other errors can also be seen in Tables 14 and 15.

Table 13: Mean Number of Borrowing Problems Containing Computational Errors

		prestest	posttest #1	posttest #2
Unsatisfactory				
CST	mean	.64	1.64	1.18
	sd	1.21	1.75	1.08
	n	11		
CRT	mean	1.18	1.36	1.73
	sd	1.60	1.36	1.42
	n	11		
Control	mean	4.00	3.40	2.10
	sd	5.01	3.95	3.38
	n	10		
Satisfactory				
CST	mean	.90	.70	1.20
	sd	1.29	1.06	2.15
	n	10		
CRT	mean	3.09	2.36	2.64
	sd	2.02	3.23	3.70
	n	11		
Control	mean	2.64	2.00	1.64
	sd	2.34	2.41	1.29
	n	11		

Table 14: Mean Number of Borrowing Problems Containing Borrowing Errors

		prestest	posttest #1	posttest #2
		Unsatisfactory		
CST	mean	3.09	1.27	1.45
	sd	5.30	2.69	3.39
	n	11		
CRT	mean	5.09	2.36	4.09
	sd	7.11	4.43	6.09
	n	11		
Control	mean	3.40	2.90	1.00
	sd	3.34	5.88	1.56
	n	10		
		Satisfactory		
CST	mean	1.90	1.20	.20
	sd	2.18	1.62	.42
	n	10		
CRT	mean	3.00	2.55	1.27
	sd	2.93	4.70	2.33
	n	11		
Control	mean	2.91	2.27	2.36
	sd	3.11	2.76	3.67
	n	11		

Table 15: Mean Number of Borrowing Problems Containing Other Errors

		prestest	posttest #1	posttest #2
		Unsatisfactory		
CST	mean	1.55	.64	.73
	sd	3.21	.92	1.56
	n	11		
CRT	mean	4.64	3.09	3.00
	sd	6.83	5.39	4.82
	n	11		
Control	mean	6.60	5.10	3.30
	sd	6.31	6.72	5.77
	n	10		
		Satisfactory		
CST	mean	1.30	.40	.60
	sd	2.50	.70	.97
	n	10		
CRT	mean	.18	1.18	1.00
	sd	.60	3.60	2.49
	n	11		
Control	mean	1.45	1.00	1.27
	sd	2.11	2.41	2.20
	n	11		

Controls

In an attempt to understand why treatment versus control differences disappeared at posttest #2 for Unsatisfactory subjects, the performance of control subjects was examined. Table 11 shows that while treatment means for Unsatisfactory subjects remained fairly stable during the two weeks between posttest #1 and posttest #2, the number of problems solved correctly by control subjects increased. This increase appears to have been responsible for wiping out the treatment effect between posttests #1 and #2.

Some evidence is available to suggest that the improvement of the Control group may have resulted from extra instruction in borrowing procedures provided by the classroom teacher during the interval between posttest #1 and #2. Table 16 presents the individual subject data for the Unsatisfactory Control group. Subjects #12, #11 and #10 dramatically improved their performance in the two week interval following posttest #1. The errors committed by these subjects suggested that they had difficulty with borrowing procedures. Subject #12 and #10 for example, wrote "0" as the answer whenever borrowing was required, an error which was categorized as "other". These errors were eliminated by posttest #2 and were replaced by correct borrowing. Subject #11 committed borrowing errors which were also eliminated by posttest #2. All three of these children came from the same classroom and were the only control subjects from this classroom. Although the nature of the training provided for the children in Study 3 was not revealed to the teachers, they could have been informed by the children participating in the study. It is not surprising, then, that a classroom teacher would wish to provide additional instruction in borrowing once he/she became aware that some of the students had received remedial instruction (while some had not). In the same vein, this mechanism may have been instrumental in the maintenance of treatment gains for children who were assigned to the experimental training conditions.

Given the above findings, and the results of Study 1 which indicated that children having difficulty with subtraction did not improve over the course of one month despite regular classroom instruction, it appears that the effects of Component Skills Training and Criterion Training were maintained when performance was examined two weeks after training had ended.

Insert Table 16 about here

It is also interesting to note that, although subjects were placed in the Unsatisfactory category on the basis of solving fewer than 60 percent of the pretest problems correctly, 27 of 32 subjects committed errors on all borrowing problems. Of these 27 subjects, only 4 applied borrowing procedures correctly. For these 4 children, correct procedures were applied on 1, 6, 7 and 3 problems respectively (out of 17). This suggests that lack of knowledge of the complete set of borrowing skills is the most common difficulty for children who perform at a level considered unsatisfactory by most academic standards (less than 60%). Correct but inconsistent application of borrowing procedures occurs infrequently in members of this group (28% of Unsatisfactory subjects).

Remedial Program Performance

Although it was found that the remedial programs resulted in improvements of Unsatisfactory subjects as a group, some children did

Table 16: Number of Borrowing Problems Solved Correctly for Unsatisfactory Control Subjects

Subject	Pretest	Posttest #1	Posttest #2
12	0	0	12
11	0	0	16
10	0	0	18
30	0	0	0
35	0	13	14
53	0	0	0
49	0	0	0
48	8	8	11
54	0	0	0
59	0	1	0

not do very well during training. Because it would be unreasonable to expect children who had difficulty during training to demonstrate posttest improvements, the finding that some children had difficulty warrants further investigation. In order to identify children who had difficulty with the remedial training, a median split was done on the basis of the number of criterion problems solved correctly. This median split was done for subjects in Unsatisfactory and Satisfactory conditions. Recall that the criterion measure consisted of the last 10 problems within each training program.

This median split yielded two subsamples of children within each of the Unsatisfactory and Satisfactory conditions. Those above the median, considered to have successfully completed the programs, were labelled Tutorial-High, and those below the median, considered to have been less successful, were labelled Tutorial-Low.

As suggested above, only those children who complete remedial training successfully would be expected to show improvements in classroom performance. Those Unsatisfactory children who would be successful in training (and presumably later in the classroom) could not be identified on the basis of data obtained at the pretest. For example, the number of borrowing and non-borrowing problems solved correctly at the pretest did not differ between Tutorial-High and Low groups. Further, scores on the criterion measure which was used to determine success during training did not significantly correlate with pretest score, $r = .22$, $df = 20$. There also did not appear to be any difference in the number or types of errors most frequently committed by Tutorial-High and Low subjects.

In the hope of shedding some light on why Tutorial-Low children did not fare as well as their Tutorial-High counterparts, performance during training was examined. Because of the small number of subjects involved and the post-hoc nature of this examination, it was decided to employ analysis of trends for Tutorial-High and Tutorial-Low children. Material contained in the Component Skills and Criterion Training programs was divided into small units as described below, and accuracy of performance was examined using a repeated measures ANOVA across these units. Only significant F values were interpreted as representing actual fluctuations in the observed pattern of performance. Because the purpose of this type of analysis was to generate questions for further study, training units were grouped in several ways in an attempt to obtain a better understanding of what happened during training.

Criterion Training: Unsatisfactory Group

In order to investigate performance throughout the Criterion Training program, the program was divided into 9 blocks of 10 problems each. The criterion measure used to divide subjects into Tutorial-High and Low groups consisted of the last 10 problems in the program and these were, therefore, excluded from the analysis.

For Unsatisfactory subjects in the Tutorial-High condition, a one way repeated measures ANOVA with the 9 training blocks as the within subject factor indicated a significant main effect for block, $F(8, 32)=7.56$, $p<.001$. This main effect showed that the number of problems solved correctly at block #7 was greater than at block #2. A Bonferroni t comparison between blocks #1 and #9 showed that the number of problems solved correctly for Unsatisfactory subjects in the Tutorial-High condition was significantly higher at the end of the program (on block

#9) than at the beginning (on block #1), t(4)=3.38, p<.05. Tests for trend indicated significant linear, F(1, 4)=10.56, p=.0314, and quadratic, F(1, 4)=12.39, p=0224 trends. In order to further understand these data, a 3 (day) X 3 (blocks) analysis of variance with repeated measures on both days and blocks was performed (the training was conducted over 3 days with 3 blocks completed per day). This analysis revealed a significant main effect of day, F(2, 8)=11.79, p=.0041, and a significant day X block interaction, F(4, 16)=4.60, p=.0116. The main effect of day was accompanied by a significant linear component, F(1, 4)=11.50, p=.0275. Bonferroni t comparisons between means on days 1 and 2, days 1 and 3, and days 2 and 3 indicated that the number of problems solved correctly increased from day 1 to day 2, t(4)=3.58, p<.05. This increase was maintained at day 3, t(4)=3.40, p<.05 (day 1 versus day 3). In all, these results describe a steady increase in the number of problems solved correctly from block #2 through block #5, followed by a levelling off through block #9. Examination of Fig. 1 suggests that the increase occurred on day 2 and was maintained during day 3. The previous analysis also showed that this increase in the number of problems solved correctly during training continued on through to the posttest.

A one way repeated measures ANOVA across the 9 training blocks for Tutorial-Low subjects in the Unsatisfactory group again indicated a significant main effect for blocks, F(8, 40)=2.47, p=.0282. This shows that subjects solved significantly more problems correctly on block #1 than on block #9. A significant cubic trend was also found, F(1, 5)=13.87, p=.0136. To assist in the interpretation of this trend, a 3 (days) X 3 (blocks) repeated measures ANOVA was performed, as above.

This analysis yielded a significant main effect for blocks, $F(1, 5) = 13.19$, $p = .0015$, a significant linear trend for blocks, $F(1, 5) = 26.54$, $p = .0036$, as well as a significant day X block interaction, $F(4, 20) = 2.99$, $p = .0435$. Bonferroni t comparisons between means on days 1 and 2, days 1 and 3, and days 2 and 3 indicated that there was a significant decrease in the mean number of correct solutions per block from day 1 to day 3, $t(4) = 4.57$, $p < .05$. No differences were significant within days. These analyses show a decrease in performance from day 1 to day 3. This decrease did not carry over to the posttest, however, as the previous analysis showed no pre-post differences for this group. These data are presented in Fig. 2.

Insert Fig. 2 about here

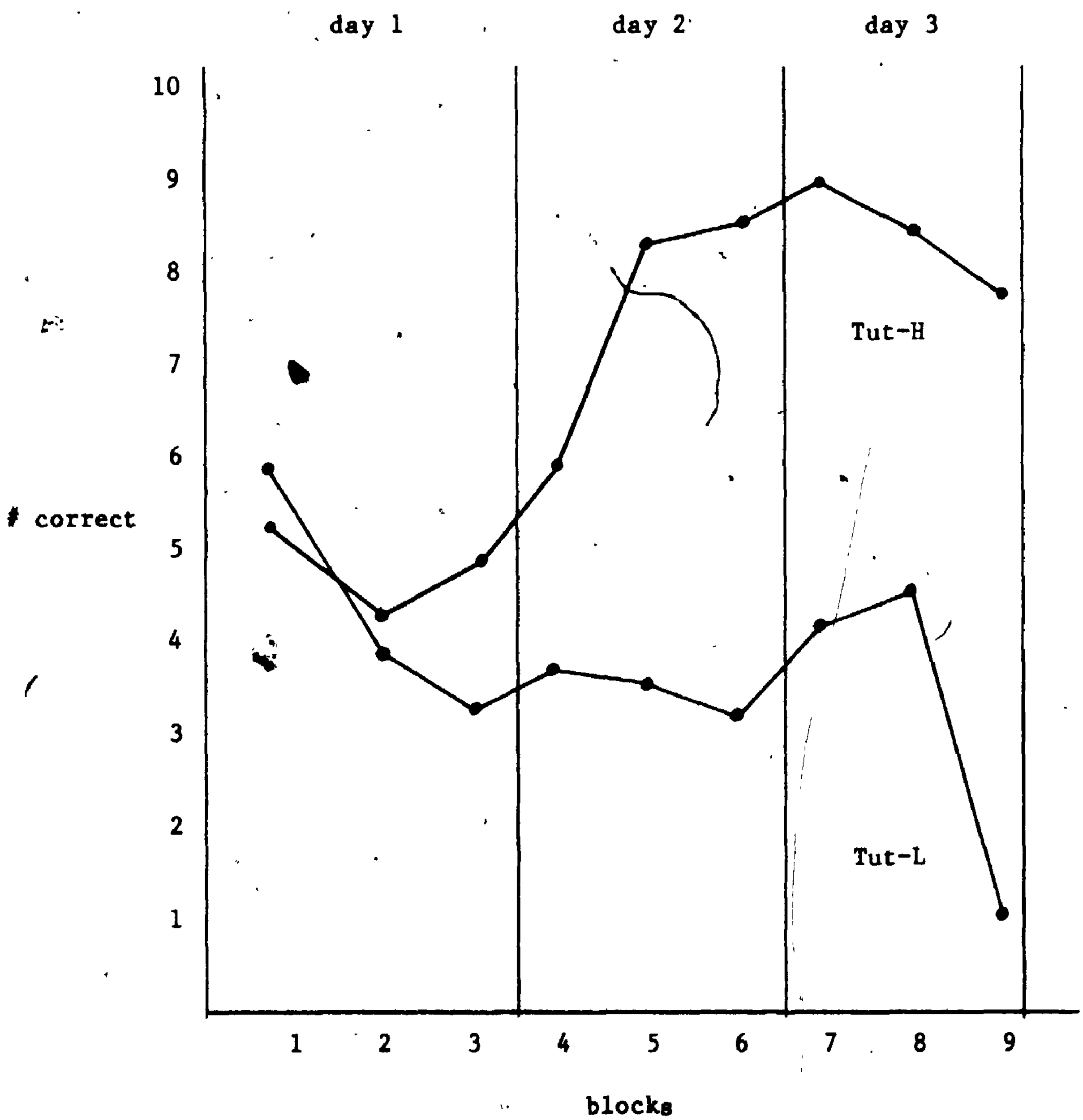
When the trends for subjects in the Tutorial-High and Tutorial-Low conditions are examined visually, it appears that differences between the two groups' performance trends occurred at block 4 and became more pronounced with subsequent blocks. For subjects in the Tutorial-Low group, it appears that decreases in the number of training problems solved correctly roughly coincided with days.

Satisfactory Group

For Satisfactory subjects in the Tutorial-High condition, no effect for any of the above factors was found. As can be seen in Fig. 3, performance for these subjects appears to commence at a fairly high level and to remain stable across the 9 training blocks.

For Satisfactory subjects in the Tutorial-Low condition, the above one way ANOVA across blocks yielded a significant main effect, $F(8, 24) = 3.6$, $p = .007$, a significant quadratic trend, $F(1, 3) = 19.12$, $p = .0221$, as

Figure 2: Mean Number of Problems Solved Correctly Across Training Blocks in the Criterion Training Condition, Unsatisfactory Group



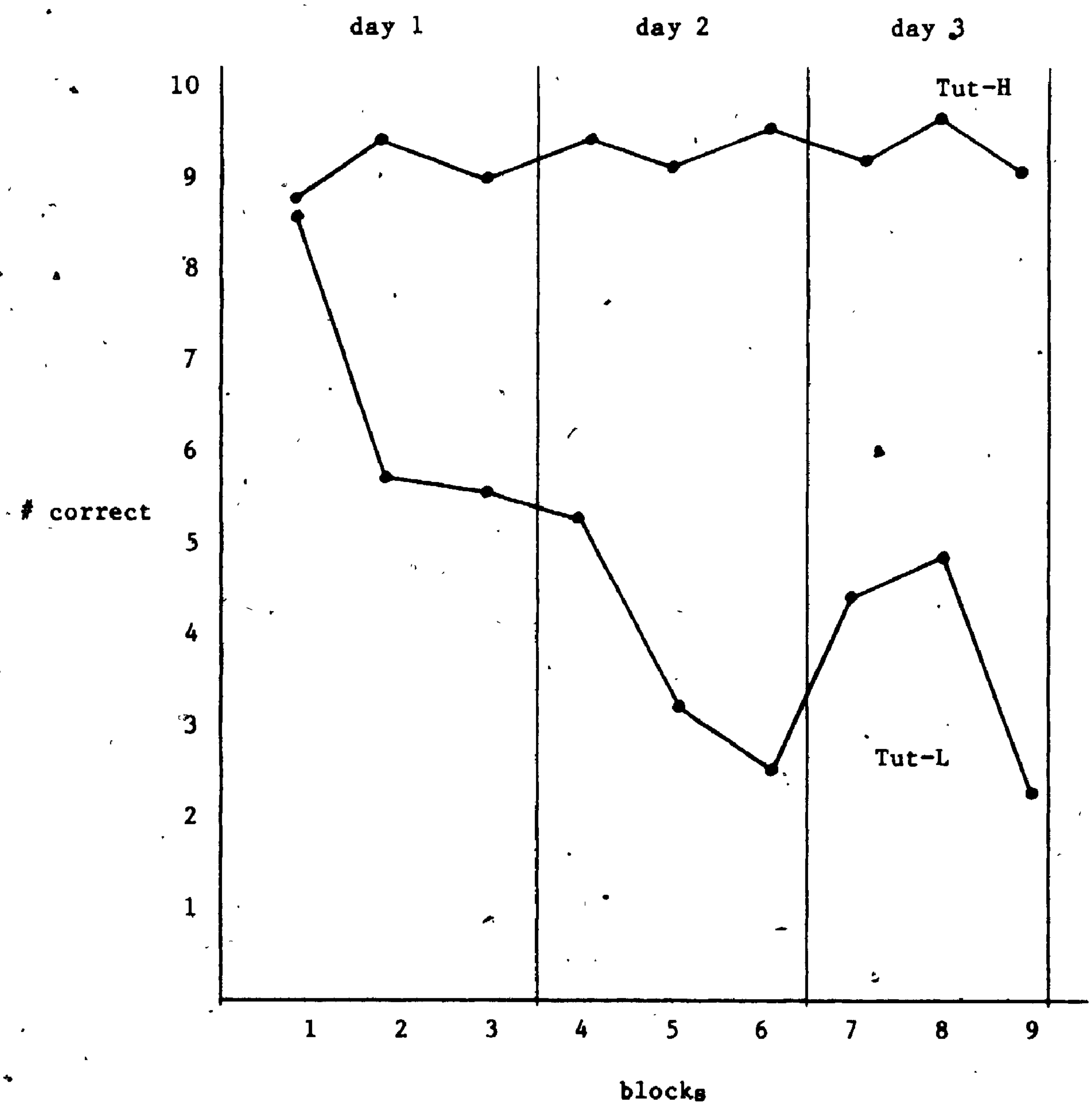
well as a significant cubic trend, $F(1, 3)=37.87$, $p=.0086$. The main effect reflects a significant reduction in the number of problems solved correctly from block #1 to block #9. A 3 (days) X 3 (blocks) ANOVA yielded a main effect for blocks only, $F(2, 6)=14.25$, $p=.0053$, as well as significant linear trend for blocks, $F(1, 3)=23.38$, $p=.017$. Bonferroni t comparisons between blocks #1 and #3 for day 1, day 2 and day 3 showed that the number of problems solved correctly decreased from block #1 to #3 at day 1, $t(4)=4.23$, $p<.05$, and at day 2, $t(4)=3.85$, $p<.05$. These results suggest that for Satisfactory subjects in the Tutorial-Low condition, performance decreased fairly steadily through days 1 and 2 (blocks #1 through #6). Although the differences were not statistically significant, visual inspection of Fig. 3 shows that the number of problems solved correctly increased at the beginning of the third day (blocks #7 and #8), and then decreased again (block #9).

Insert Fig. 3 about here

Borrowing Problems Only: Unsatisfactory Group

When constructing the Criterion Training program, it was thought desirable to include in each block a small number of problems which did not require borrowing. The purpose was to shape correct responses to both borrowing and non-borrowing types of problems. It was also thought best to vary the number of borrowing and non-borrowing problems from block to block, to try to ensure that children relied only on the critical features of the problems when deciding when to borrow. However, the change in the number of borrowing and non-borrowing problems from block to block presents some interpretational difficulties. Fluctuations in the number or problems solved correctly

Figure 3: Mean Number of Problems Solved Correctly Across Training Blocks in the Criterion Training Condition, Satisfactory Group

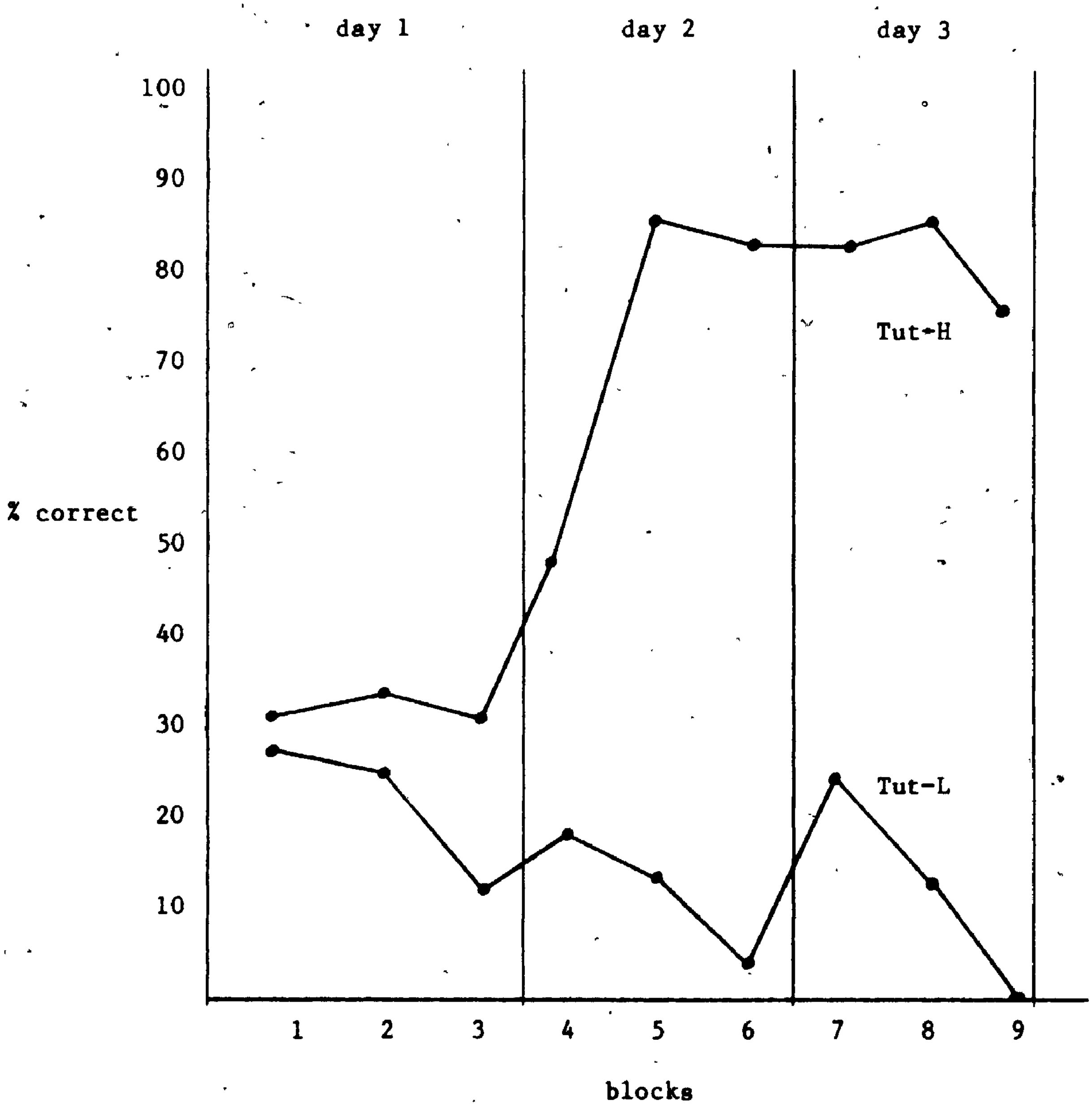


from block to block, seen especially in the performance of Tutorial-Low children, could coincide with the number of borrowing problems contained in each block. If Tutorial-Low children (at least in the Unsatisfactory group) had difficulty solving problems with borrowing, the number of problems solved correctly might be greater in those blocks containing more non-borrowing problems. In order to examine this question, all problems which did not require borrowing were eliminated and the above analyses were repeated with the percentage of borrowing problems solved correctly as the dependent measure.

For Unsatisfactory subjects in the Tutorial-High condition, a significant main effect across all 9 blocks was found, $F(8, 32)=6.40$, $p=.0001$, along with a significant linear trend, $F(1, 4)=7.51$, $p=.0519$. To better understand these results, a 3 (day) X 3 (block) ANOVA was performed. A significant main effect of day, $F(2, 8)=8.49$, $p=.0105$ was found. This effect was partialled into a significant linear component, $F(1, 4)=8.09$, $p=.0466$. A significant day X block interaction was also obtained, $F(4, 16)=3.15$, $p=0434$. Bonferroni *t* tests between the mean percentage of problems solved correctly on days 1 and 2, days 1 and 3, and days 2 and 3 indicated that these percentages increased from day 1 to day 2, $t(4)=3.25$, $p<.05$, and from day 1 to day 3, $t(4)=4.93$, $p<.01$. These results basically replicate those of the earlier analyses which included problems that did not require borrowing. Both analyses showed a linear increase in the number of problems solved during day 2, followed by a levelling off on day 3, as seen in Fig. 4.

<u>Insert Fig. 4 about here</u>

Figure 4: Mean Percentage of Borrowing Problems Solved Correctly Across Training Blocks in the Criterion Training Condition, Unsatisfactory Group.



For Unsatisfactory subjects in the Tutorial-Low condition, a one way repeated measures ANOVA across the 9 training blocks revealed no significant effects. A 3 (days) X 3 (blocks) ANOVA also failed to show any significant effects, although the main effect for block approached significance, $F(2, 10)=3.75$, $p=.0611$. Inspection of Fig. 4 shows that the percentage of borrowing problems solved correctly appears to decline each day with some recovery between days. This was the pattern observed when non-borrowing problems were included, as shown in Fig. 2. The earlier analysis produced a significant decrease in the number of problems solved correctly from day 1 to day 3. Although similar in trend to the earlier results, floor effects appear to have prevented this comparison from reaching statistical significance when borrowing problems only were included.

Borrowing Problems Only: Satisfactory Group

For Satisfactory Tutorial-High subjects, the absence of significant effects in both the one way and 3 X 3 ANOVAs is consistent with performance which begins at a high level and remains high. This result is identical with that obtained in the previous analysis of borrowing and non-borrowing problems.

For Satisfactory Tutorial-Low subjects, a one way ANOVA across all 9 blocks yielded a significant main effect for block, $F(8, 24)=2.36$, $p=.0493$, and a significant quadratic trend, $F(1, 3)=20.95$, $p=.0196$. A 3 X 3 ANOVA yielded a significant main effect of block, $F(2, 6)=6.06$, $p=.0363$, and a significant linear trend for block, $F(1, 3)=20.95$, $p=.0196$. Bonferroni t comparisons between means for blocks #1 and #2, blocks #1 and #3, and blocks #4 and #6 (as suggested by visual inspection) indicated a significant drop in the percentage of borrowing problems solved correctly at day 1, $t(4)=3.88$, $p<.05$ for blocks #1 and #2. No other differences between days were significant. These results show that the percentage of problems solved correctly decreased fairly steadily through days 1 and 2 and then increased on day 3. These results are again fairly consistent with those of the earlier analyses which showed a significant drop between blocks #1 and #3 on day 1, and blocks #1 and #3 on day 2. Visual inspection of Figs. 3 and 5 confirms this similarity for Tutorial-Low subjects.

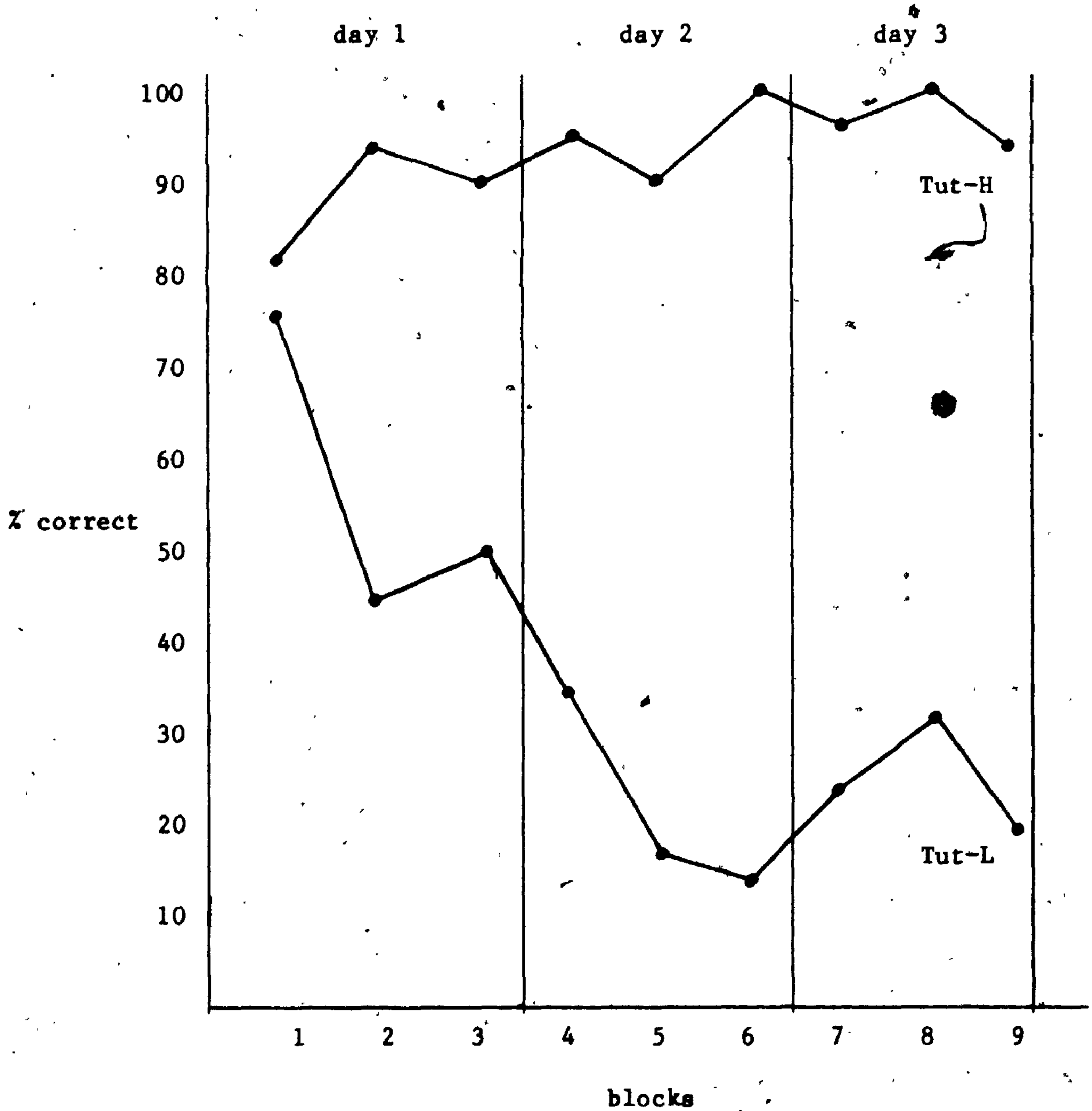
Insert Fig. 5 about here

In summary, patterns in the number of problems solved correctly for subjects in the Criterion Training condition were similar whether non-borrowing problems were included or not. Increases in the number of problems solved correctly for Unsatisfactory Tutorial-High subjects occurred at day 2. Within day decreases were observed for Unsatisfactory and Satisfactory Tutorial-Low subjects, as well as some between day recovery. These patterns were seen when non-borrowing problems were included in the analyses and when they were omitted. Differences in the number of borrowing problems contained in each block, therefore, cannot explain the fluctuations in performance described above. Instead, changes in the number of problems solved correctly during training appear to be in some way related to days.

Component Skills Training

Unlike the Criterion Training program which was similar throughout, the Component Skills Training program consisted of 5 booklets, each unique in content. It seemed reasonable to assume that the content of each booklet, rather than the day of administration, would have the

Figure 5: Mean Percentage of Borrowing Problems Solved Correctly Across Training Blocks in the Criterion Training Condition, Satisfactory Group.



strongest effect on performance. As we described earlier, the first 3 booklets dealt with pre-borrowing skills such as column recognition, location of the largest digit in a column, etc. Booklets 4 and 5 contained items which presented the actual borrowing procedures. Because borrowing posed more difficulty than any other subtraction skill for many of the children in this study, booklets 4 and 5 were expected to be the most difficult.

In order to evaluate the performance of subjects as they worked through the Component Skills program, each booklet was divided into two blocks, the first block containing the first half of the items and the second block containing the second half of the items. There were 10 blocks in all, two for each of the 5 booklets. As stated above, each block was composed of a unique set of items. The number of items within each booklet also varied. As in the Criterion Training program, the 10 criterion items used to group subjects as either Tutorial-High or Low were located at the end of booklet 5. These were not included in the formation of the item blocks and were, therefore, excluded from the analyses.

Unsatisfactory Group

A 5 (booklet) X 2 (block) ANOVA with repeated measures on both booklet and blocks was performed with the percentage of items answered correctly per block as the dependent variable. Booklet was included as a factor in order to determine subjects' performance over a particular content domain. Block was included as a factor in order to observe possible performance variations within these content domains. For Unsatisfactory subjects in the Tutorial-High condition, a significant booklet X block interaction was found, $F(4, 16)=4.38$, $p=.014$. Bonferroni t comparisons, however, failed to reveal any significant differences in the percentage of items answered correctly between blocks #1 and #2 of each booklet. Inspection of Fig. 6 shows a somewhat fluctuating pattern, with performance remaining above 75 percent accuracy throughout all 5 booklets.

Insert Fig. 6 about here

For Unsatisfactory subjects in the Tutorial-Low condition, the same analysis yielded a main effect for booklet, $F(4, 20)=5.27$, $p=.0046$, as well as a significant linear trend for booklet, $F(1, 5)=7.15$, $p=.0041$. The main effect indicated that performance was lower in booklet 5 than in booklet 1. Bonferroni t comparisons between booklets 4 and 5, 3 and 5, and 2 and 5 (as suggested by visual inspection of Fig. 6) indicated that the percentage of items answered correctly in booklet 5 was lower than that in booklet 3, $t(6)=3.92$, $p<.05$. This analysis, along with the inspection of Fig. 6, suggests that performance remained fairly stable until the beginning of booklet 4, where it began to decline steadily through to the end of booklet 5. Differences in trends for High and Low Tutorial groups became apparent at booklet 4 where the High group remained stable while the Low group began to deteriorate.

Satisfactory Group

As in the Criterion Training condition, Tutorial-High subjects in the Satisfactory group began at a high level of performance (greater than 90 percent accuracy) and maintained that level throughout the program. These data are presented in Fig. 7. No further analyses were conducted.

Insert Fig. 7 about here

Figure 6: Mean Percentage of Items Answered Correctly Across Training Blocks in the Component-Skills Training Condition, Unsatisfactory Group.

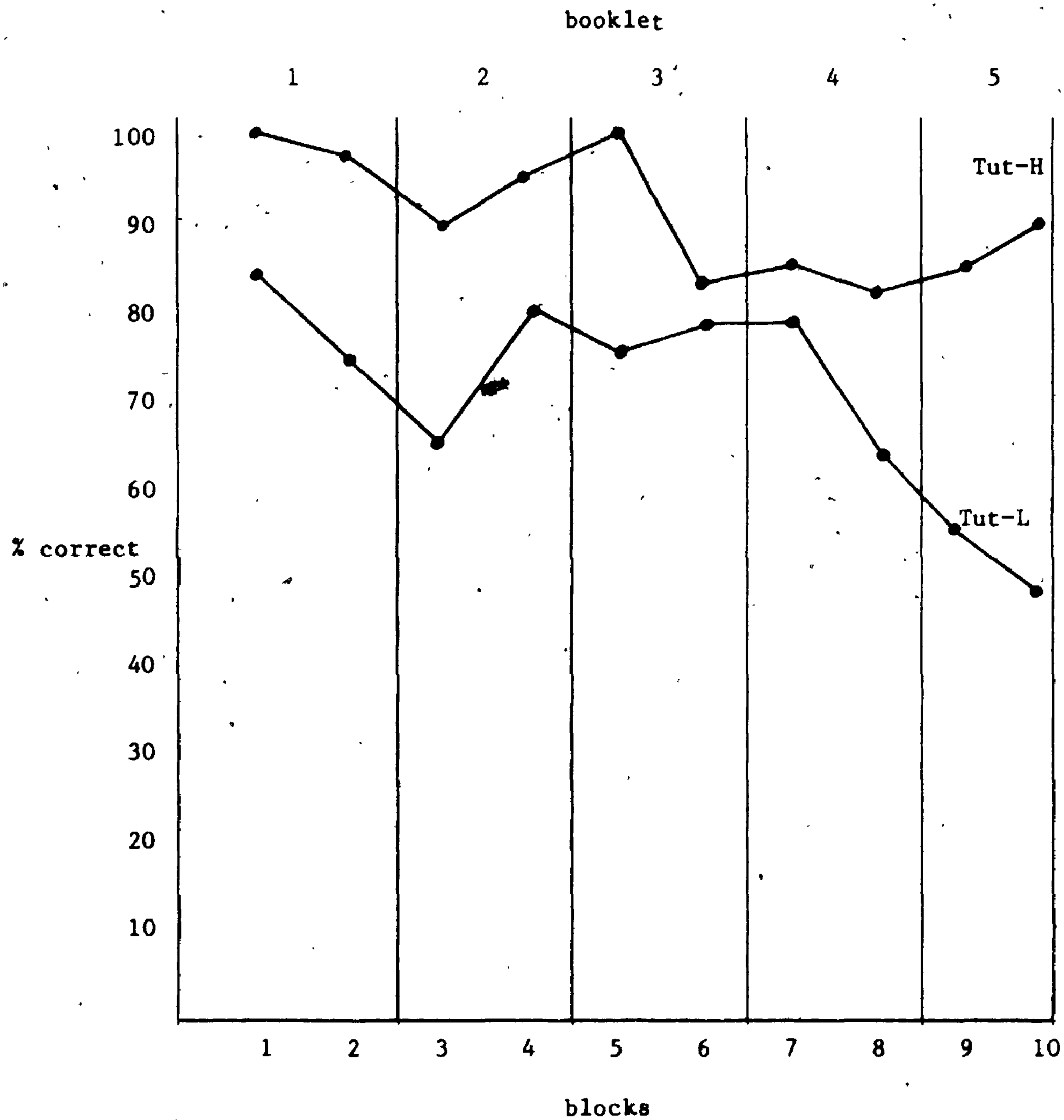
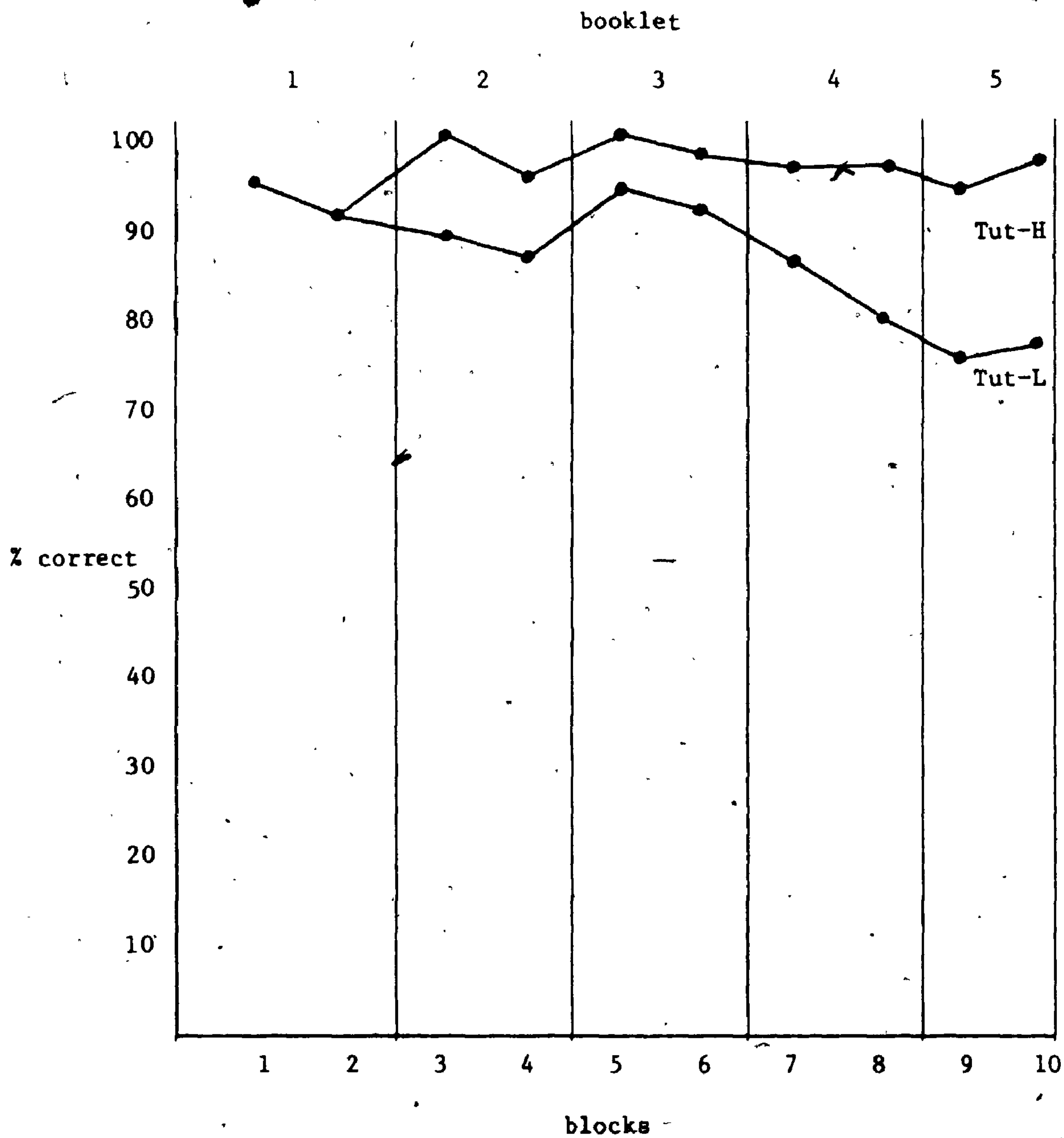


Figure 7: Mean Percentage of Items Answered Correctly Across Training Blocks in the Component Skills Training Condition, Satisfactory Group.



The same 5 X 2 repeated measures ANOVA for Tutorial-Low subjects yielded no significant effects. Bonferroni t comparisons between booklets 1 and 5, and 3 and 5, indicated that, as was the case for Unsatisfactory subjects in the Tutorial-Low condition, subjects answered fewer items correctly in booklet 5 than in booklet 1, $t(4)=3.79$, $p<.05$, and than in booklet 3, $t(4)=4.67$, $p<.05$. It appears that booklet 5 is more difficult for these children than the first 3, and differences in trends can be detected between High and Low subjects by booklet 4. These data are also presented in Fig. 6. Profiles for individual subjects are presented in Figs. 8 through 15.

Insert Figs. 8 through 15 about here

Sex

Of the 64 children examined in Study 3, 38 percent of the sample was male. Forty-seven percent of the Unsatisfactory condition consisted of males, while 28 percent of the Satisfactory condition was male. On a percentage basis, it appears that males are slightly more likely to perform below the acceptable academic standard of 60 percent. For Unsatisfactory and Satisfactory Tutorial-High conditions, 40 and 36 percent respectively were male. These percentages reflect the sex distribution found in the original sample. For Unsatisfactory and Satisfactory Tutorial-Low conditions, 67 and 30 percent respectively were male. This suggests that, of those children who were good candidates for remedial work, males were more likely than females to experience some difficulty with training.

Figure 8: Number of Problems Solved Correctly During Criterion Training for Unsatisfactory Tutorial-High Subjects

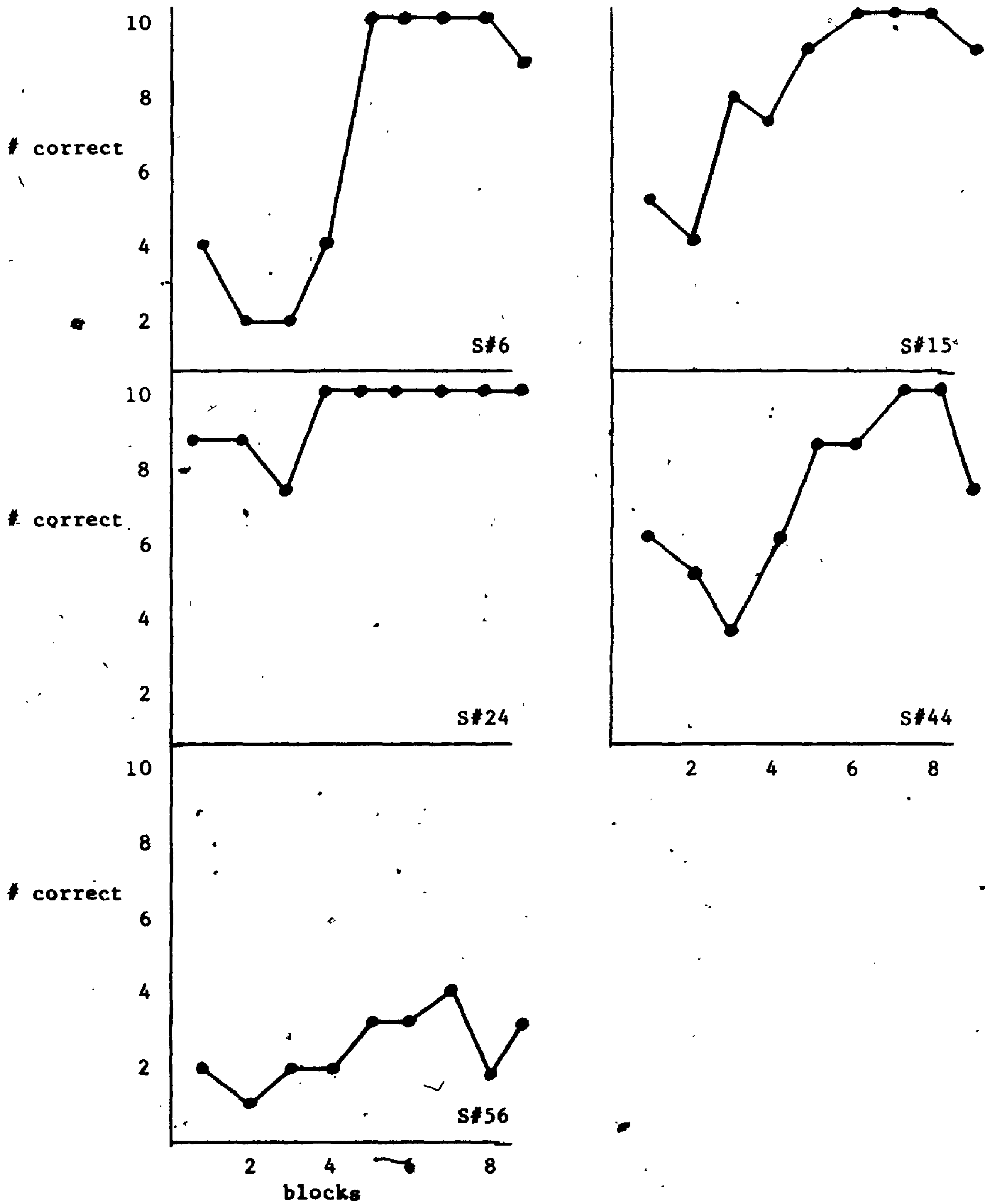


Figure 9: Number of Problems Solved Correctly During Criterion Training for Unsatisfactory Tutorial-Low Subjects

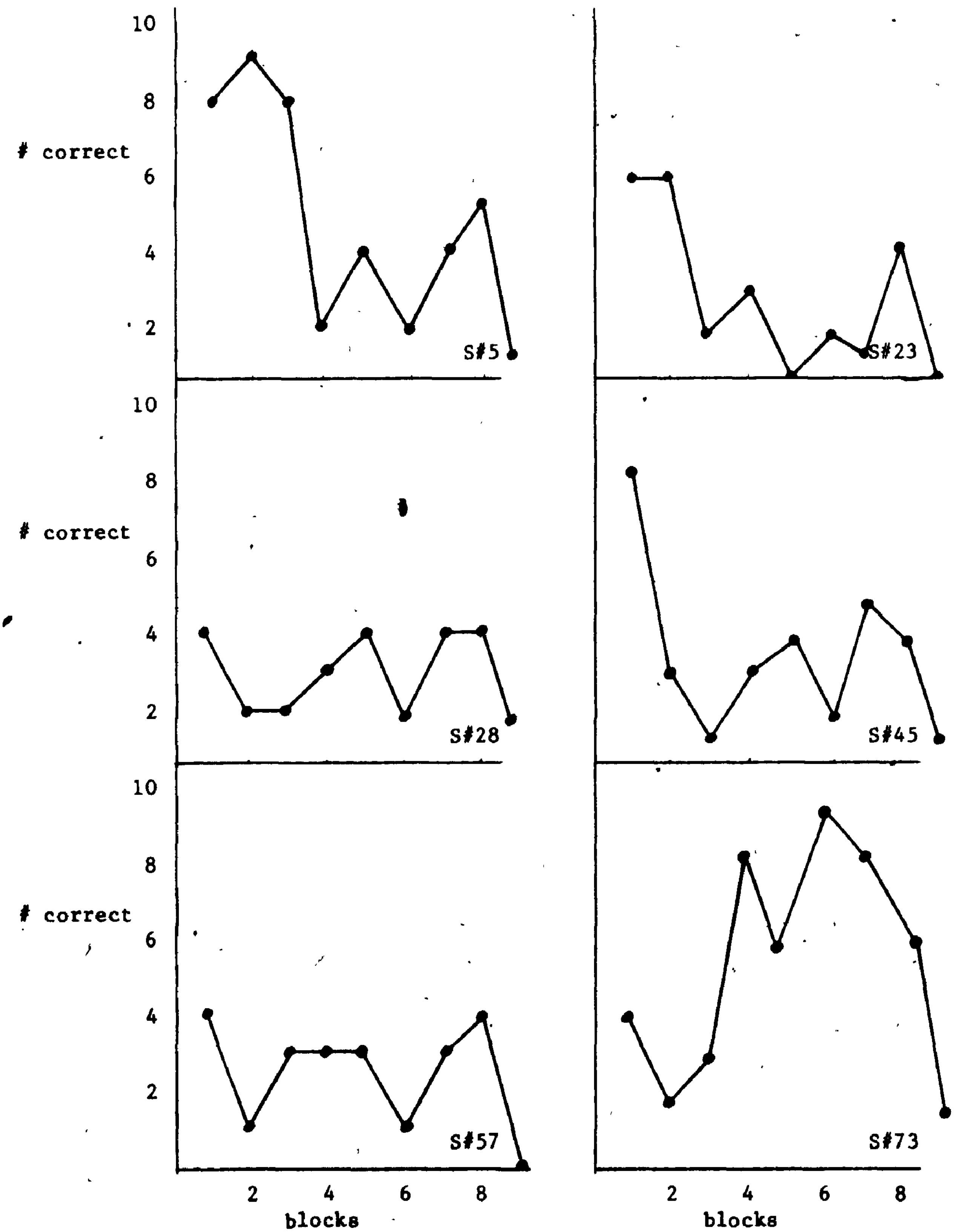


Figure 10: Number of Problems Solved Correctly During Criterion Training for Satisfactory Tutorial-High Subjects

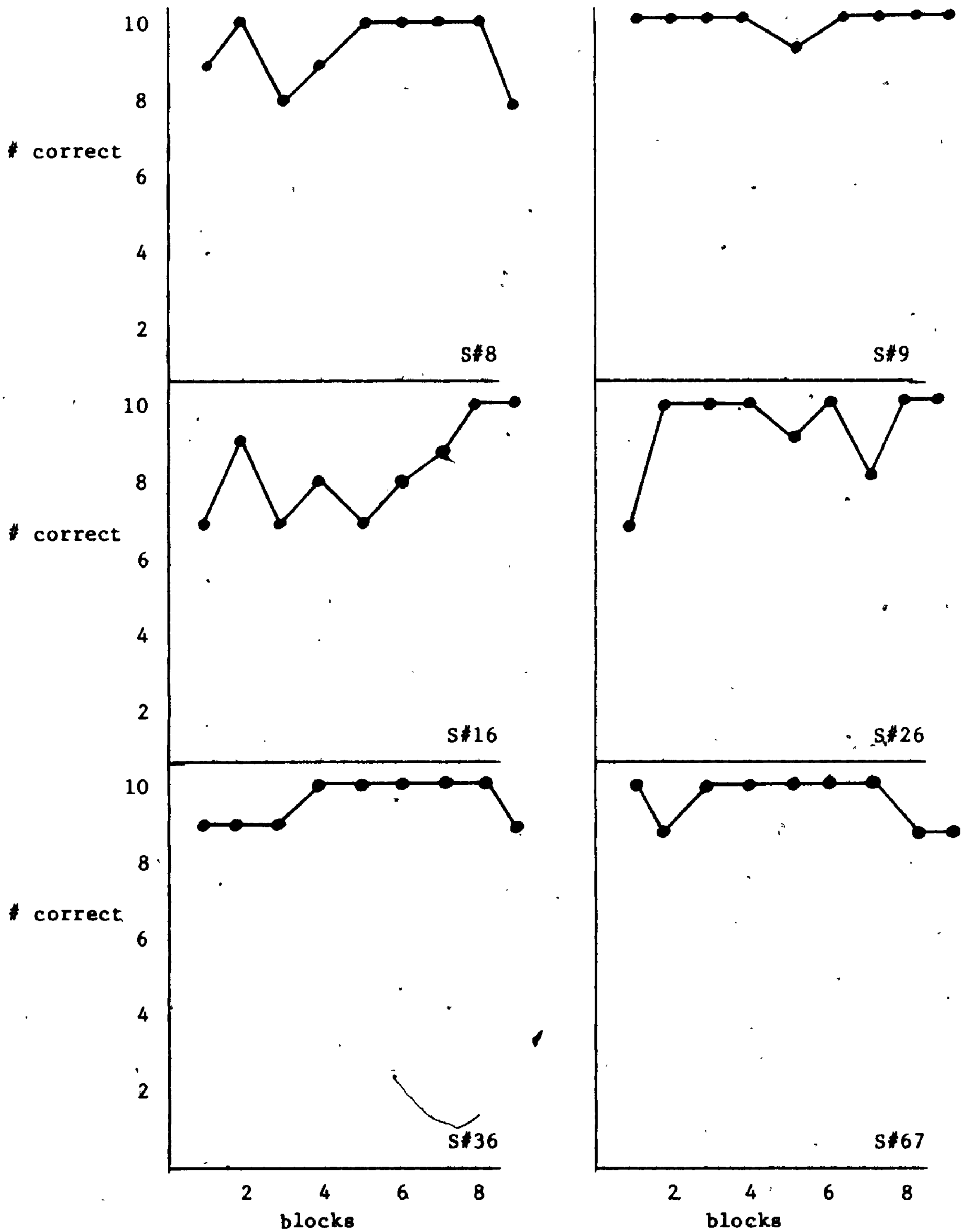


Figure 11: Number of Problems Solved Correctly During Criterion Training for Satisfactory Tutorial-Low Subjects

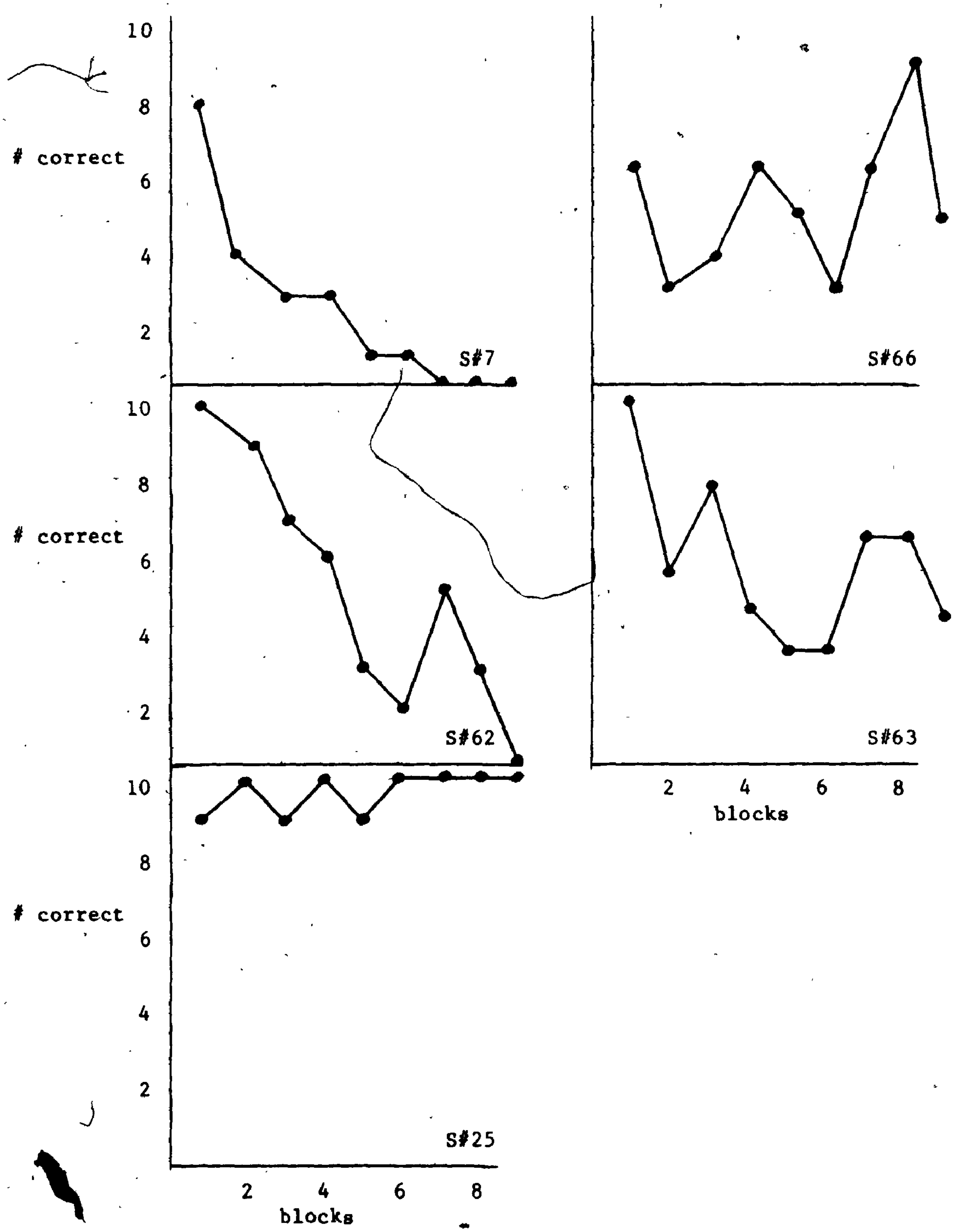


Figure 12: Percentage of Items Answered Correctly During Component Skills Training for Unsatisfactory Tutorial-High Subjects

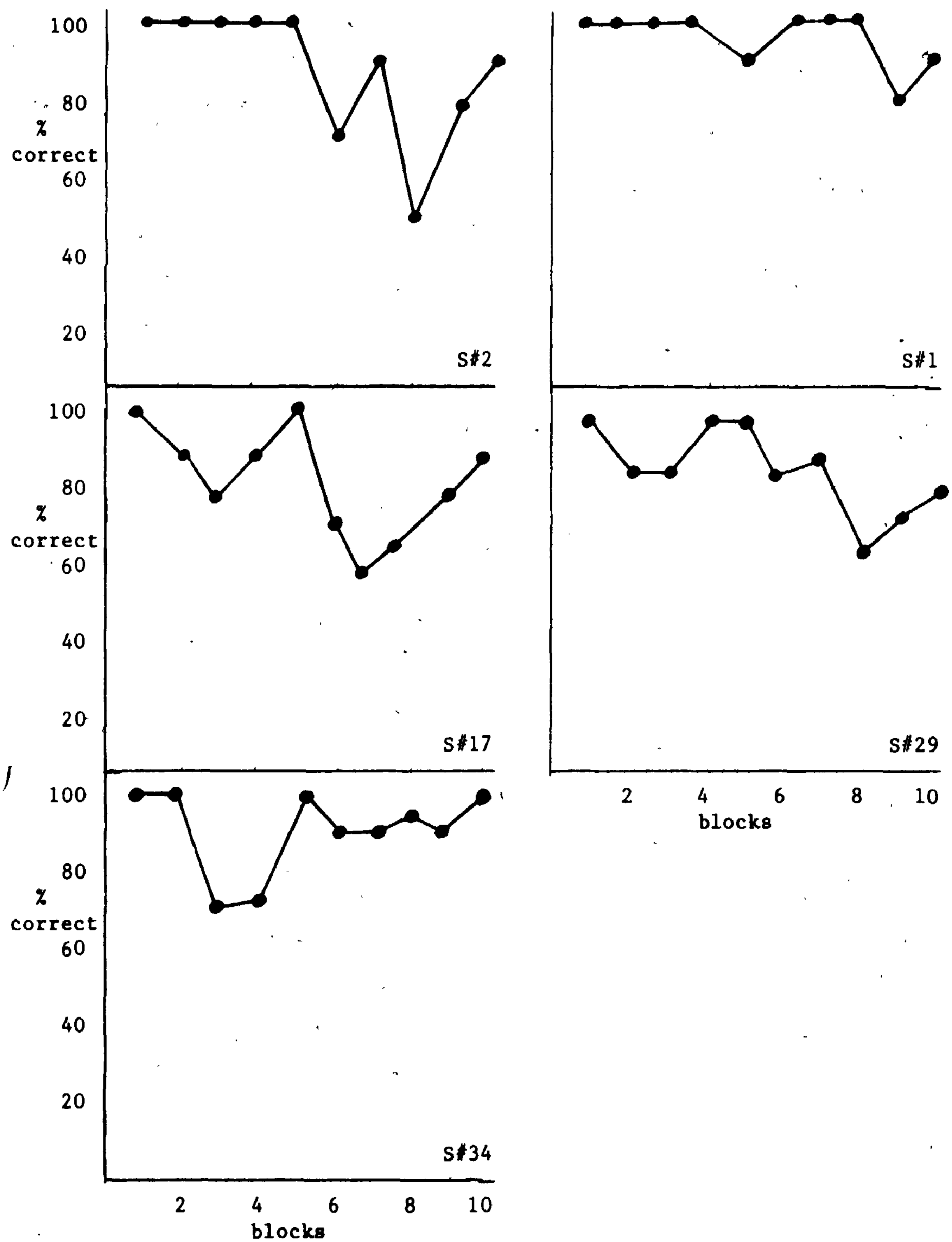


Figure 13: Percentage of Items Answered Correctly During Component Skills Training for Unsatisfactory Tutorial-Low Subjects

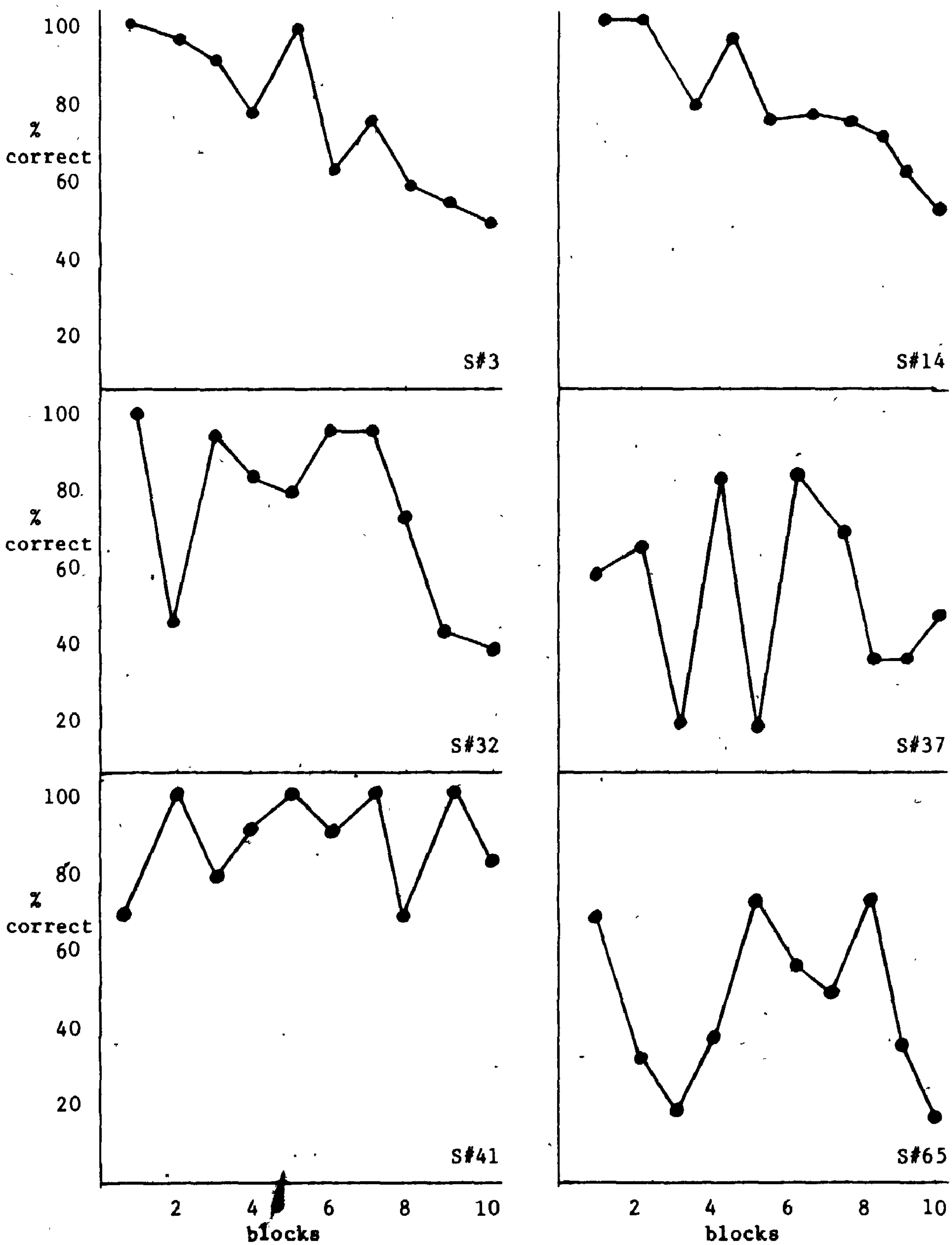


Figure 14: Percentage of Items Answered Correctly During Component Skills Training for Satisfactory Tutorial-High Subjects

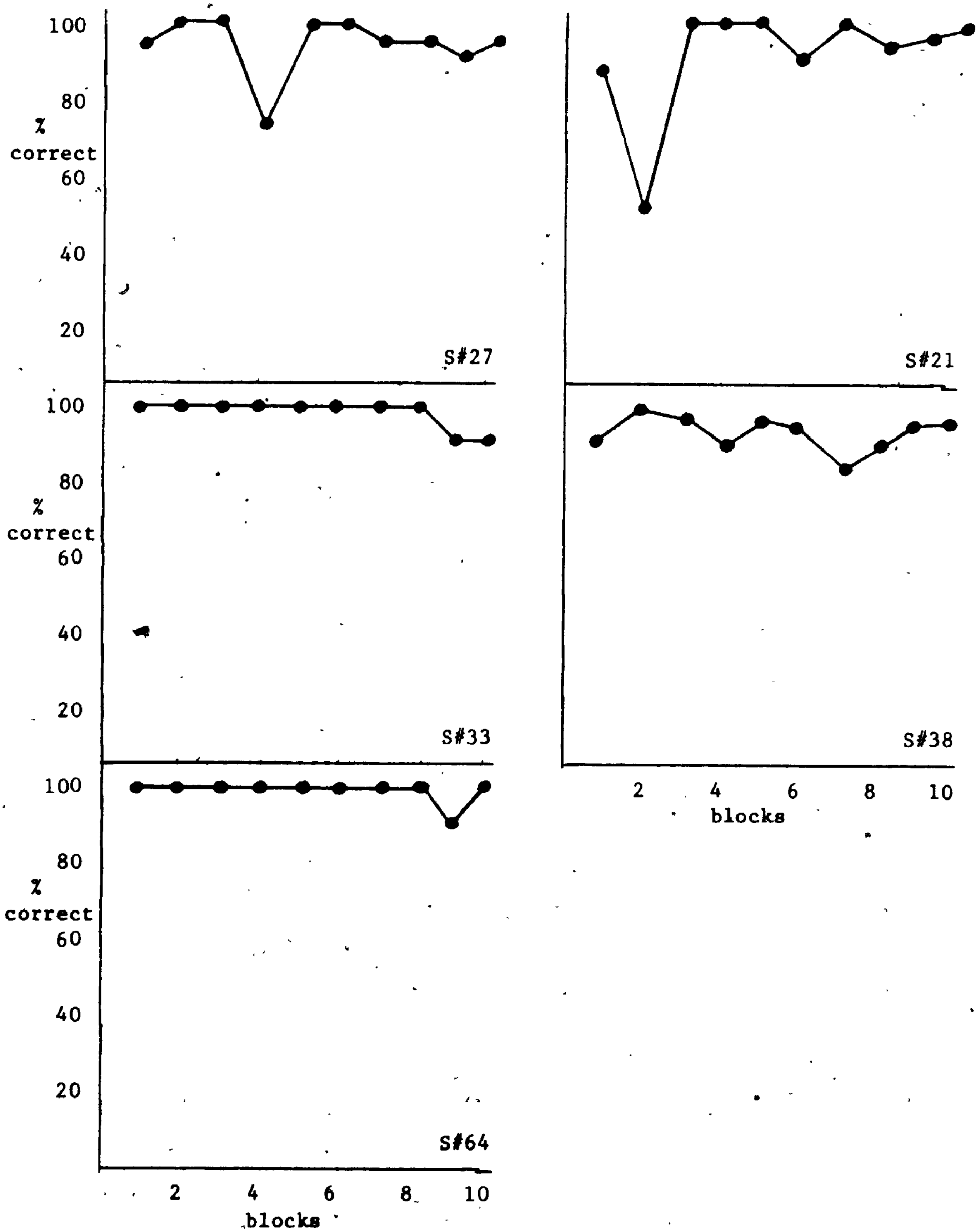
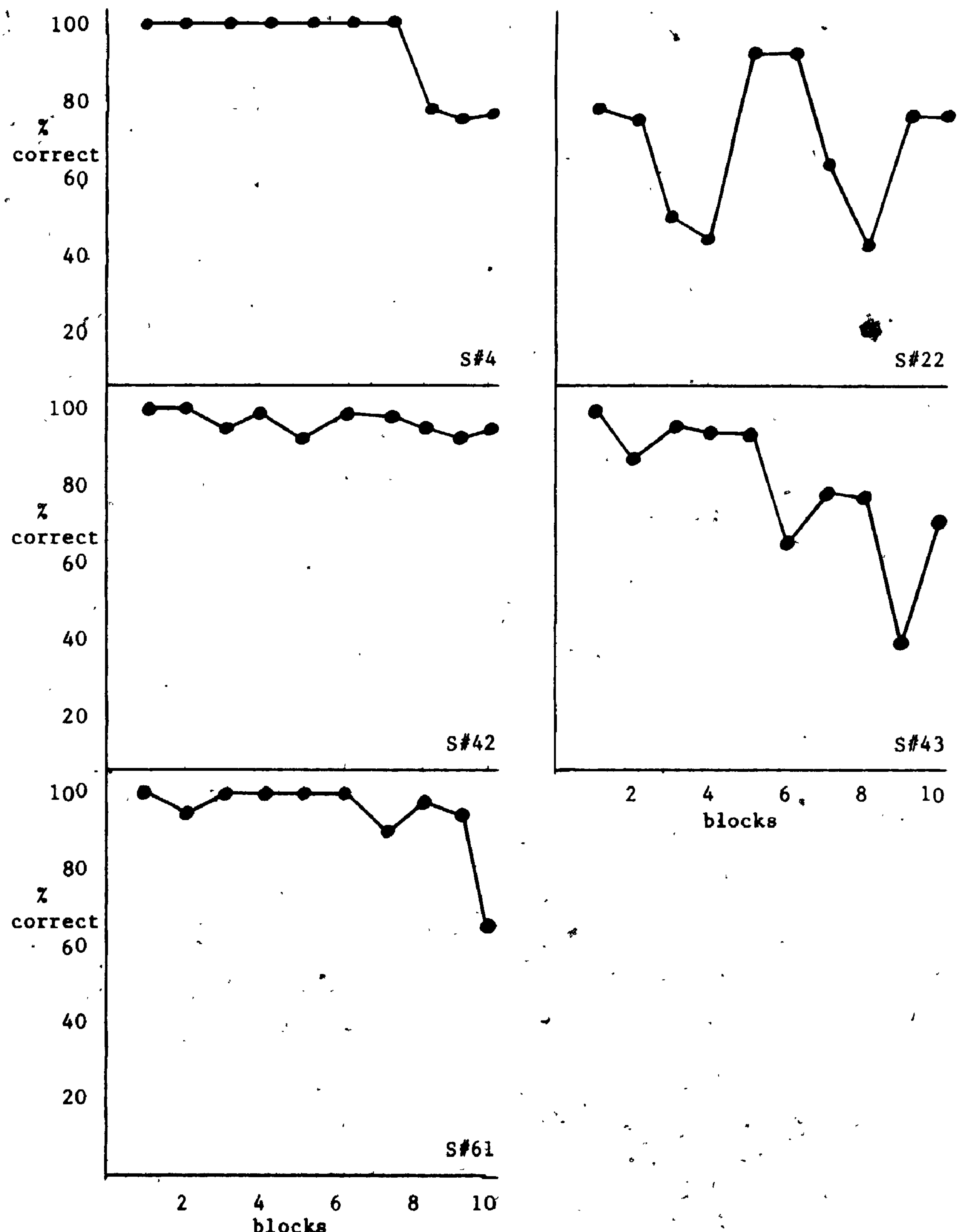


Figure 15: Percentage of Items Answered Correctly During Component Skills Training for Satisfactory Tutorial-Low Subjects



Discussion

Grade 3 children's performance on a series of subtraction problems was examined following training with two different types of remedial programs. The Criterion Training program presented children with subtraction problems and feedback in the form of correctly worked solutions. The Component Skills Training program presented the procedures required for borrowing in a step by step fashion, along with feedback in the form of correct answers.

The subjects were divided into two groups based on pretest performance. Children who solved more than 60 percent of the pretest problems correctly were classified as Unsatisfactory subjects. Those who solved between 60 percent and 80 percent (inclusive) of the problems correctly were classified as Satisfactory subjects. Both Component Skills Training and Criterion Training resulted in increases in the number of borrowing problems solved correctly relative to controls for subjects in the Unsatisfactory condition. These two training conditions appeared to have been successful in reducing errors in the borrowing process and in increasing the frequency with which borrowing procedures were correctly applied. Remedial training had no significant effects, however, on the performance of Satisfactory subjects.

While some control subjects improved in the two week interval following remedial training, it appeared that the performance improvements seen for experimental subjects were maintained during this time. The changes observed in the control subjects appeared the result of a specific classroom related phenomenon which probably did not occur in the other classrooms. It is felt, however, that the remedial programs examined in this study would be most effectively utilized as techniques for elevating performance quickly. Performance gains could subsequently be maintained through contingencies arranged by the classroom teacher. As a teacher implemented strategy, this remediation - maintenance intervention could be carried out easily and with little disruption to class members not requiring such additional assistance.

An unfortunate complication, however, is that some children failed to progress successfully through the remedial programs. Decreases in the number of problems solved correctly were observed both within and between training days, although the pretest and posttest scores of these children were not different. In the Criterion Training Condition examination of the patterns of errors of individuals who did not complete training successfully suggests that: 1) reinforcement used in this study was not sufficient to maintain initially moderate to high levels of accuracy S#5, #23 and #4) or to maintain improvements in performance S#73); 2) subjects did not appear to be utilizing feedback. Error patterns did not change following the presentation of correct solutions, and subjects #28 and #57 failed to even attempt to borrow, although this requirement was shown in the feedback presented.

Performance in Component Skills Training is more difficult to interpret because the items differed in content and were designed to produce a high level of accuracy. The consistently high level of performance seen in the individual subject profiles showed that this goal was achieved for some subjects (Tutorial-High). The performance of others, however (Tutorial-Low), was again characterized by fluctuations in the number of items answered correctly from block to block as well as by decreases throughout the last two booklets of the program. Ineffective reinforcers, or particular subject characteristics may have been responsible for these performance fluctuations. The similarity of Tutorial-Low profiles in both training conditions strongly suggests that similar factors contribute to the patterns observed for subjects who do not train well.

This failure of some children to benefit from training indicates that, if these remedial programs are to be effective in the classroom, the identification of subjects who are likely to fail is crucial. Differences between potentially successful and unsuccessful training candidates were apparent sometime during the second day of training for Criterion Training, and during presentation of the fourth training booklet for Component Skills Training. At this stage of the research, attention to performance while subjects are working through the programs and offering assistance when deterioration is noticed may be the optimal method to encourage successful remediation of all children. Ideally, the identification of those children likely to have difficulty prior to training would allow assignment of these students to programs perhaps more suited to their needs.

During Criterion Training, Unsatisfactory subjects who were successful in training (Tutorial-High) showed marked increases in the number of problems solved correctly to a high of 88 percent. This increase remained fairly stable for the remainder of training. In fact, all but one child in this condition attained 100 percent correct on at least two blocks of problems. So it seems that some Unsatisfactory children were able to use feedback to achieve accurate performance. In fact, this type of treatment was as effective as a more carefully programmed one in improving the performance of these subjects.

In evaluating the relative efficacy of the two training programs, factors such as the relative difficulty of designing the programs, ease of administration, and length of time before acceptable levels of performance on the target skill are reached, need to be considered in addition to the magnitude of treatment gains. Although there are many areas for improvement, the Component Skills Training program required much planning and pilot testing. The subjects involved in this program tended to ask many more clarification questions throughout, and the level of performance on the target skill could not be assessed until the end of day 3 when the program was completed. In the Criterion Training program, high levels of performance were attained by day 2 for many children. They had no difficulty following directions, and the program was easy to construct and implement. The repetitive nature of this program, however, may have resulted in fatigue or boredom which could have undermined its effects for some children and produced the between and within-day decrements discussed above. With some attention to the introduction of novelty and of more powerful reinforcers, the Criterion Training program appears to have good potential as a remedial technique, at least from the standpoint of training time and diagnostic flexibility.

General Discussion

The aim of this thesis was to find a remedial technique which could be used to help children who experience difficulty in subtraction. The technique was judged to be more effective if it could be implemented by a teacher in the regular classroom and if it could yield performance improvements within a short period of time. Two types of training procedures appeared to fit the bill, training in the component skills required for solving subtraction problems and feedback in the form of correctly worked solutions.

When the performance of children whose performance was erratic was compared with that of children who always committed errors, no clear cut differences between the two groups were found. While the errors of the latter group tended to more consistently suggest problems with the procedures required for borrowing, many members of the former group demonstrated similar difficulties. These findings suggest that performance could be understood as falling along a mastery continuum. The consistent inversion errors committed by some children represented failure to invoke the procedures necessary to solve the problems. That is, these children had difficulty at the extreme end of the continuum. It also appeared that when these children were prompted to invoke these procedures, large gaps in their knowledge were present. Other children, while recognizing the conditions where it was necessary to engage in borrowing, also demonstrated gaps in knowledge. These children were located slightly to the right of the first group on the mastery continuum. A third group of children made procedural errors but also solved problems correctly. This group was even further to the right on this continuum, but tended to move around a good deal. When tested at a later point in time, the performance of some of these children was consistently accurate, while other children failed to attempt any of the borrowing procedures. Lastly, some children almost always solved problems correctly and were located at the extreme right on the mastery continuum.

In terms of a cognitive model of subtaction performance, these studies were not designed to evaluate the contributions of particular components to processing competency. I would, however, like to address the developmental progression issue. First of all, the contribution of short term memory improvements (or of individual differences in short term memory) to the type of performance examined here is difficult to assess. The decrements in performance during training of Tutorial-Low subjects and their recovery between days, however, suggests that short term memory failure is not responsible for poor training performance of these children. For Tutorial-High children, the apparent gaps in their knowledge were quite easily remediated using techniques which would not be expected to address aspects of short term memory functioning (for example, information chunking).

I suggest that, in addition to the developmental progression proposed for the processing of mental arithmetic problems (Ashcraft, 1982), and the contribution made by short term memory changes to this development (Brainerd, 1981), a second developmental progression occurs which begins at the point where children appear to rely more on stored representations of number facts for the solutions to simple problems in mental arithmetic. This progression would deal with children's developing ability to solve problems requiring the application of a large body of rules and heuristics. By the time children are introduced to the process of subtraction with borrowing, their short term memory capacity would be expected to be reaching its upper limits (Brainerd, 1981). Further improvements in short term memory would be the result of more efficient processing techniques, such as chunking and imagery representation. When introduced to the complexities of borrowing, children are faced with two tasks. They must learn to apply the procedures correctly and they must be able to store and retrieve these procedures in the correct sequence. As in the case of mental arithmetic, I suggest that the development of the use of procedures preceeds a stored representation of these procedures.

In the first stage of learning how to use the procedures, children are at the extreme left on the mastery continuum. The consistency of their errors reflect gaps in their knowledge of these procedures. While they practice and receive additional instruction in these procedures, their performance takes on a more erratic quality and they move to the right on the mastery continuum. Errors are made, but from time to time problems are also solved correctly. It is at this point that improvements in organization of information occur to facilitate processing in short term memory and subsequent storage and retrieval. At this stage, the frequency of procedural errors decreases, and factors which interfere with processing or with information retrieval are primarily responsible for errors. These factors include fatigue, boredom, behavioral styles, incompatible behavior, etc.

If this outline represents the development of children's ability to solve these types of problems, three different types of treatments would be expected to achieve different results for different groups of children. For those children at the procedure mastery level, interventions which provide them with information concerning their performance should result in improvements. Both feedback and instruction in component skills should fulfill this function. For children at the information organization stage, interventions which facilitate this process such as chunking strategies, mnemonic aids, cue cards, etc., would be expected to be more helpful. The lack of these types of interventions in Study 3 may explain the failure of Satisfactory subjects to demonstrate performance gains. A third type of intervention would be intended for those children whose errors are the result of factors which interfere with processing at some level. These factors could prevent the learning of rules at the procedure mastery stage, or could interfere with the execution of these procedures at later stages of mastery. Contingency management, introduction of novelty, or interventions focused on undesirable behavior are possibilities. Again, failure of some children to benefit from training offered in Study 3 may be the result of failure to include these types of interventions, and the failure to identify the children for which these techniques would be appropriate.

It is of course recognized that without the necessary longitudinal (and cross sectional) data, this developmental progression and implications for remediation remain hypothetical. Further investigations might take the following course.

A longitudinal study of children's solutions to subtraction problems with borrowing could provide information regarding the development of mastery. Examination of errors and speed of performance might be especially useful in plotting changes in competency. Single subject experimental designs might be used to examine the effects of different types of teaching strategies at different stages in mastery development. This approach could be used to examine the development of skill in borrowing, discussed earlier. Strategies used to promote one kind of development would only be expected to facilitate performance at that particular point in development. Also, an examination of characteristics of those children who have difficulty in training and of those whose development fails to progress past a particular point may suggest appropriate intervention strategies or remedial techniques that could be implemented at early stages of learning. Both remedial and preventative applications would be of interest.

It is understood that the failure of some children to successfully complete the remedial programs could simply reflect inadequacies in program design. Sidman and Stoddard (1966 and 1967), with retarded, nonverbal children found that although some children made errors on a circle/elipse discrimination task, they learned reasonable things. Children appeared to be attentive to features of the stimulus materials which were not critical for the discriminations the experimenter wanted these children to make. The children's behavior suggested that they continued to attend to these "irrelevant" features, obtaining intermittent reinforcement pretty much by chance. The authors pointed out that, under the circumstances, the contingency structure promoted and fostered this type of behavior. While the children committed errors, from the experimenter's point of view, subjects' behavior was consistent with learning principles and they were in fact learning. They were not, however, learning what the experimenter wanted them to.

Modifications in the instructional sequence resulted in better experimental control over these children's behavior.

In Study 3, it also appeared that the contingencies, at least for some children, were not associated with correct responses. Children often did not appear to attend to the feedback provided, initially high rates of correct responding were not maintained, and in some cases children appeared to have been responding in a random fashion. In short, as with Sidman and Stoddard's subjects, the behavior of some children did not appear to be controlled by the contingencies desired by the experimenter.

It may be that problems in program design interacted with certain characteristics of the children who composed the Tutorial-Low group. The programs designed in Study 3 did not ensure that the desired target behavior was the only behavior that was reinforced and that children attended to the essential aspects of stimuli and feedback. The programs did not offer re-instruction immediately following errors, and the programs did not allow children to progress to the next step only after a specified performance criterion had been reached. For children whose behavior is not reinforced by knowledge of performance, or for those who are particularly subject to the effects of fatigue or boredom, these problems in program design may have resulted in the observed deterioration and low levels of performance. For the purpose of effective remediation it would be desirable to eliminate individual differences in children's responses. Careful attention to the control of specific behaviors required for learning seems required. That is, the remedy may involve a "tightening up" of these programs in terms of the application of contingencies and immediate awareness of errors. Or, the remedy may involve the addition of components which could assist in the development of more efficient information processing.

Because of its generally unprogrammed nature, it was surprising that Criterion Training was, on its own, so effective for some children. Referring to the sequencing of instructional components, Skinner (1968) made the following observation:

> The logical structure of a subject matter is not always relevant. There may be many reasons why programs should be designed on logical principles, but they are not always logical reasons. The well known case system, for example, is recommended because specific instances are usually easier to teach and remember, and are inherently more reinforcing than general principles; it is nevertheless often easier to teach precept than practice, or rule than example, and the specific-general order is then reversed. A logical order is not the order in which most behavior is acquired and is therefore not necessarily the best order in which it is to be taught (p.222).

In line with Skinner's suggestion, it appears that at least some children could extract the information needed to solve borrowing problems from the specific examples presented. That is, the specific-general order seemed to work in this case, and "logical" sequencing of material did not seem to offer any advantages. Whether those who benefitted from Component Skills Training would also have benefitted from Criterion Training is not clear. It may be that Tutorial-Low children in the Component Skills program would have had more success in the Criterion Training program and vice versa. That is an empirical question which may not be answered until more is known about why some children succeed on these programs and others do not.

At any rate, the difficulties encountered in trying to eliminate children's errors were unexpected. This is perhaps the most important result of this research, in that it suggests a simple problem which has been shown to be easily remediable (by studies cited in Study #2 and Study #3) is not. At least, techniques like instructions and promise of reward, and carefully programmed instruction and feedback were not effective for everyone. Along with even more careful attention to the technology of teaching, incorporation of findings concerning children's cognitive and skill development and concerning individual differences in these developmental areas may be necessary. This suggests that applied researchers would be well advised to incorporate and utilize more of the information available from the cognitive laboratory when developing behavior change methodology. Traditional teaching and remediation may not be effective for all children until researchers understand not only how to control behavior but why certain classes of behavior appear when they do. With the advancement of legislation to protect the rights of exceptional children, the next few decades should see the focus shift from the development of methods and techniques to attempts to understand differential responsiveness of individual children to treatment variables.

Appendix I

Statistical Analyses

In Study 2 and Study 3, subjects in each experimental condition were given a pretest prior to treatment. Following treatment, subjects were retested on one or two occasions. Those in the control condition were tested on the same schedule but were not exposed to the experimental conditions.

Three types of analyses have been applied to this type of design in the literature. The first type is a repeated measures analysis of variance where the pretest--posttest factor is the repeated measure, commonly labeled "trials". The between subject factor is experimental condition. The second type of analysis involves the use of gain scores. In this case, the dependent measure is used to reflect the amount of improvement occurring in each subject from pretest to posttest. This score is calculated by subtracting the pretest score from the posttest score. The gain scores are then analyzed using a between subjects analysis of variance. The third type of analysis is an analysis of covariance. Here, posttest scores are adjusted on the basis of their regression on the pretest scores. The only difference between the last two analyses is that in the case of gain scores, the slope of the within treatment regression line is arbitrarily set at 1, while in the case of ANCOVA the slope is determined by the data.

For a pre-post design where the subjects were randomly assigned to treatments and where the treatments did not affect the covariate, ANCOVA was selected as the analysis of choice. The reasons for this selection are summarized below.

Huck and McLean (1975) argue that, in the case of the pre-post ANOVA, the linear model which this analysis assumes is incorrect. Both treatment and interaction effects do not have any effect on pretest scores. Therefore, posttest scores cannot be thought of as having an interaction component. By including the variance associated with this interaction component, the F test for the main effect of treatment is "spread out" over pre and posttests. In this case, the interaction actually becomes the measure of the main effect. In addition to the resulting loss of power, the inclusion of a "trials" factor does not appear to be meaningful. A main effect of trials does not address the issue of treatment effects and is open to confounding by regression, maturation, history, etc.

Although the use of gain scores eliminates the trials factor and the interaction component from the linear model, this analysis normally produces a less sensitive test of treatment effects than does a covariance analysis. This is because the slope of within class regressions is usually not equal to 1. In these cases, arbitrarily setting the slope equal to 1 results in greater experimental error than would be the case if the slope was determined on the basis of the data (Huck and McLean, 1975). The use of ANCOVA would result in a more sensitive test of the treatment effects by removing variations in the covariate from experimental error (Cochran, 1957).

Overall and Woodward (1977a) suggest that excess caution exercised in the literature with regard to the use of ANCOVA is unwarranted. They argue from a theoretical viewpoint as well as from the results of Monte Carlo studies (Overall and Woodward, 1977b) that ANCOVA can give unbiased results of treatment effects when groups differ in mean scores on the covariate, even in cases where subjects were assigned to groups on the basis of the observed covariate scores. They also showed that unbiased estimates of treatment effects can be obtained in cases where the covariate is not perfectly reliable and when within and between group regressions are not equal. These conclusions refer to situations where treatments do not affect the covariate.

In the two studies described in this dissertation, treatment effects had no effect on the covariate (pretest scores) and Unsatisfactory and Satisfactory subjects were randomly assigned to treatments. That means that differences between groups on the covariate were a result of chance sampling variability and, as such, pose no problem for ANCOVA (Overall and Woodward, 1977a). It appears, therefore, that analysis of covariance would result in a more sensitive and theoretically defensible test of treatment effects than either a gain score analysis or use of a repeated measures analysis of variance.

References

Ashcraft, M.H. and Battaglia, J. Cognitive arithmetic: Evidence for retrieval and decision procedures in mental addition. Journal of Experimental Psychology: Human Learning and Memory, 1978, 4, 527-538.

Ashcraft, M.H. and Fireman, B.A. Mental addition in third, fourth and sixth grades. Journal of Experimental Child Psychology, 1982, 33, 216-234.

Ashcraft, M.H. and Stazyk, E.H. Mental addition: A test of three verification models. Memory and Cognition, 1981, 9,, 185-196.

Baxter, M.M. Prediction of error and error type in computations of sixth grade mathematics students. Doctoral Dissertation, Penn State University, 1973

Blankenship, C.S. Reduction of systematic inversion errors in the subtraction algorithm. Doctoral Dissertation, University of Washington, 1976.

Brainerd, C.J. Young children's mental arithmetic errors: A working memory analysis. Research Bulletin #547, Dept. of Psychology, University of Western Ontario, 1981.

Broughton, S.F. and Lahey, B.B. Direct and collateral effects of positive reinforcement, response cost and mixed contingencies for academic performance. Journal of School Psychology, 1978, 16, 126-136.

Brown, J.S. and Burton, R.R. Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 1978, 2, 155-192.

Brueckner, L.J. and Bond, G.L. The diagnosis and treatment of learning difficulties. NY: Appleton Century Crofts, 1955.

Burrows, J.K. A review of the literature on computational errors with whole numbers. Math Education Diagnostic and Instructional Centre (MEDIC), 1976.

Carpenter, T.P. The effect of instruction on first grade children's initial solution procedures for basic addition and subtraction problems. Paper presented at the annual meeting for the American Education Research Association, Boston, 1980.

Chadwick, B.A. and Day, R.C. Systematic reinforcement: Academic performance of underachieving students. Journal of Applied Behavior Analysis, 1971, 4, 311-319.

Chronbach, L.J. and Firby, L. How should we measure "change" - or should we? Psychological Bulletin, 1970, 74, 68-80.

Cochran, W.G. Analysis of covariance: Its nature and use. Biometrics, 1957, 13, 261-281.

Conlon, M.F., Hall, C. and Hanley, E.M. The effects of a peer correction procedure on the arithmetic accuracy for two elementary school children. In G. Semb and H.W. Hopkins (eds) A new direction for education: Behavior Analysis 1972. Lawrence, Kansas: University of Kansas, 1973.

Copeland, R.E., Brown, R.E. and Hall, R.V. The effects of principal implemented techniques on the behaviour of pupils. Journal of Applied Behavior Analysis, 1974, 7, 77-86.

Cox, L.S. Analysis, classification and frequency of systematic error computational patterns in the addition, subtraction, multiplication and division vertical algorithms for grades 2-6 and special education classes. Kansas City Medical Centre, 1974.

Ellis, L.C. A diagnostic study of whole number computations of certain elementary students. Doctoral Dissertation, Louisiana State University and Agricultural and Mechanical College, 1972.

Engelhardt, J.M. Analysis of children's computational errors: A qualitative approach. British Journal of Educational Psychology, 1977, 47, 149-154.

Ferritor, D.E., Buckholdt, D., Hamblin, R.L. and Smith, L. The noneffects of contingent reinforcement for attending behavior on work accomplished. Journal of Applied Behavior Analysis, 1972, 5, 7-17.

Fink, W.T. and Carnine, D.W. Control of arithmetic errors using informational feedback and graphing. Journal of Applied Behavior Analysis, 1975, 8, 461.

Fuson, K.C. Towards a model for the teaching of mathematics as a goal directed activity. In K.C. Fuson and W.E. Geeslin (eds) Explorations in the modeling of the learning of mathematics. Centre for Science and Mathematics Education, Ohio State University, 1979.

Gagne, R.M. and Brown, L.T. Some factors in the programming of conceptual learning. Journal of Experimental Psychology, 1961, 62, 313-321.

Gelman, R. and Gallistel, C.R. The child's understanding of number. Cambridge: Harvard University Press, 1978.

Gelman, R. and Tucker, M.F. Further investigations of the young child's conception of number. Child Development, 1975, 46, 167-175.

Ginsburg, H. Children's arithmetic: The learning process. Toronto: Van Nostrand Co., 1977.

Graeber, A.D. and Wallace, L. Identification of systematic errors: Final report. Research for better schools inc., Philadelphia, 1977. National Institute for Education (DHEW), Washington, DC.

Groen, G.J. and Parkman, J.M. A chronometric analysis of simple addition. Psychological Review, 1972, 79, 329-343.

Groen, G.J. and Poll, M. Subtraction and the solution of open sentence problems. Journal of Experimental Child Psychology, 1973, 16, 292-302.

Harris, V.W. and Sherman, J.A. Effects of peer tutoring and consequences on the mathematics performance of elementary classroom students. Journal of Applied Behavior Analysis, 1973, 6, 587-597.

Harvey and Kyte. Zero difficulties in multiplication. The Arithmetic Teacher, 1965, 12, 45-50.

Holland, J. Research on programming variables. In R. Glaser (ed) Teaching machines and programmed learning II, data and directions. US: National Education Association, 1965.

Horowicz, L.M. Elements of Statistics for Psychology and Education. New York: McGraw Hill, 1974.

Huck, S.W. and McLean, R.A. Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. Psychological Bulletin, 1975 82, 511-518.

Kirby, F.D. and Shields, F. Modification of arithmetic response rate and attending behavior in a seventh grade student. Journal of Applied Behavior Analysis, 1972, 5, 79-84.

Logan, H.L. Study of the effects of examples, non examples and a mix of the two on computational skills in subtraction for systematic and non-systematic error groups among fourth grade children., Doctoral dissertation, U of Iowa, 1976.

Lovitt, T.C. and Esveldt, K.A. The relative effects on math performance of single versus multiple ratio schedules. A case study. Journal of Applied Behavior Analysis, 1970, 3, 261-270.

Marholin, D. and Steinman, W.M. Stimulus control in the classroom as a function of the behavior reinforced. Journal of Applied Behavior Analysis, 1977, 10, 465-478.

Mick, H.W. and Brazier, G.D. A logical-cognitive view of learning and teaching mathematical concepts. In K.C. Fuson and W.E. Geeslin (eds) Explorations in the modeling of the learning of mathematics. Centre for Science and Mathematics Education, Ohio State University, 1979.

Mierkiewicz, D.B. Instructional and theoretical implications of a mathematical model of cognitive development. In K.C. Fuson and W.E. Geeslin (eds) Explorations in the modeling of the learning of mathematics. Centre for Science and Mathematics Education, Ohio State University, 1979.

Morton, R.L. An analysis of errors in the solution of arithmetic problems. Educational Research Bulletin, 1925, 4, 187-190.

Myers, J. Fundamentals of experimental design. Toronto: Allyn and Bacon, Inc., 1979.

Overall, J.E. and Woodward, J.A. Nonrandom assignment and the analysis of covariance. Psychological Bulletin, 1977a, 84, 588-594.

Overall, J.E. and Woodward, J.A. Common misconceptions regarding the analysis of covariance. *Journal of Multivariate Behavioral Research*, 1977b, *12*, 171-185.

Parsons, J.A. The reciprocal modification of arithmetic behavior and program development. In G. Semb and W.L. Hopkins (eds) *A new direction for education: Behavior analysis 1972*. Lawrence, Kansas. University of Kansas, 1973.

Post, T.R. and Reys, R.E. Abstraction, generalization and the design of mathematical experiences for children. In K.C. Fuson and W.E. Geeslin (eds) *Explorations in the modeling of the learning of mathematics*. Centre for Science and Mathematics Education, Ohio State University, 1979.

Reiss, A. An analysis of children's number responses. *Harvard Educational Review*, 1943, *13*, 149-162.

Sagotsky, G., Patterson, C.J. and Lepper, M.R. Training children's self control: A field experiment in self monitoring and goal setting in the classroom. *Journal of Experimental Child Psychology*, 1978, *25*, 242-253.

Sidman, M. and Stoddard, L.T. The effectiveness of fading in programming a simultaneous form discrimination for retarded children. *Journal of the Experimental Analysis of Behavior*, 1967, *10*, 3-15.

Sidman, M. and Stoddard, L.T. Programming perception and learning in retarded children. *International Review of Research in Mental Retardation*. 1966, *2*, 151-208.

Skinner, B.F. *The technology of teaching*. New Jersey: Prentice Hall Inc., 1968.

Smith, C.W., Jr. A study of constant errors in subtraction and in the application of selected principles in the decimal numeration system made by third and fourth grade students. Doctoral Dissertation, Wayne State University, 1968.

Smith, D.D., Lovitt, T.C. and Kidder, J.D. Using reinforcement contingencies and teaching aids to alter subtraction performance of children with learning disabilities.
In G. Semb and W.L. Hopkins (eds) A new direction for education: Behavior analysis 1972. Lawrence, Kansas: University of Kansas, 1973.

Sternberg, S. High speed scanning in human memory. Science, 1966, 153, 652-654.

Sternberg, S. The discovery of processing stages. Extensions of Donders' method. In W. Koster (ed) Attention and Performance II. Acta Psychologicia, 1969, 30, 276-315.

Usiskin, Z. Uses of the fundamental operations. Chicago University, Dept. of Education, National Science Foundation, Washington, DC, 1974.

Welkowitz, J., Ewen, R.B. and Cowen, J. Introductory Statistics for the Behavioral Sciences. NY: Academic Press, 1976.

Williams, C.L. and Whitaker, R.L. Diagnosis of arithmetic difficulties. Elementary School Journal, 1937, 37, 592-600.

Winer, B.J. Statistical principles in experimental design. NY: McGraw Hill, 1971.

Woods, S.S., Resnick, L.B. and Groen, G.J. An experimental test of five process models for subtraction. Journal of Educational Psychology, 1975, 67, 17-21.

Young, R.M. and O'Shea, T. Errors in children's subtraction. Cognitive Science, 1981, 5, 153-177.

END

06-09-84

FIN