DOCUMENT RESUME

ED 199 060 SE 034 282

AUTHOR	Greeno, James G.; And Others
TITLE	Individual Differences and Selective Processes in Cognitive Procedures.
INSTITUTION	Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	LRDC-1979/15
PUB DATE	79
NOTE	56p.: Not available in hard copy due to small print throughout entire document.
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Abstract Reasoning; Algebra; Arithmetic; *Cognitive Processes; Educational Research; Flow Charts; Individual Differences; *Learning Theories; *Mathematical Concepts; *Mathem... Education; Models; *Problem Solving; Second... ...ication; Secondary School Mathematics
IDENTIFIERS	*Mathematics Education Research

ABSTRACT

Problems were presented to 13 university students in which letters were to be added or subtracted (e.g., B + D = ? or F - ? = D). After each problem, each subject gave a retrospective protocol indicating the way in which the problem was solved. Models of performance by each subject in each experimental session shared major properties: choices by all subjects depended on a few features of the problems. Individual differences consisted in the features that were included in a subject's choice procedure and the criteria applied in choices based on the features that were used. (Author)

LEARNING RESEARCH AND DEVELOPMENT CENTER

INDIVIDUAL DIFFERENCES AND SELECTIVE PROCESSES IN COGNITIVE PROCEDURES

1979/15

JAMES G. GREENO, JOHN G. GREENO, AND CHARLES A. PERFETTI

University of Pittsburgh

INDIVIDUAL DIFFERENCES AND SELECTIVE PROCESSES IN COGNITIVE PROCEDURES

James G. Greeno, John G. Greeno, and Charles A. Perfetti

Learning Research and Development Center

University of Pittsburgh

1979

The research reported herein was supported by the Learning Research and Development Center, supported in part by funds from the National Institute of Education (NIE), United States Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of NIE, and no official endorsement should be inferred.

Abstract

Problems were presented to 13 subjects in which letters were to be added or subtracted (e.g., B + D = ? or F - ? = D). After each problem, each subject gave a retrospective protocol indicating the way in which the problem was solved. Models of performance by each subject in each experimental session shared major properties; choices by all subjects depended on a few features of the problems. Individual differences consisted in the features that were included in a subject's choice procedure and the criteria applied in choices based on the features that were used.

INDIVIDUAL DIFFERENCES AND SELECTIVE PROCESSES IN COGNITIVE PROCEDURES

James G. Greeno, John G. Greeno, and Charles A. Perfetti

Learning Research and Development Center
University of Pittsburgh

Many analyses of cognitive procedures are composite models, based on trends that are observed in performance of groups of subjects. Occasionally, models have been fit to data of individual subjects, using trends that are observed on a heterogeneous set of problems. In these studies, various models are formulated, representing differing ways in which a task can be performed. The models imply different patterns of difficulty in the set of problems given to subjects, and data in the form of latencies or frequencies of errors are used to select the model that best fits the data that are obtained. However, in relatively complex tasks it seems likely that individuals perform according to different procedures, or that a single individual uses different procedures to solve different problems. To investigate differences between procedures used by different individuals, or by a single individual on different problems, it is helpful to obtain more detailed observations, such as thinking-aloud protocols, that make it possible to identify the procedure used by each subject on each problem.

Introduction

The task that we examined involves arithmetic problems presented using letters of the alphabet. Processes involved in doing arithmetic the ordinary way, with numbers, have been studied extensively. In several studies, measurements of latency have been the basis for concluding that simple addition and subtraction problems are solved by a variety of procedures based on counting. For simple addition problems of the form $m + n = ?$, the model that agrees best with latency data is one in which a counter is initiated at the value of the larger addend, and then incremented the number of times indicated by the smaller addend. This model implies that latency should be approximately a linear function of the smaller addend, and this agrees with data obtained with adult subjects and fifth-grade subjects (Groen & Parkman, 1972) and with some preschool subjects who were taught to count both addends, but who apparently invented the procedure of counting up from the larger addend during a series of practice problems (Groen & Resnick, 1977). For simple subtraction problems of the form $s - n = ?$, a mixture of procedures is apparently used. According to one model, a counter is initiated at n, then incremented the number of times needed to reach s, and the answer is the number of increments that occurred. According to a second model, the counter is initiated at s, then decremented a number of times equal to n, and the answer is the value reached by the counter. (An alternative is to start the counter at a trial value a few numbers below s, count up to s, and if the number of increments equals n, report the trial value as the answer, otherwise adjust the trial value appropriately.) The first model implies that latency should be approximately a linear function of the differences between s and n; the second model implies that latency should be approximately a linear function of n. Data obtained with second- and fourth-grade subjects were fit best by a regression line where latency is a function of the smaller of n and $s - n$; that is, for most subjects the procedure that was used for a given problem was the one that involved the smaller number of increments or decrements (Woods, Resnick, & Groen, 1975). In Groen and Poll's (1973) study of third-grade subjects solving missing addend problems, latencies on problems of the form $m + ? = s$ were consistent with the findings for simple subtraction problems. Latencies on problems of the form ? + n = s were not consistent with any simple model.

The letter-arithmetic task that we studied is analogous to ordinary arithmetic. For example, for the problem B + D = ? the answer is F, which is two letters (B) beyond D. For the problem G - ? = D, the answer is C. (D is three letters [C] before G.) This task can be done using a variety of procedures and the general approach used for a problem can be reported by a subject. For example, after solving B + D = ?, a subject can easily report whether the solution was obtained: (a) by translating B to 2 and D to 4, finding the numerical sum, 6, and then translating back to a letter; or (b) by translating B to 2 and counting up that number of times in the alphabet, starting with D; or by some other method. On the other hand, although letter-arithmetic problems are complex enough to be performed by differing procedures that seem available to introspection, they are simple enough to allow quite a large number of problems to be performed by individual subjects. This permitted us to obtain enough data from each subject to allow inferences about the procedures used by each subject, and for subjects who used a variety of procedures, to identify conditions in problems that influenced selection of procedures for the various problems.

Our decision to study the letter-arithmetic task was motivated partly by an interest in a more detailed study of processes of solving simple arithmetic problems. Difficulties of obtaining thinking-aloud protocols from young children made it seem desirable to use a task in which adult subjects could provide useable data. It seems a reasonable conjecture, supported by results cited above, that children's solutions of addition and subtraction problems depend on their knowledge of the ordered sequence of numbers. Letter-arithmetic might provide an analogous task for adults, since our knowledge of the alphabetic sequence is strong, but we have not acquired an elaborate set of relations on the alphabetic sequence as we have on the numeric sequence. We also hoped that analysis of processes of solving a variety of problems, including addition and subtraction sentences with unknown terms in all three positions, might shed some light on the process of comprehension involved in selection procedures to solve problems requiring different operations.

Arithmetic Methods

There are several possible ways of solving problems of addition and subtraction with letters. The most obvious method, for someone who knows ordinary arithmetic well, is to translate each letter of a problem into its corresponding number, perform the operation of addition or subtraction that is called for in the problem, and finally translate the result back into its corresponding letter, which is the answer. This is shown as a flow chart labeled Method I in Figure 1 for problems requiring addition, and in Figure 2 for problems requiring subtraction. In these and later flow chart representations the input presumes a notation that varies among problem formats. For the methods shown in Figure 1, m is the first addend given in an addition problem presented with a plus sign, but if the problem has a minus sign with the first term unknown, m is the term to the right of the equal sign. For example, the problem B + E = ? would be encoded with m as B and n as E; however, the problem ? - B = E would be encoded with m as E and n as B. In problems requiring subtraction, the notation must specify which term is to be subtracted from the other and the larger term is always encoded as s in the notation of Figure 2. Another notational matter concerning Figures 1 and 2 is the use of arrows in the boxes representing operations. The arrows denote assignment operations in which a variable is given a value. For example, the fourth box of Method I represents the operation of adding the numbers x and y and assigning the result as the value of z.

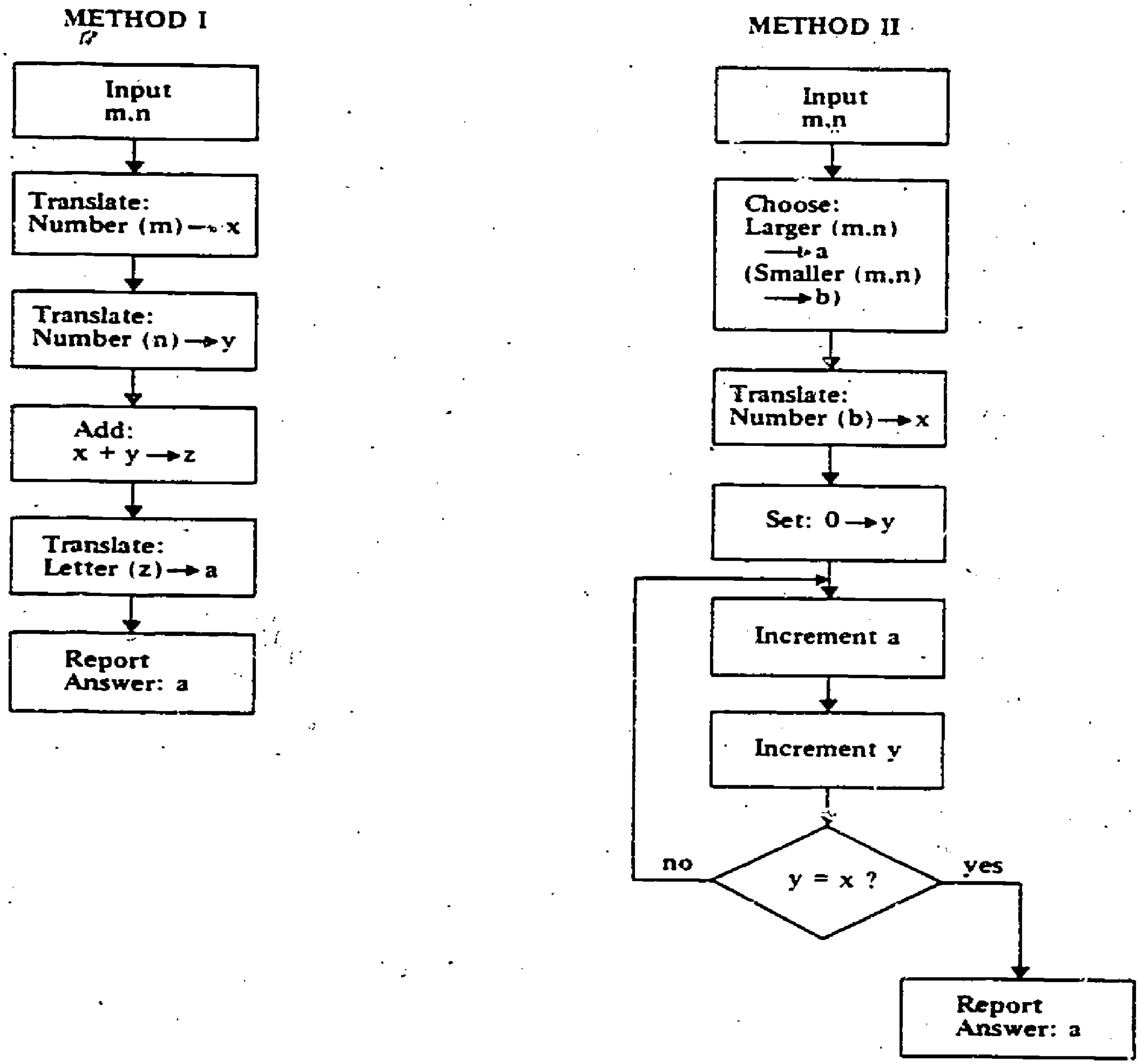


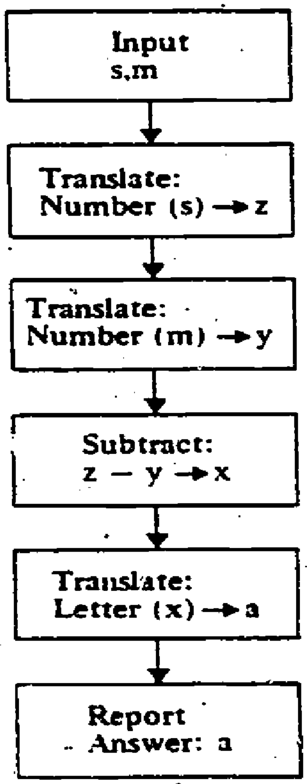
Figure 1. Methods for solving problems: $m + n = ?$ and $? - n = m$.

Translation operations are represented as functions. For example, the second box of Method I represents the operation of finding the number that corresponds to m and assigning that number as the value of x. The operations of adding and subtracting numbers are left unanalyzed here. For adult subjects, these probably involve relatively direct retrieval of stored information, although the time required for retrieval seems to depend on the sizes of the numbers involved in the retrieval (Groen & Parkman, 1972).

Method II in Figure 1 solves problems requiring addition by counting up from the larger of two addends. The main part of this processor is a counter that can be initiated at any letter and incremented to give the next letter in the alphabet. This is done in conjunction with a numeric counter so the number of increments is known. Suppose the problem is E + B = ? . E and B are encoded as the values of m and n, respectively. The larger addend, E, is identified; the smaller addend, B, is translated to its corresponding number, 2. The alphabetic counter is initialized at E, and the numeric counter at zero. Then both are incremented until the numeric counter reaches 2, and the letter that was reached is the answer. Subjects often use their fingers in using this method; for example, saying "E, F (first finger), G (second finger), G!"

A closely related method for subtraction problems is shown as Method II in Figure 2. Here the counter is used to find the length of the interval separating two letters. The letter that is to be subtracted is identified and is used to initialize the alphabetic counter. The numeric counter is initialized at zero. Then the counters are incremented until the alphabetic counter has reached the other letter in the problem. The number of increments that occurred is translated into a letter, which is the answer. For example, if the problem is E - B = ? , the alphabetic counter starts at B, the numeric counter starts at

METHOD I



METHOD II

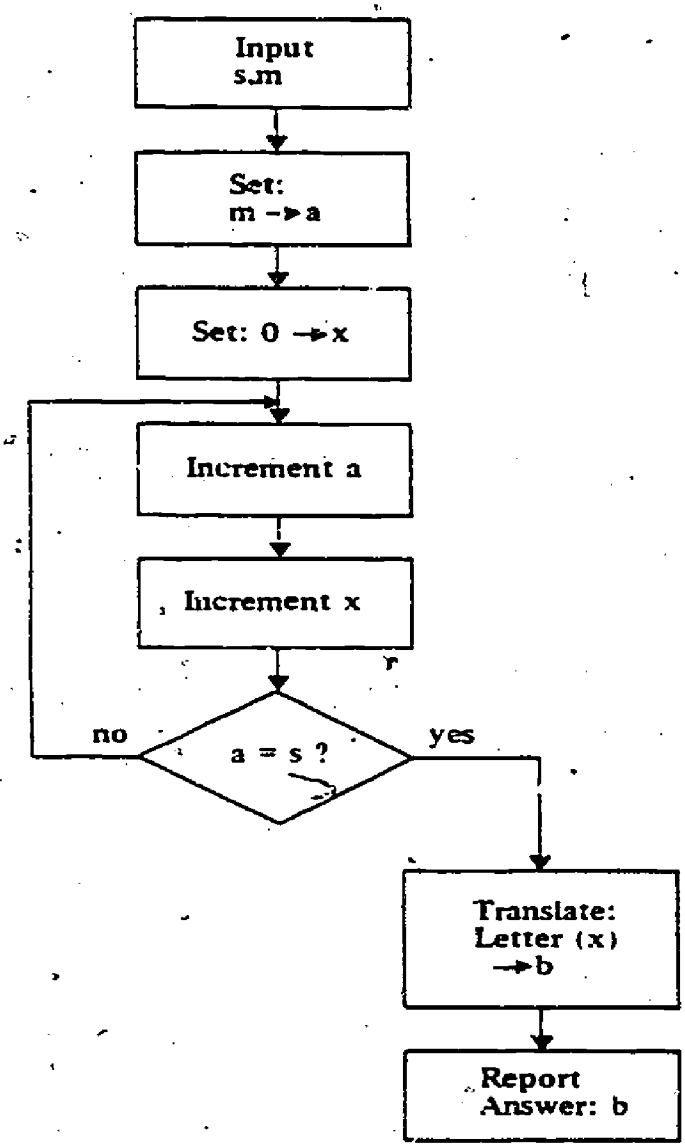


Figure 2. Methods for solving problems: ? + m = s, m + ? = s, s − ? = m, and s − m = ? —continued—

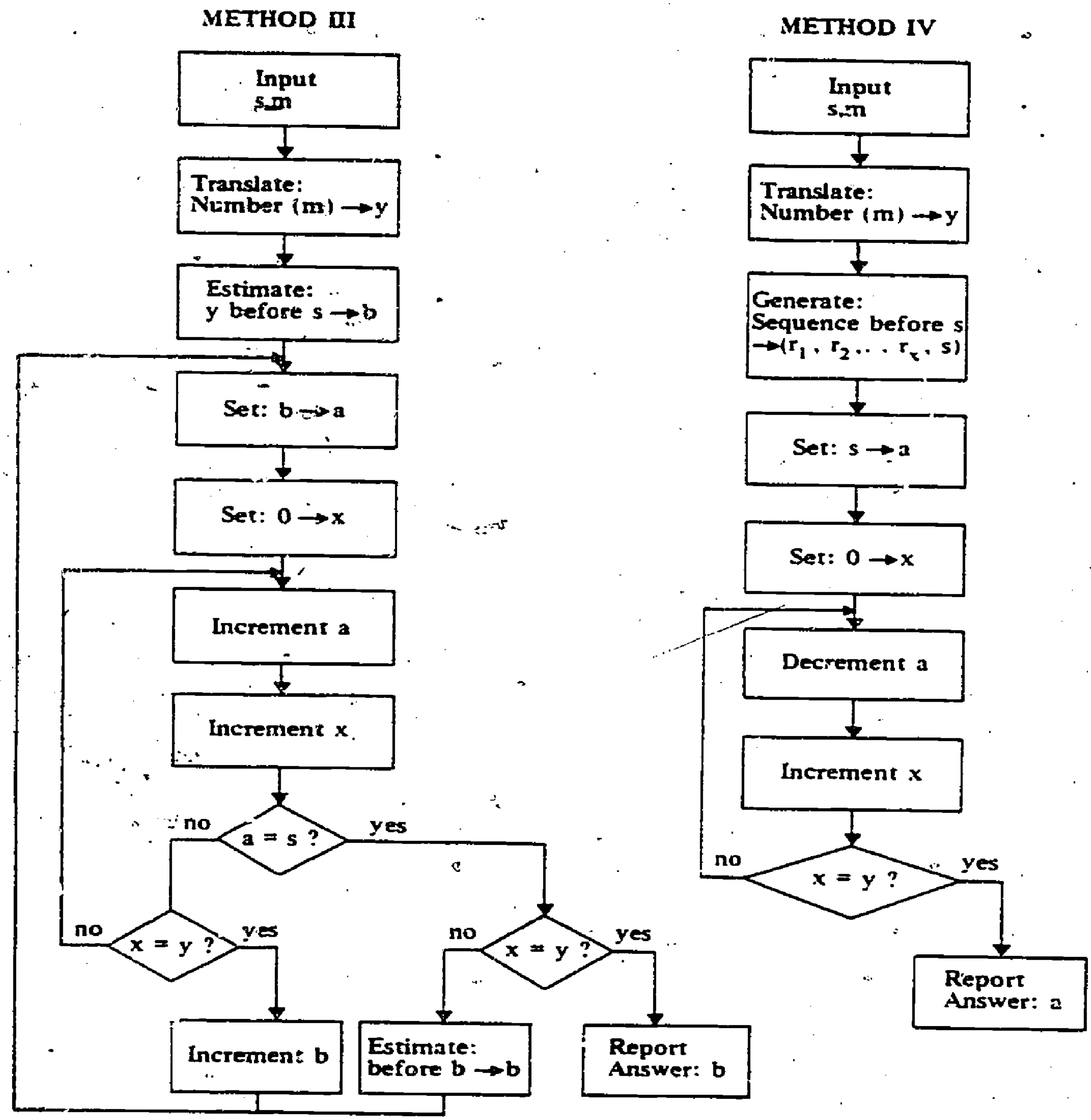


Figure 2 (continued).

zero. Each counter is incremented until the alphabet counter has reached E. The numeric counter will have reached 3. Three is translated to C, which is the answer. Fingers are often used here also for the numeric counting; for the example, "B, C (first finger), D (second finger), E (third finger), C!"

A third method for subtraction is shown as Method III in Figure 2. This is equivalent to counting backward to find the answer, but our introspections, as well as those of the subjects who were in our experiment, suggested that counting backward in the alphabet of letters is difficult and infrequent. Note that if it were possible to count backward, a method for subtracting would be like Method II for addition. The letter to be subtracted would be translated to a number, and the alphabetic counter would be decremented a number of times equal to that number. For example, for H - B = ?, "H, G (first finger), F (second finger), F!" Without the capability of counting backward, a system can begin a forward count at some reasonable point ahead of the target letter and see whether the interval turns out to be correct. For example, for H - B = ?, the letter to be subtracted, B, is translated to 2. Then an estimate is made--it might be E. The alphabetic counter is initialized at E and incremented. After each increment, two tests are made: "Have I reached H?" and "Have I made two increments yet?" For H - B = ?, starting with E, two increments will occur first. In the version of Method III shown in Figure 2, a new starting point will be chosen by incrementing E to F, and testing that possibility. There are plausible alternative remedies to those shown in Figure 2; for example, the discrepancy could be counted and used to adjust the initial trial letter. Another plausible idea is that if the initial try is wrong, another method, such as Method I or Method II, could be used to find the answer. However, the basic method of estimating and testing is the characteristic that distinguishes Method III from the other methods.

We anticipated the occurrence of Methods I, II, and III on the basis of our own introspections and earlier work (especially Woods, Resnick, & Groen, 1975). An additional method was used by subjects in our experiment often enough to merit showing it. The main innovation is a mechanism for generating a short sequence of letters that can be recorded in working memory and used as a way of locating the letter that is in a specified position. It is apparently difficult to count backward in the letter alphabet by retrieving information from long-term memory. However, if a short sequence of letters is put into short-term memory, it seems easier to determine which letter is in a specified position. For example, to solve the problem H - B = ? using Method IV, B would first be translated to 2. Then a short sequence would be generated, perhaps "E, F, G, H." With this sequence at hand, the system would find the letter that occurs two positions before H. In the version shown in Figure 2, this is done by counting backward. Note that this method can fail. The sequence that is generated may not be long enough to contain the answer. Figure 2 does not show a recovery procedure for this eventuality. One possibility would be to try again with a longer sequence. A more likely possibility would be to use a different method.

Experiment

Given a variety of possible methods for solving problems in letter-arithmetic, it seemed worthwhile to observe a few subjects to see whether there would be stable individual differences and whether different problems would be solved in reliably different ways.

Method

Materials. A set of 27 problem facts was constructed. Each fact consisted of a triple of letters including two addends and their sum. The first nine letters of the alphabet were divided into three subsets: small (A, B, C), medium (D, E, F), and large (G, H, I). Each of the letters in every subset was combined with one of the letters from each subset to form the 27 problem facts. Thus, there were three facts having the first and second addends from the small subset, three facts with the first addend small and the second medium, three facts with the first addend small and the second large, and so on. The individual letters occurred with approximately equal frequency in the 27 facts.

Each of the problem facts was used to form six problems. Denote the two addends as m and n and the sum as s. The six problems then have the forms: (a) ? + n = s; (b) m + ? = s; (c) m + n = ? ; (d) ? - n = m; (e) s - ? = m; and (f) s - n = ? . A total of 162 problems were constructed from the 27 facts.

The 162 problems were divided into three sets of 54 problems each. Each set of problems had nine examples of each problem form, and an approximate equalization was made of the occurrences of the various letters in each set.

Design, subjects, and procedure. Subjects were 12 University of Pittsburgh students who received course credit in Introductory Psychology for their participation. There were two conditions with six subjects run in each condition. In one condition subjects gave thinking-aloud protocols while solving a set of 54 problems immediately after being introduced to the task. Four of these six subjects returned for a second experimental session the following day when they received a set of 108 practice problems and then gave protocols on a second set of 54 problems. Subjects S5, S6, S7, S8, S9, and S13 were in this condition, and thus gave a set of protocols before having a session of practice problems. Subjects S5, S6, S7, and S9 returned for the second session, and thus gave a second set of protocols after having practice both on the initial 54-item problem set and on a 108-item practice set.

In the second condition, subjects were given a series of 108 practice problems after being introduced to the task. These practice problems were worked with pencil and paper during the first experimental session. These six subjects, S1, S2, S3, S10, S11, and S12, returned for a second experimental session the following day when they gave protocols on a set of 54 problems. Thus, these subjects only gave protocols after having practice on the 108-item practice set.

In the sessions in which subjects gave protocols, each problem was presented on an index card with the unknown term indicated by an asterisk. The subject was asked to give the answer to the problem and then to explain how the answer was obtained, usually with the question, "How did you get it?" If the subject's response did not seem sufficiently specific, the experimenter asked further questions until he was satisfied that enough information had been obtained to know whether individual letters had been translated into numbers and what counting operations had been used. The sessions were recorded on audiotape.

In the sessions involving practice problems, subjects were given two sets of 54 problems as a paper-and-pencil test. In the condition where practice problems were given in the first session, subjects were simply asked to be as fast and accurate as possible and the exercise was untimed. In the condition where practice problems were given in the second session, the problems were presented without a time limit, but subjects were informed at the end of each minute of work and were asked to mark the problem they had reached then. In the condition with practice given in the first session, the 108 practice problems and the 54 protocol problems were all different, constituting the three sets of 54 problems that were constructed for the experiment.

In the condition with protocols in the first session, one problem set was used for the initial set of protocols, the other two sets were used for practice, and one of the sets used in practice was also used for the final set of protocols so that subjects did not give both sets of protocols on the same set of problems.

Results

Each problem solved by each subject was classified as to the method used according to the subject's retrospective protocol. A solution was classified as Method I if the subject mentioned either the numerical value or a method of translation for finding the numerical value for each of the letters given in the problem. For example, Method I would be assigned to the solution of H - D = * if the subject said, "D is four, and I remembered G is seven, so H is one more. Then I subtracted." A solution was classified as Method II for addition if some reference was made to counting up from one of the terms; for example, for the problem * - D = H, "I counted up four from H." A solution was classified as Method II for subtraction if reference was made to counting between the given terms; for example, for the problem H - D = *, "I counted from D to H and got four." Method III or Method IV was assigned if a subject referred to a relation involving the size of the smaller term; for example, for the problem D - A = *, "C and D are right next to each other," or for the problem I - * = B, "I thought of G, H, I, so it's G."

Initially, solutions were classified into Methods I, II, and III. However, further analysis led to the conclusion that Method IV also occurred sufficiently often to merit its inclusion in the analysis, and the recorded protocols were examined again and reclassified. It was possible to arrive at reasonably definite classification of 92% of the solutions, including four solutions in which it was clear that either Method III or Method IV had been used, though we could not determine which. The 8% of the solutions that could not be classified included some where the subject's explanation was insufficient, but also included other forms of solution than the four categories we were using. Some solutions were based on recall of nearly the exact problem from earlier trials, such as having * - I = F by recalling * - F = I. Other solutions involved transformation of a problem into a more elaborate form, such as a solution for I + D = * that involved counting from D to I (five steps) and adding 5 to 8, which was obtained by doubling 4, the value of D.

Table 1 shows the number of solutions of each type given by each subject. (The four uncertain solutions are given as Type III.) Choice of methods did not appear to be changed substantially by practice.

Table 1
Frequencies of Solution Methods

Subject	Before Practice					After Practice				
	I	II	III	IV	Unclassifiable	I	II	III	IV	Unclassifiable
S5	41	5	2	0	6	45	7	0	2	2
S6	13	31	0	4	6	0	37	3	9	5
S7	52	0	0	0	2	46	0	0	0	8
S9	44	2	0	0	8	40	0	0	1	13
S8	43	8	0	1	2					
S13	4	34	12	0	4					
S1						39	10	0	0	5
S2						48	4	2	0	0
S3						20	26	0	8	0
S10						0	50	2	2	0
S11						9	36	0	3	4
S12						50	0	0	0	4

Two of the subjects who gave protocols without initial practice, Subject S6 before practice and Subject S13, used Method II on a majority of problems, and three of the subjects who practiced on 108 problems before giving protocols, Subjects S3, S10, and S11, used Method II on the majority of problems. The four subjects who practiced extensively between two protocol sessions used the same method in a majority of cases in both of the problem sets in which protocols were given.

Analyses were performed to identify features of problems that influenced choice of solution methods. First, consider the different forms of problems. Table 2 shows the frequencies of solution methods for problems of different form. There were small differences between the methods chosen for problems stated with plus rather than minus signs. The apparent difference between the two problem forms requiring addition was not significant, $t(15) = 1.6$, $p > .10$. There was a significant tendency to choose Method I more often for subtraction problems with minus signs than with plus signs (last two lines of the table versus lines 3 and 4), $t(15) = 3.1$, $p < .01$. However, the position of the unknown term clearly had no substantial effect, as seen in the virtual equality of line 3 with line 4 and of line 5 with line 6.

Table 2
Frequencies of Solution Methods

Problem Form	I	II	III	IV	Unclassifiable
m + n = ?	90	41			13
? − n = m	80	48			16
? + n = s	73	44	8	3	16
m + ? = s	75	48	3	7	11
s − ? = m	89	29	5	11	10
s − n = ?	87	36	5	8	8

In all remaining analyses, data are presented only from sessions in which the subject used more than one solution method on a substantial number of problems. We defined substantial use as solution of six or more problems with a method, and we omitted data from sessions in which only one method had a frequency of six or greater. The sessions thus selected for further analysis were the following: Subject S5 after practice, Subject S6 before practice, Subject S6 after practice, Subject S8, Subject S13, Subject S1, Subject S3, and Subject S11.

The various methods are easier to use for some problems than for others. Method II is easy to use for problems requiring a small number of counts. We defined the number of counting spaces as the smaller of two addends in a problem of form $\underline{m} + \underline{n} = ?$ or $? - \underline{n} = \underline{m}$ and as the number of letters between the two given letters in the other forms. The number of counting spaces thus equals the number of times that the incrementing loop would be used in solving problems using Method II. The mean number of counting spaces was calculated for problems solved in sessions where different methods were used on a substantial number of problems. The average number of counting spaces was calculated for the problems solved by each method that was used for at least six problems. The results are in Table 3.

Table 3
Mean Number of Counting Spaces for Problems
Solved Using Different Methods

Subject	Method I	Method II	Method III	Method IV
S5, After Practice	5.02	1.14		
S6, Before Practice	5.62	3.35		
S6, After Practice		4.49		5.89
S8	5.14	1.88		
S13		4.26	5.92	
S1	4.51	4.80		
S3	5.70	2.81		4.00
S11	4.78	3.92		

Note that for most of the subjects, the problems solved using Method II were substantially easier using that method than were the problems solved using other methods. The single exception appears to be Subject S1, for whom the number of counting spaces did not have a substantial effect on the overall frequencies of solution methods.

Theoretical Analysis

Models of Individual Performance

Performance in each session where more than one solution method was used on at least six problems was examined in detail and used as a basis for formulating a model of the process of selecting a method. These models should be considered as summary descriptions of the data; they are not strongly constrained by theoretical considerations, and the data do not provide strong tests of their validity. However, they do provide a description of the choices of solution methods in a way that relates the method to features of the problems and that permits relatively systematic comparisons between individual subjects.

The models were derived by informal examination of the features of problems that were solved by various methods by each subject. A list of the problems solved in each way was made, and these lists were scanned to find shared features. The features that seemed salient were then included in a model of the selection process, formulated as a decision network. Our major effort was to account for the choices of methods that occurred a substantial number of times. However, we included features in the model for choice of methods that occurred rarely if the choices seemed consistent with general principles that characterized other subjects.

Figure 3 shows the model induced from the performance of two subjects: Subject S5, in the session following practice, and Subject S8. These two subjects used Method I primarily, but used Method II

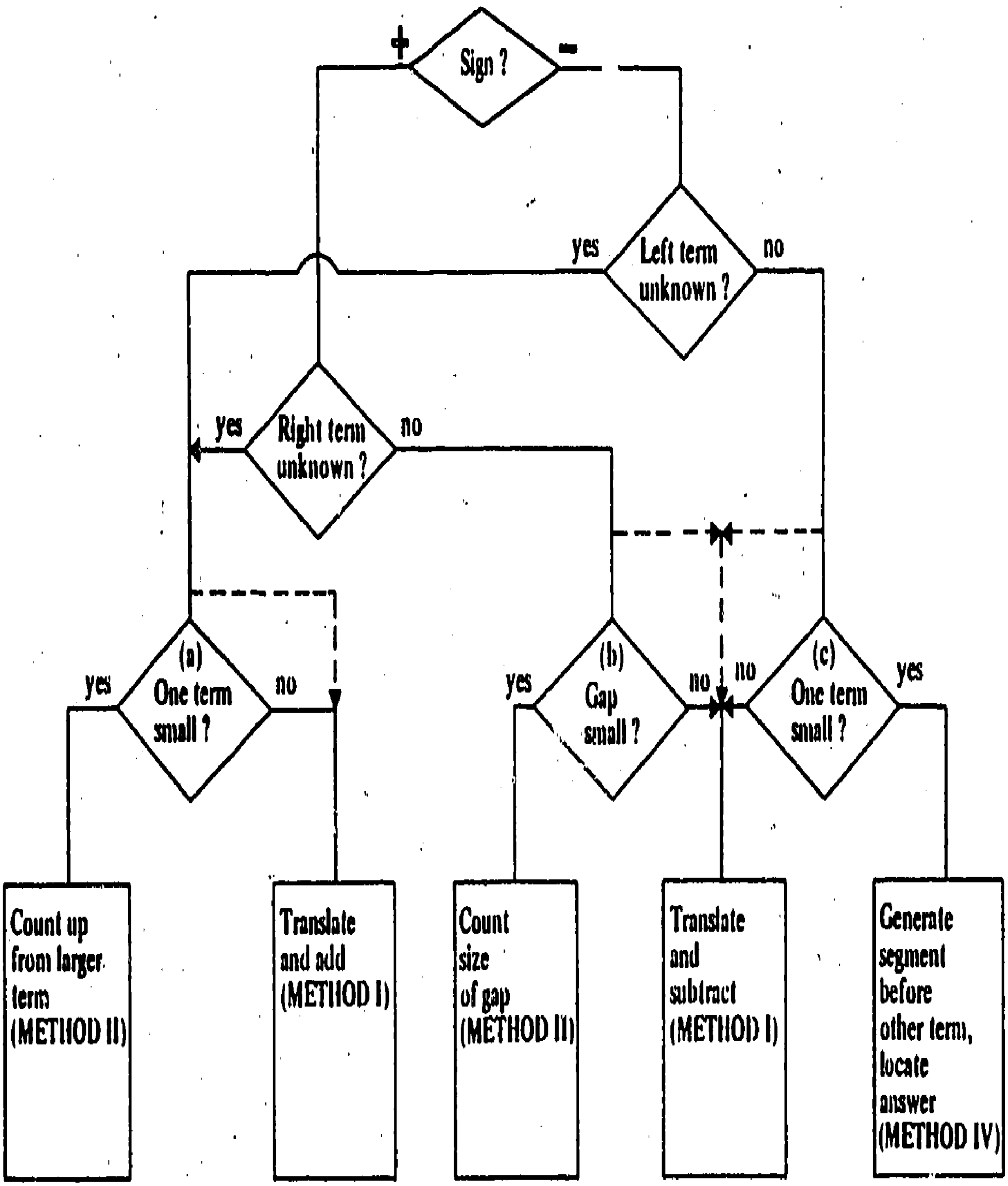


Figure 3. Model for Subjects S5 after practice and S8.

(counting up) several times each, and used Method IV (generate and locate answer) a few times.

The first two tests assumed to occur are tests that determine whether the problem requires addition or subtraction. In all of the models, this is assumed to occur through a simple process of pattern recognition, involving the sign of the expression and the position of the unknown term. Addition is required when the expression has a plus sign and the unknown is to the right of the equal sign and when the expression has a minus sign and the unknown is to the left of the minus sign.

Subjects S5 and S8 were consistent with the general trend discussed earlier in using Method II when there was a small number of counting spaces. For addition problems, this corresponds to having a small term given in the problem, and for subtraction problems, it corresponds to having the terms close together--that is, having a small gap between the terms. The tests for these two features are labeled (a) and (b) in Figure 3, with Method II chosen if these tests are positive and Method I chosen if they are negative. The process in Figure 3, like most of the models we will present, does not differentiate between the two forms of addition problems, m + n = ? and ? - n = m. Recall from Table 2 that overall performance was very similar for these two problem forms. Note that Figure 3 does not include a test for a small gap for subtraction problems when the expression has a minus sign. Subjects S5 and S8 never used Method II with subtraction problems having minus signs, so a better fit to their data is obtained if the test for small gap is assumed to occur only for those subtraction problems resulting from missing addends. In the subtraction problems where Method IV was used, the term to be subtracted was small--in fact, it was always A. This test is indicated as Test (c) in Figure 3. Note that this test is not reached unless the expression of the problem has a minus sign. Again, this agrees with the data obtained with these two subjects; Method IV was not used for any problems having plus signs.

How well does Figure 3 predict the performance of Subjects S5 and S8? This depends on how one interprets the tests for small terms and small gaps. One possibility would be to choose a cutoff value for term size and gap size and assume that the test would succeed whenever the term or gap was smaller than that critical value. This approach seems inappropriate for the present situation; subjects were not consistent in that way. Another approach is to examine the frequency with which different terms and gap sizes occurred in problems where according to the model the tests would have been applied. These are given in Table 4. The data are pooled from the two subjects.

Table 4
Frequencies of Outcomes of Hypothetical Tests
by Subjects S5 After Practice and S8

Minimum Term	Test (a)		Test (c)		Gap Size	Test (b)	
	Yes	No	Yes	No		Yes	No
A	4	6	3	1	1	5	0
B	4	8	0	0	2	0	2
C	0	2	0	2	3	2	2
D	0	6	0	8	4	0	0
E	0	3	0	4	5	0	6
F	0	0	0	6	6	0	6
G	0	0	0	6	7	0	6
H	0	2	0	2	8	0	4
I	0	0	0	4	9	0	2

The results show that these features were relevant factors. No minimum term larger than B and no gap larger than three were ever identified as small. However, there were instances in which small letters and gaps apparently were not identified as small by the subjects. One interpretation is that these subjects had a relatively strong preference for using Method I and used that method unless they happened to notice that a particularly small letter or gap was present in a problem. This idea corresponds to an assumption that the tests shown in Figure 3 are not applied to every problem. Suppose that the dashed lines in Figure 3 represent possible paths through the decision network that are taken with some probability. Since the dashed lines avoid the feature tests, it follows that the features are not tested on some trials. When they are tested, the process can terminate with a choice of Method II or Method IV, but a choice of Method I will occur for some problems in which a small term or a small gap is present.

Figure 4 shows a model that accounts for the performance of Subject S6 in the problems solved before practice. S6 had a preference for Method II rather than Method I and also used Method IV on a few problems. On the problems involving addition, S6 always used Method II when the smaller addend was A or B and used a mixture of Method I and Method II for larger addends; see the first columns of Table 5.

Table 5
Outcomes of Hypothetical Tests by Subject S6 Before Practice

Minimum Term	Test (a)		Test (c)		Gap Size	Test (b)	
	Yes	No	Yes	No		Yes	No
A	4	0	3	0	1	3	0
B	2	0	1	2	2	2	0
C	1	3	0	2	3	0	0
D	1	0	0	1	4	1	0
E	2	0	0	2	5	2	1
F	0	1	0	2	6	0	1
G	0	0	0	2	7	1	1
H	0	1	0	2	8	0	2
I	0	0	0	1			

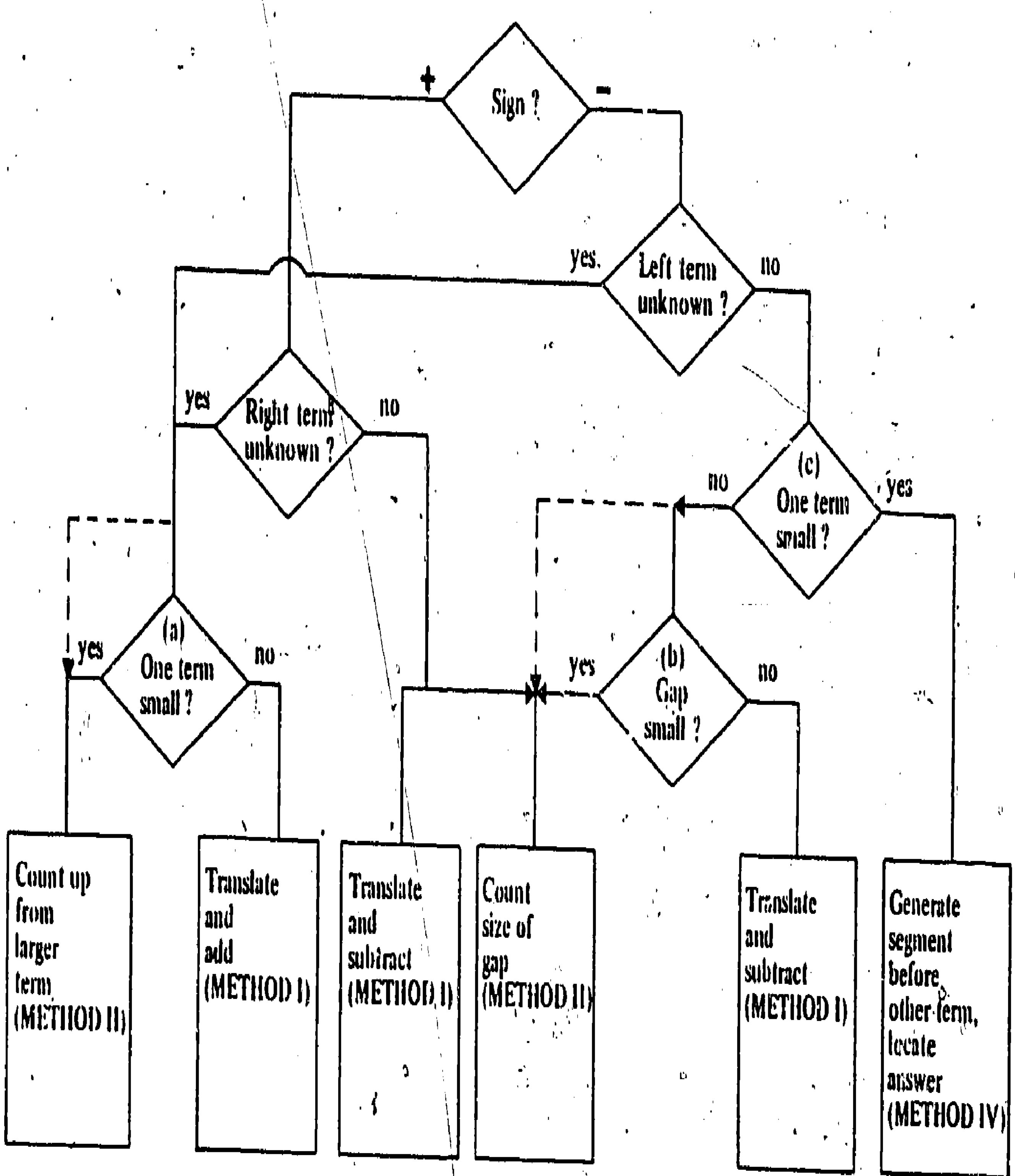


Figure 4. Model for Subject S6 before practice.

For problems with a plus sign and a missing addend, S6 usually used Method II; three occurrences of Method I were not explicable. The arrangement of tests shown for subtraction problems with minus signs was chosen because S6 chose Method IV for all these problems in which an A occurred and then appeared to discriminate on the basis of the gap size; see the two sets of columns in Table 5 for Tests (b) and (c).

A somewhat simpler model accounted for Subject S6's performance on the problems worked after practice. Recall from Table 1 that Subject S6's overall frequencies of solution methods were somewhat different after practice than before. The model for performance after practice is in Figure 5. After practice, S6 solved all addition problems using Method II. On subtraction problems, if the size of the minimum term was not small, Method II was used. If the size of the minimum term was small, S6 used either Method IV (nine times) or Method III (three times). Table 6 shows that the discrimination made in the test of term size was quite sharp. Table 6 also shows the discrimination that would be made at Test (a) if it were assumed that the size of the gap, rather than the size of the subtracted term, were tested.

Table 6
Outcomes of Hypothetical Tests by Subject S6 After Practice

Minimum Term	Test (a)		Gap Size	Position of Test (a)	
	Yes	No		Yes	No
A	6	0	1	2	0
B	4	0	2	3	0
C	1	3	3	2	2
D	1	0	4	4	3
E	0	6	5	3	1
F	0	4	6	2	1
G	0	2	7	2	0
H	0	4	8	2	2
I	0	3	9	2	3

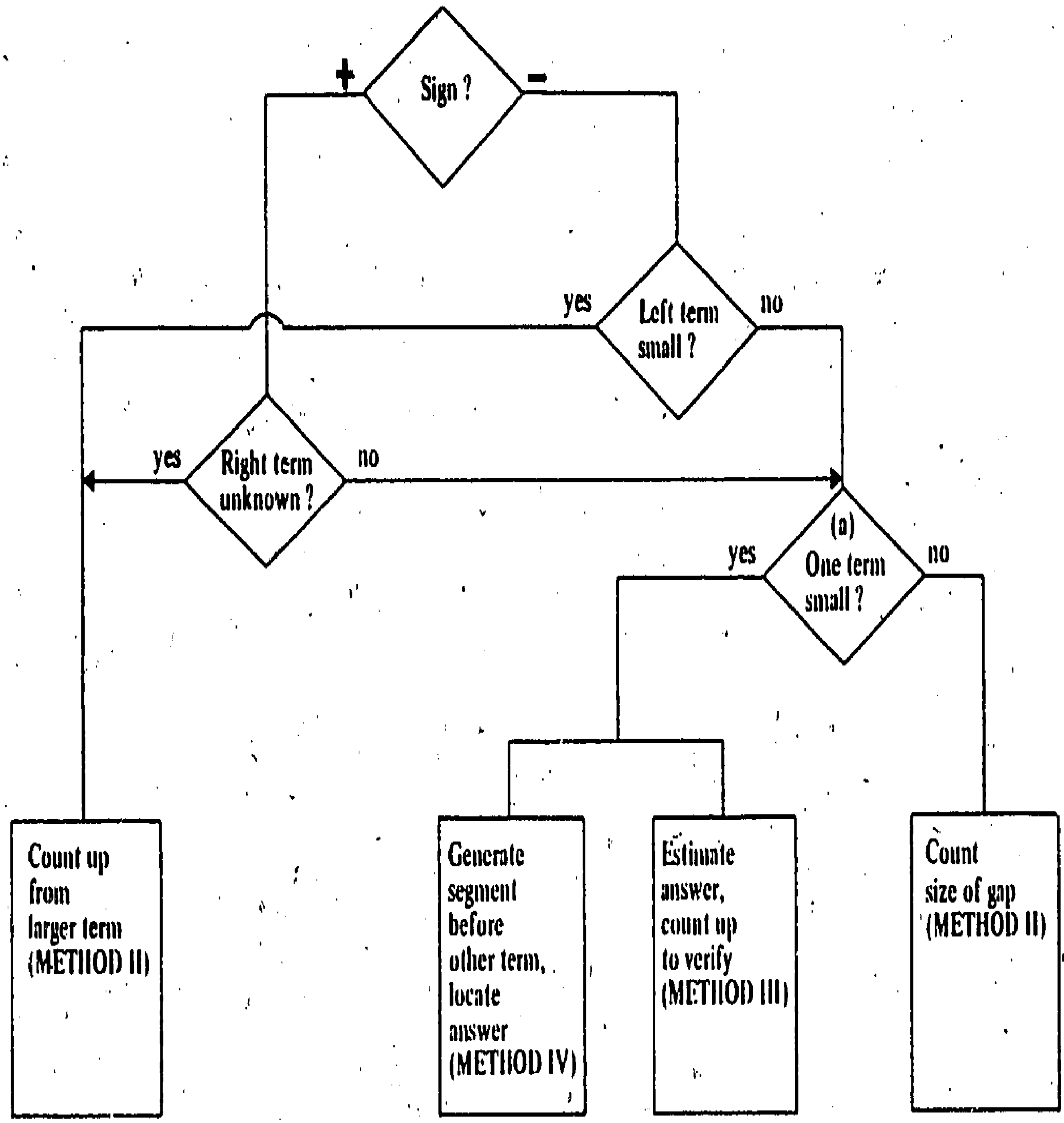


Figure 5. Model for Subject S6 after practice.

It seems quite likely that the decision to use Method III or Method IV was based on the size of the minimum term rather than on the size of the gap.

Figure 6 shows the model that accounts for the performance of Subject S13, who mainly used Method II, used Method III on 12 problems, and used Method I on four problems whose salient features were not clear to us. It seemed to us that the decision to use Method III rather than Method II depended jointly on finding that the gap between terms was not small, but that a small term was to be subtracted, as shown on the right side of Figure 6. The occurrences of this joint outcome are shown in Table 7 as x's; note that they tend to be concentrated in the upper right portion of the matrix, where small minimum terms and large gaps are located.

Table 7

Outcomes of Test Pairs: Negative for Small Gap and Positive for a Small Term by Subject S13

Gap Size	1	2	3	4	5	6	7	8	9
Minimum Term									
A				x				x	
B					x	x	x		
C	o		o		x		x		
D	o				ox	ox		o	x
E			o				x		
F		o					o		oo
G			x			o	oo		
H			o					oo	o
I				o		o		o	

Note: x denotes choice of Method III, an outcome in agreement with features tested; o denotes disagreement with one or both features.

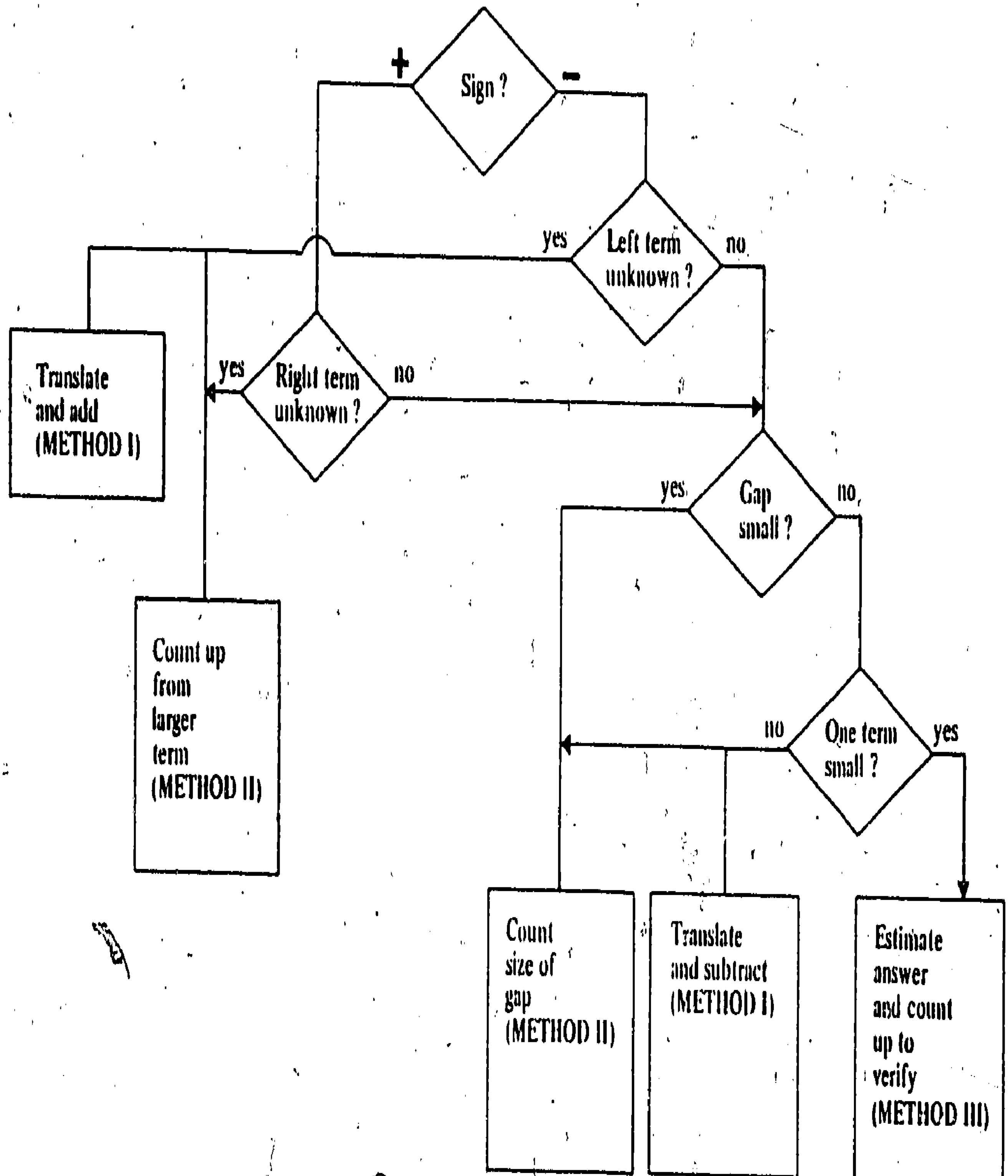


Figure 6. Model for Subject S13.

Our effort to represent Subject S1's performance in Figure 7 was not very successful. S1 had a strong preference for Method I. Method II was used on two problems of the form ? - $\underline{n}$ = $\underline{m}$, but these two were not similar in any noticeable way. Method II was also used on nine subtraction problems, and although there appeared to be some relationship between use of Method II and the size of the term to be subtracted, the relation was not a strong one, as Table 8 shows. A possibility is that S1 tended to use Method II when retrieval of the numerical equivalents of the letters seemed difficult. Practice in the task would be expected to lead to storage in memory of a fairly large set of associations between letters and numbers, especially if Method I was used for most problems. If the use of Method II was based on a failure to find an association in memory during a brief search, there would be a tendency for this to happen more frequently for problems with relatively large terms, as occurred for Subject S1. Recall from Table 3 that Subject S1's choices of solution methods did not depend strongly on the number of counting spaces in the problems. The model shown in Figure 7 is consistent with that general feature of this subject's performance.

Table 8
Outcomes of Hypothetical Test by Subject S1

	Test (a)	
Minimum Term	Yes	No
A	2	1
B	4	0
C	2	1
D	5	2
E	3	0
F	0	1
G	3	1
H	1	2
I	1	2

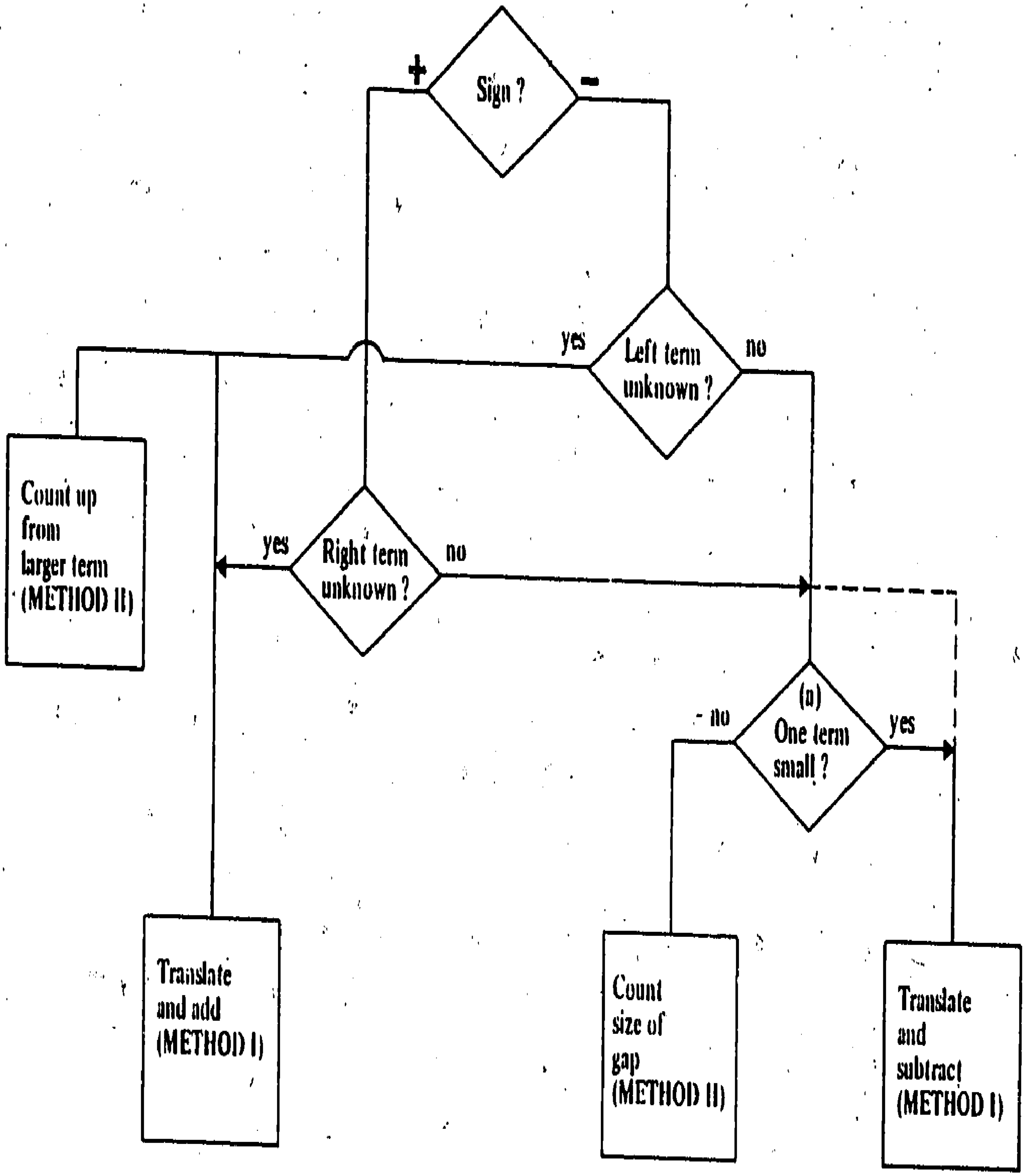


Figure 7. Model for Subject S1.

The model developed for Subject S3 is in Figure 8. As Table 9 shows, there was rather good discrimination in the hypothetical tests for the size of the minimum-term and rather poor discrimination of the size of the gap. Our conjecture is that for problems that did not have a small term, S3 may have used Method II whenever the gap was very small and, if it was not, considered Method I and used it if the numerical values of the letters in the problem were easy to retrieve, otherwise resorting to Method II.

Table 9

Outcomes of Hypothetical Tests by Subject S3

Minimum Term	Test (a)		Test (c)		Gap Size	Test (b)	
	Yes	No	Yes	No		Yes	No
A	4	0	[illegible]	0	1	6	0
B	2	0	2	2	2	4	1
C	2	1	1	2	3	1	1
D	0	3	1	3	4	2	0
E	0	1	0	3	5	1	1
F	0	3	0	3	6	1	2
G	0	0	0	6	7	2	2
H	0	1	0	5	8	0	3
I	0	0	0	5	9	1	1

The model developed for Subject S11 is in Figure 9. S11 used Method II on most problems. Method I was used on seven addition problems and two subtraction problems. Although there was a relationship between choice of Method I and the size of the minimum addend, there was some indication that S11 chose Method I when the numerical values of the terms were known, as Table 10 suggests.

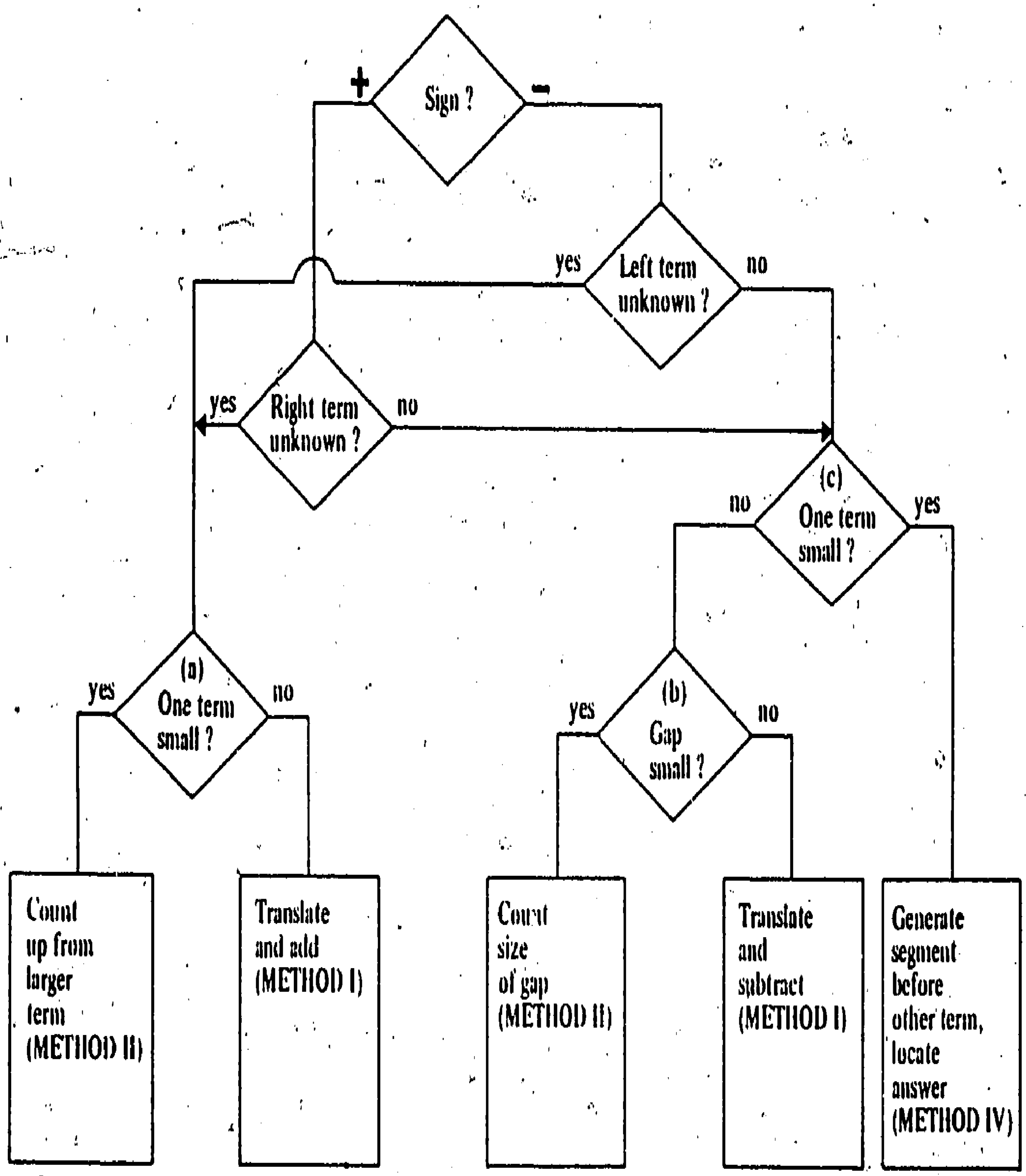


Figure 8. Model for Subject S3.

31

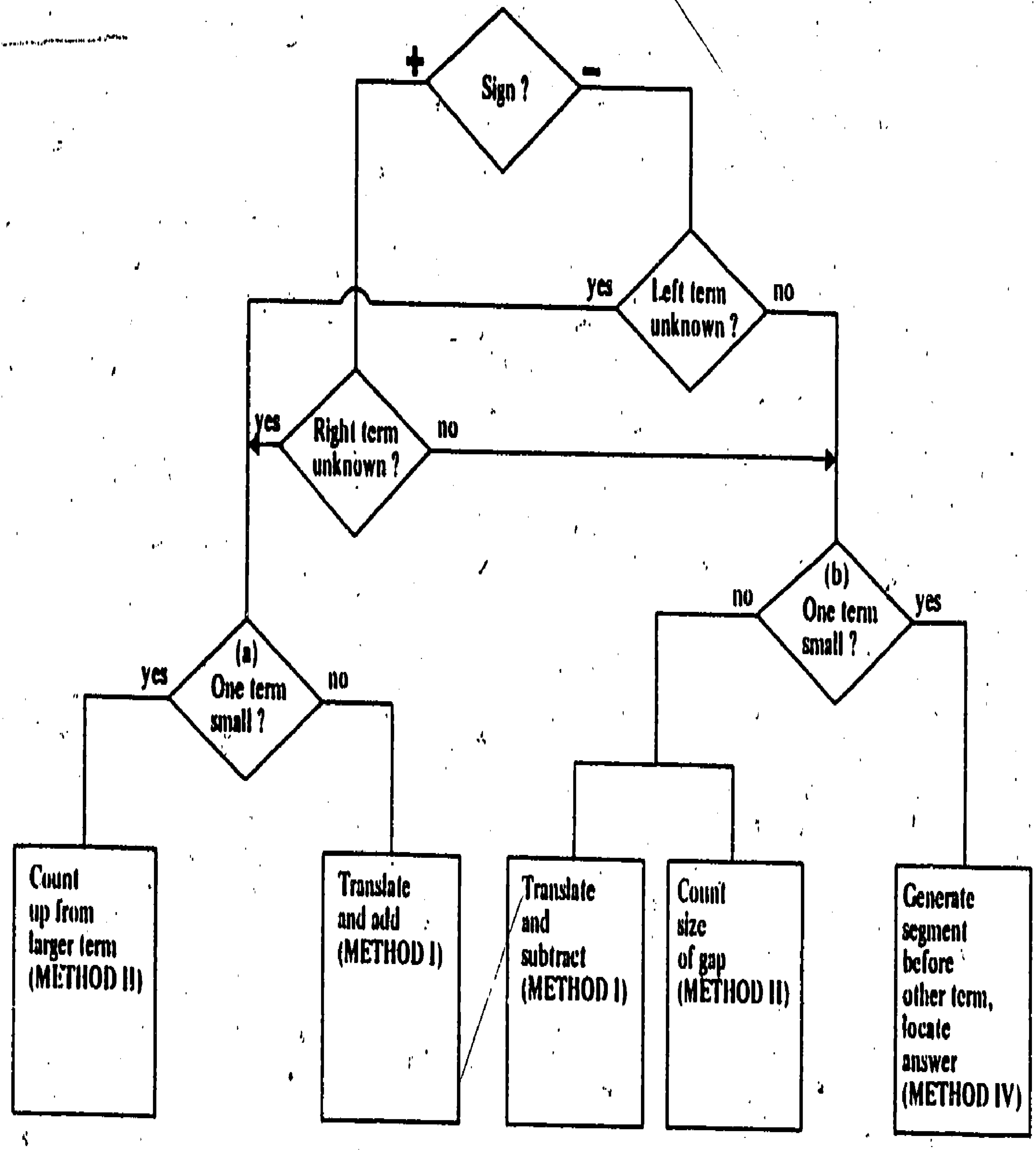


Figure 9. Model for Subject S11.

Table 10
Outcomes of Hypothetical Tests by Subject S11

Minimum Term	Test (a)		Test (b)	
	Yes	No	Yes	No
A	3	1	2	0
B	1	0	1	0
C	2	2	0	3
D	2	1	0	3
E	0	1	0	3
F	0	1	0	3
G	0	0	0	6
H	0	1	0	4
I	0	0	0	5

A Composite Model

While the details of the models for individual subjects differed, it turned out that performance of all the subjects could be explained reasonably well. The set of arithmetic procedures that we considered initially was nearly complete; Method IV was added in response to performance of subjects. The selection of procedures by subjects who used mose than one method also involved a considerable amount of similarity across subjects. In all the models we have presented here, the features tested are the size of the smaller term in the problem and the size of the gap between the two given terms. During the development of these models, various different features were included in tentative versions of models. For example, we considered the possibility that some subjects were testing the size of a specific term, such as the middle term in an addition problem, rather than the size of the smaller term regardless of its position. However, when we examined the performance of subjects systematically. we always came to the conclusion that the data were not explained any better by the inclusion of these alternative features.

It seems worthwhile to examine the relationship between procedural selection and the sizes of the minimum term and the gap between terms combining data across subjects. Combined data are more stable, and although the conclusion drawn from them may not actually represent the processes of any single subject, they should provide a useful characterization of group trends. The data used for this analysis were the methods chosen in the eight sets of protocols for which individual-subject models were constructed. In the other eight cases, a single method was used almost uniformly, so little information could be obtained about the process of selecting a procedure. The composite model is shown in Figure 10.

Reliability of effects in these data was estimated using chi-square tests based on frequencies pooled across subjects. These test statistics are probably not distributed as chi-square under the null hypothesis, since each subject contributed several observations to the data, and hence the observations were not independent. On the other hand, the test statistics give an indication of the relative sizes of effects compared with differences that might occur because of simple random variation. It also should be kept in mind that the composite model to be presented here is a kind of summary of the general trends in the data rather than a model that we think characterizes the performance of individual subjects.

First, consider problems that were solved by addition, with formats $\underline{m} + \underline{n} = ?$ and $? - \underline{n} = \underline{m}$. Most subjects seemed to choose Method II, counting up from the larger term, for these problems more often when the smaller of the two terms was small. Figure 11 shows the influence of the size of the minimum term on choice of a solution method for these problems. The data are the proportions

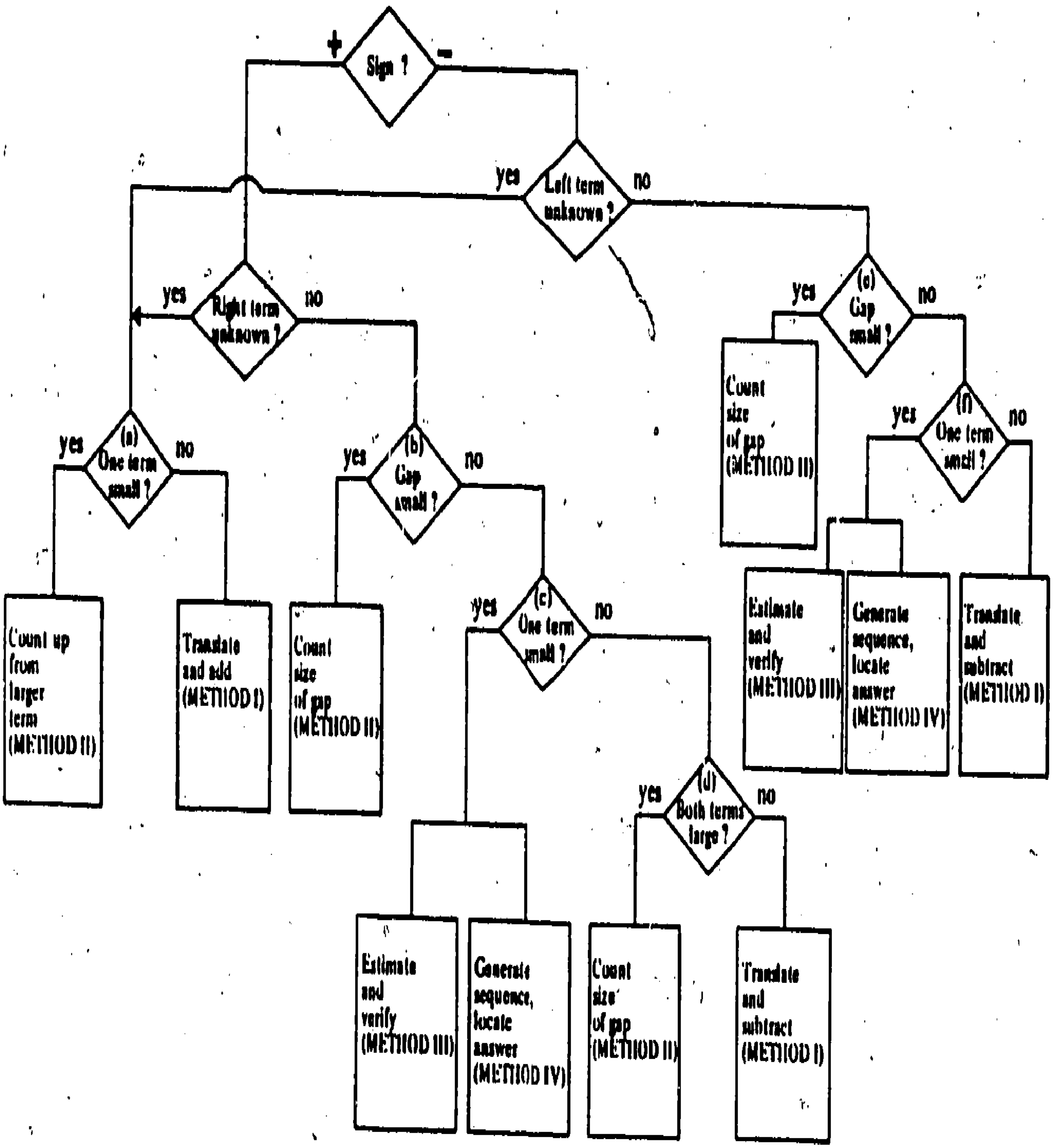


Figure 10. Composite model based on general trends in data pooled across subjects.

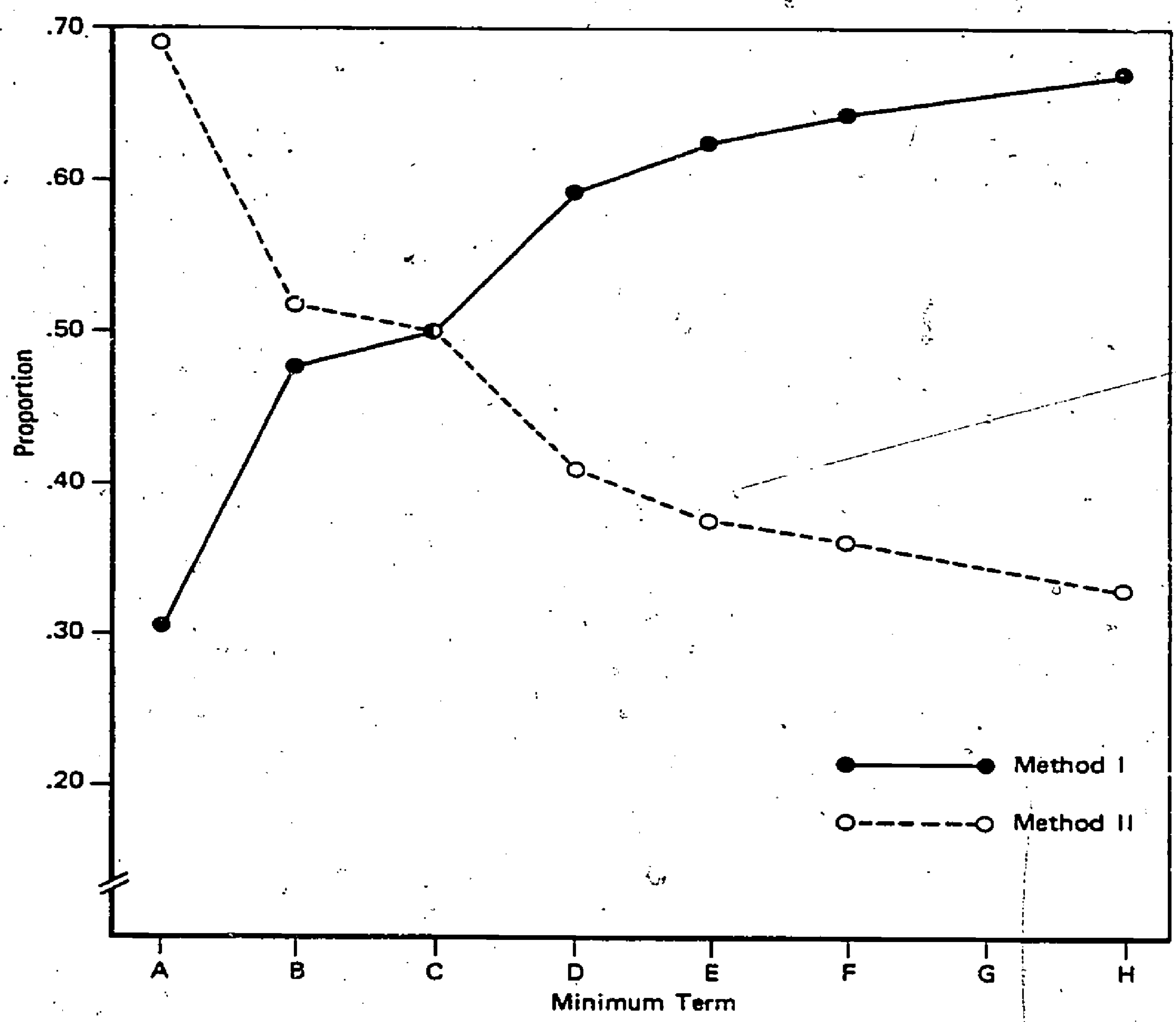


Figure 11. Proportions of choice of solution methods for addition problems: m + n = ? and ? − n = m.

of all problems having the various minimum terms that were solved, respectively, by Method I, translating and adding, and by Method II, ignoring problem solutions that could not be classified. (Note that Methods I and II were the only ones found for these problems, so the proportions must sum to 1.0.) The difference shown in Figure 11 is not large enough to be statistically significant, $\chi^2(6) = 7.58$, $p > .20$; however, the effect is systematic, and it seems reasonable to conclude that the size of the minimum term was probably a factor in at least some subjects' choices of methods for solving these problems. In Figure 10, the test for size of the minimum term for addition problems is shown as Test (a). We have no strong indications of any substantial difference between the two formats of addition problems. In problems with the format ? - n = m, problems with A or B as the minimum term were solved with Method II more often (.71) than was the case for problems with format m + n = ? (.50); however, the difference between the formats in choice of methods was clearly nonsignificant, $\chi^2(5) = 3.50$, $p > .60$. Because of this absence of a difference, the model in Figure 10 shows a single procedure that is applied for problems with both the m + n = ? and ? - n = m formats, as was true for all the individual models except those for Subjects S13 and S1.

In analyzing the problems that required subtraction, two factors appeared to determine the choices of solution methods by individual subjects. One factor was the size of the minimum term of the problem; Method III, estimating and verifying, and Method IV, generating a sequence and locating the answer, were used more often in problems with small minimum terms. A second factor was the size of the difference or gap between the two terms; Method II, counting the size of the gap, was used more often for problems in which the gap was small. Two analyses were conducted with each of these factors. With problems partitioned by their minimum terms, frequencies of Method III and IV versus Method I and II were analyzed, and frequencies of Method I versus Method II were then analyzed. With problems partitioned by the sizes of gaps, frequencies of Method II versus Methods I, III, and IV were analyzed, and frequencies of Method I versus Methods III and IV were then analyzed. No effort was made to separate Methods III and IV in the analyses due to small frequencies of their use.

There was substantial evidence that subtraction problems with minus signs, s - n = ? and s - ? = m, differed from missing addend problems, ? + n = s and m + ? = s, in choices of solution methods by subjects, so these two sets of problems were analyzed separately. Because of the differences, which will be documented below, the model in Figure 10 has different procedures for choosing methods for missing addend problems (plus sign, then negative for "right term unknown?") and subtraction problems (minus sign, then negative for "left term unknown?"). There were no substantial differences between the two formats of subtraction problems with minus signs nor between the two formats of missing addend problems; the various test statistics will be reported as we present the analyses. Thus, in Figure 10, there is a single procedure for subtraction problems with minus signs and another single procedure for missing addend problems, with no distinction depending on the the location of the unknown term.

Figure 12 shows the proportions of choice of solution methods for subtraction problems with minus signs, partitioned by the size of the gap between terms in the problem, and Figure 13 shows the data for subtraction problems with minus signs, partitioned by the minimum term of the problem. The effect in Figure 12 in which Method II was preferred for problems with small gaps, but not for

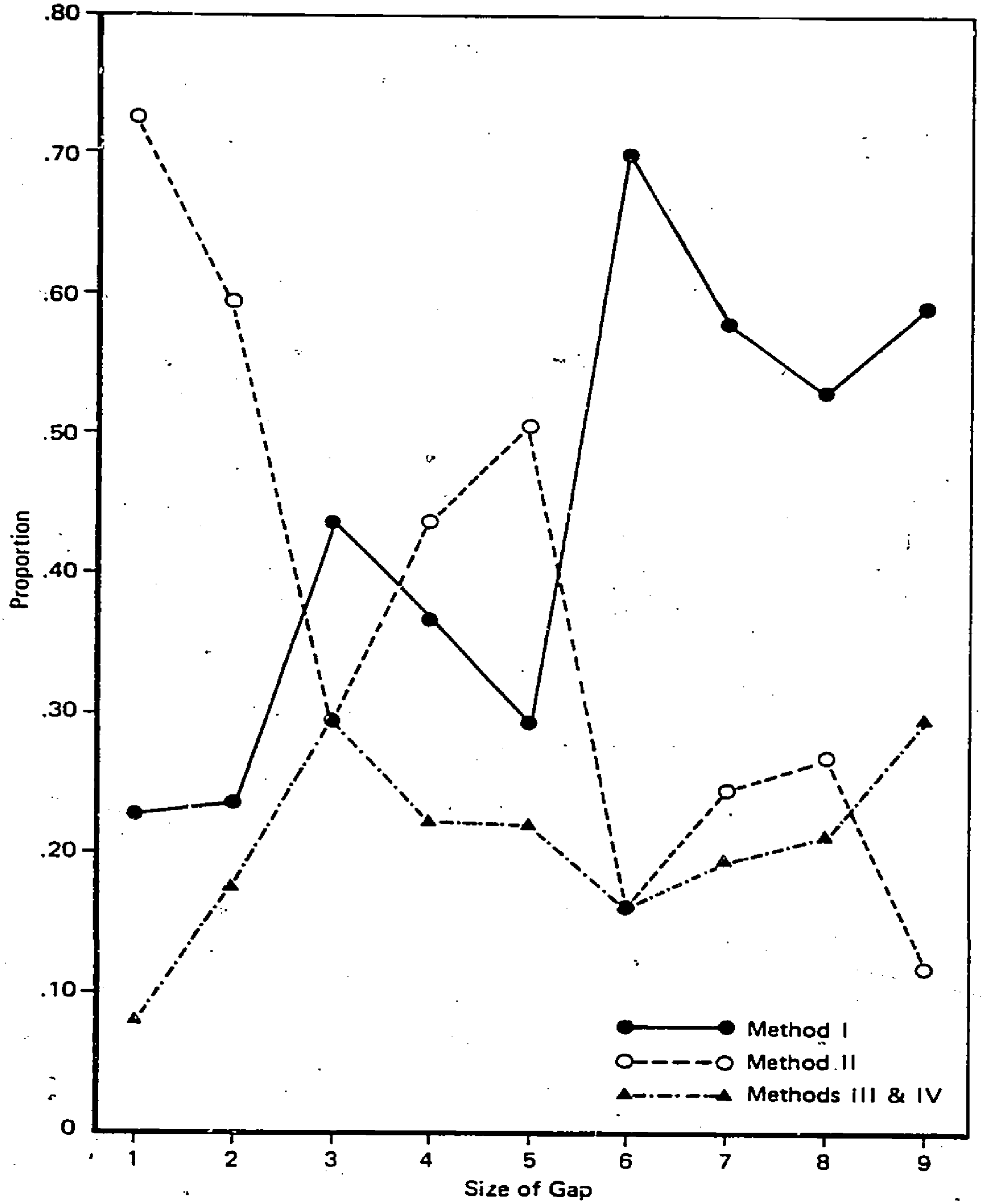


Figure 12. Proportions of choice of solution methods for subtraction problems with minus signs: s — ? = m and s — n = ?

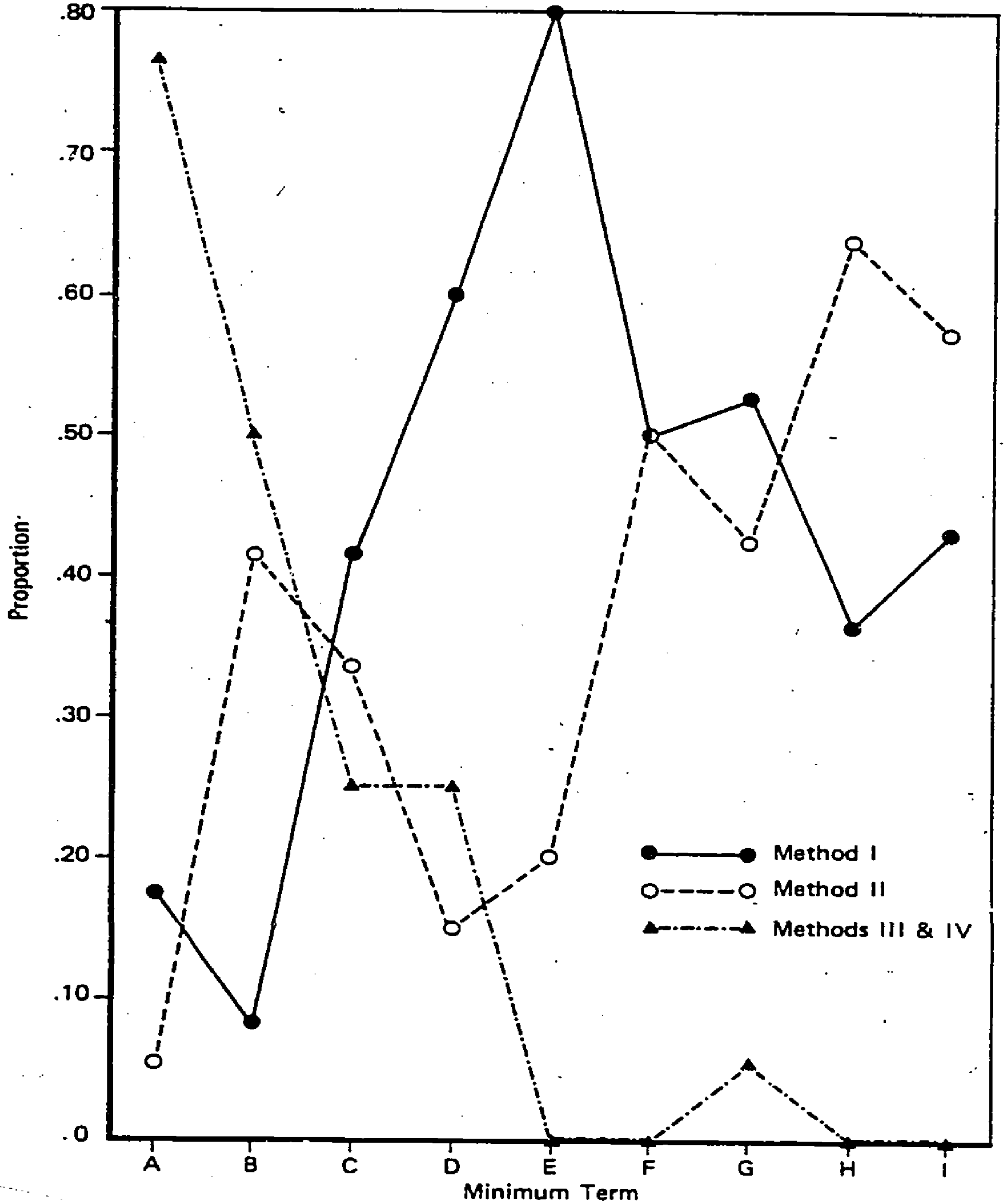


Figure 13. Proportions of choice of solution methods for subtraction problems with minus signs: s − ? = m and s − n = ?

larger gaps, was significant. A chi-square test of independence between size of gap and choice of Method II versus the other methods gave $\chi^2(8) = 21.13$, $p < .01$. (The two subtraction problem formats did not differ significantly, $\chi^2(4) = 3.11$, $p > .50$.) Choices between Method I and Methods III and IV were not influenced significantly by the size of the gap, $\chi^2(5) = 1.92$, $p > .80$. (The two formats did not differ, $\chi^2(3) = 1.69$, $p > .60$.)

In Figure 13, the effect in which Methods III and IV were used more for problems with small minimum terms was reliable, $\chi^2(3) = 54.40$, p = nil. (Difference between formats: $\chi^2(2) = 2.12$, $p > .30$.) Considering choices between Method I and Method II, there were more choices of Method I for intermediate minimum terms (C, D, and E); however, the choices between Method I and Method II did not depend significantly on the minimum term, $\chi^2(7) = 11.06$, $p > .10$, and since the apparent effect was nonmonotonic, it seems most reasonable to attribute it to chance. (Differences between formats: $\chi^2(4) = 1.66$, $p > .70$.)

The procedure for selecting methods for subtraction problems with minus signs is shown on the right side of Figure 10. A tendency to choose Method II if the gap is small is represented by Test (e), and the choice of Methods III and IV for small minimum terms is represented by Test (f).

Figure 14 shows the proportions of choice of solution methods for missing addend problems, partitioned by the size of gap, and Figure 15 shows the data for missing addend problems partitioned by the minimum term. The effect in Figure 14 in which Method II is used predominantly for problems with small gaps was significant, $\chi^2(8) = 28.52$, p = nil. (The difference between the two missing addend problem formats was not significant, $\chi^2(4) = 3.70$, $p > .40$.)

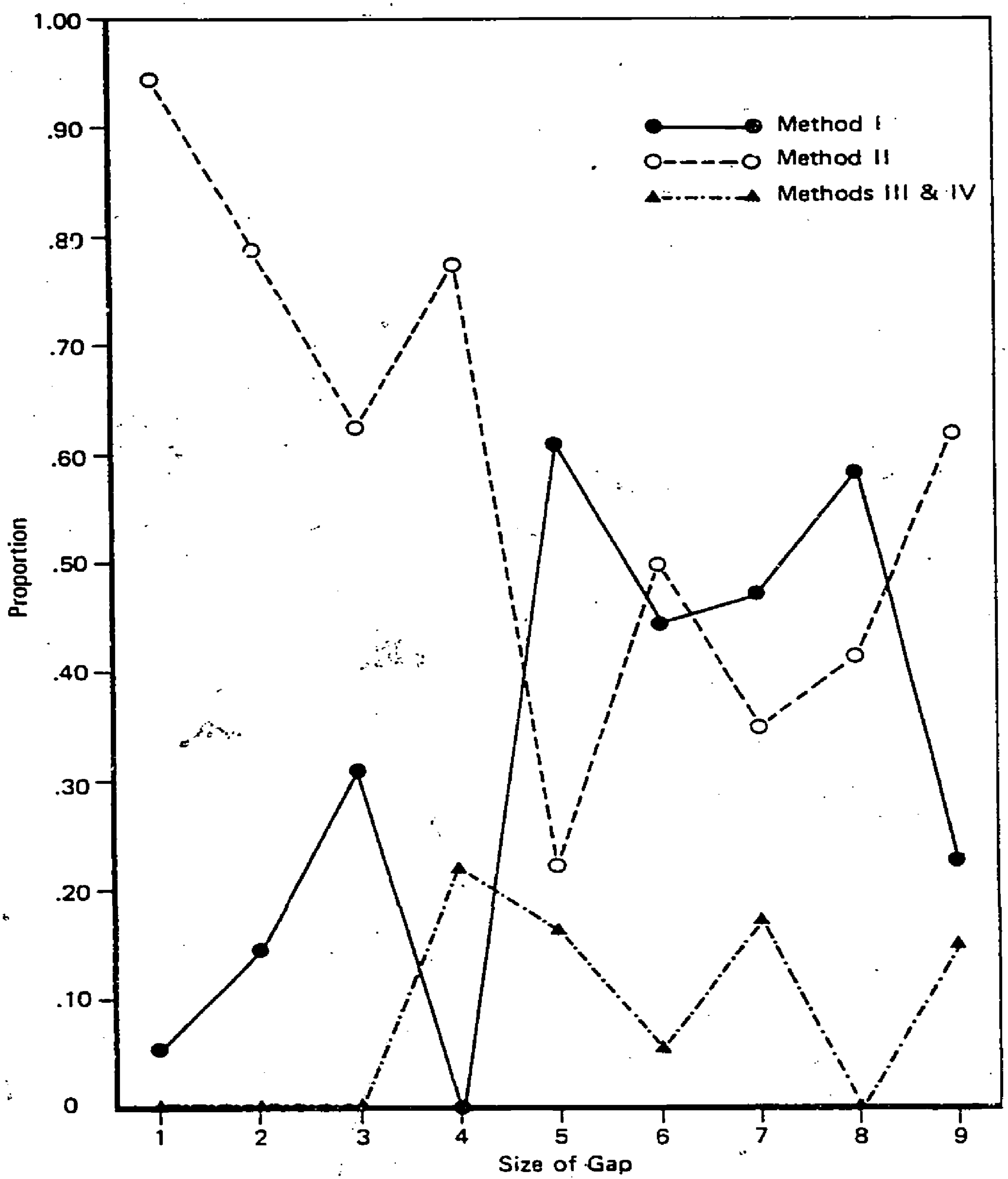


Figure 14. Proportions of choice of solution methods for missing addend problems: ? + n = s and m + ? = s.

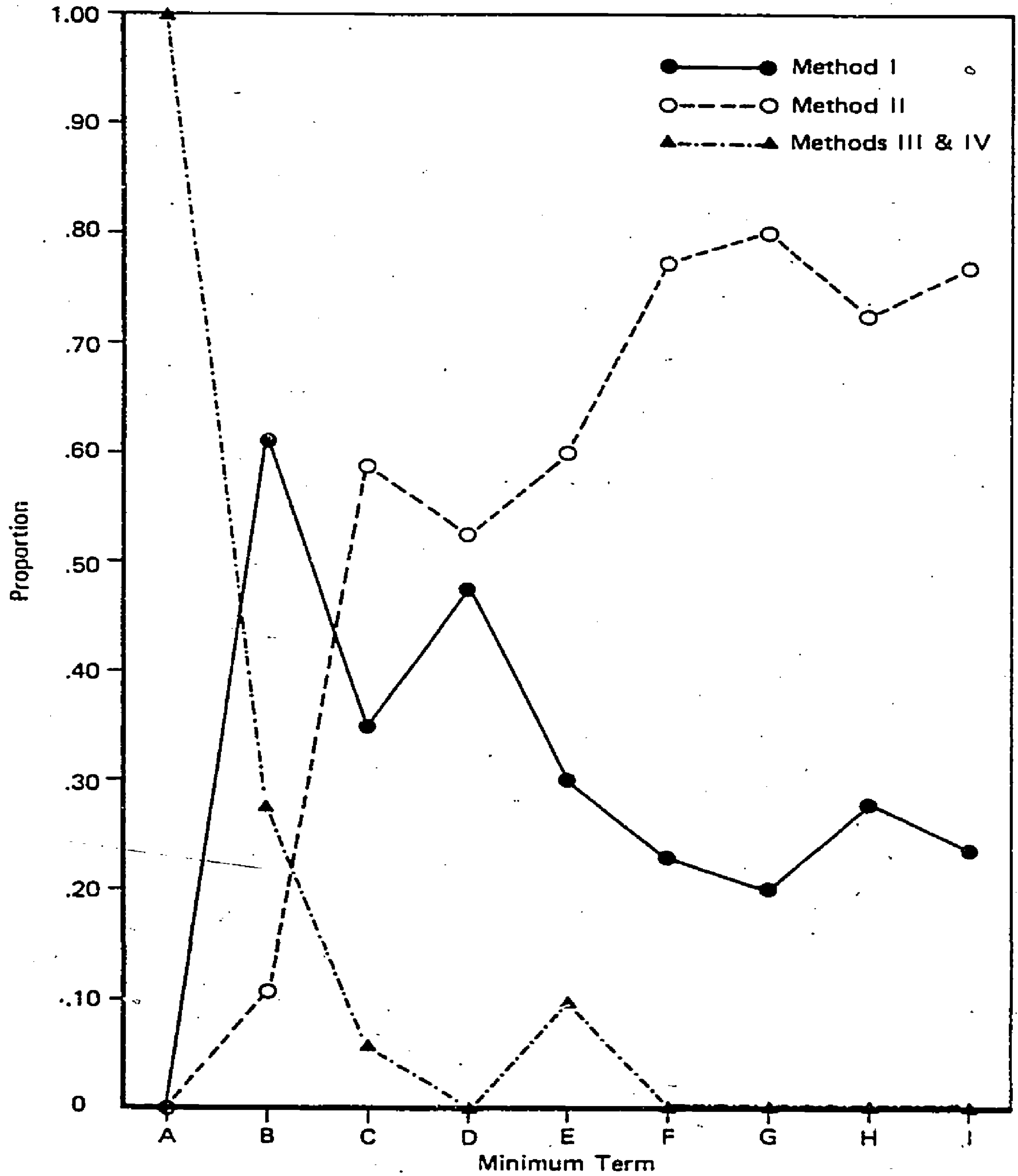


Figure 15. Proportions of choice of solution methods for missing addend problems: ? + n = s and m + ? = s.

Choices between Method I and Methods III and IV were not significantly influenced by the size of the gap, $\chi^2(1) = .55$, $p > .40$. (Difference between formats: $\chi^2(2) = .50$, $p > .60$.)

In Figure 15, the effect in which Methods III and IV were used predominantly in problems with small minimum terms was significant, $\chi^2(1) = 12.46$, p = nil. (Difference between formats: $\chi^2(2) = 1.25$, $p \not> .50$.) Choices between Method I and Method II also were significantly dependent on the minimum term of the problem, $\chi^2(6) = 18.00$, $p < .01$. (Difference between formats: $\chi^2(4) = 1.17$, $p > .80$.)

A selection procedure that represents these effects is shown in the central portion of Figure 10. Test (b) represents the tendency to choose Method II if the gap is small. Test (c) represents a tendency to choose Method III or Method IV if one of the terms is small. Test (d) represents a test with a somewhat larger letter as its criterion than the test for a small term at Test (c). For example, C and D probably would not be small according to Test (c), but also would probably not be large according to Test (d). Test (d) is included because of the effect in Figure 15 where Method II was preferred over Method I if the minimum term was large. One possible interpretation is that if both terms were large, subjects were less likely to recall the numerical value of either term directly and choose Method II since in that method neither of the terms given in the problem must be translated to its numerical equivalent.

Now consider differences between subtraction problems with minus signs and missing addend problems. Comparing Figures 12 and 14, the pattern of choice between Method II and the other methods differed significantly between the two sets of problems, $\chi^2(5) = 18.68$,

$p < .005$. The difference consisted mainly in the greater frequency of choosing Method II to solve missing addend problems generally. Recall from Table 3 that this difference occurred, and was significant, where data from all subjects were analyzed. This could be produced by the procedure in Figure 10 in two ways. First, the test for size of gap applied to missing addends might be more lenient than the test applied to subtraction problems; Test (b) might have a positive outcome more often than Test (e). A second factor is the inclusion of Test (d), which leads to use of Method II for some problems in which the test for gap size has a negative outcome.

Choices between Method I and Methods III and IV did not depend on gap size differently in addition and subtraction patterns, $\chi^2(3) = 1.09$, $p > .70$. Recall that these choices were apparently independent of gap size for both sets of problems.

Next, consider the dependence of solution method on the minimum term of the problem, comparing Figures 13 and 15. First, the choice of Method III or IV versus the other methods had a different pattern in the missing addend problems from that in the subtraction problems with minus signs, $\chi^2(2) = 6.89$, $p < .05$. In terms of the model of Figure 10, this suggests that the test applied to missing addend problems, Test (c), was a more stringent test, providing a sharper discrimination than the test applied to subtraction problems with minus signs, Test (f).

Finally, the choice between Method I and Method II depended on the minimum term in different ways between missing addend problems and subtraction problems with minus signs, $\chi^2(4) = 16.99$, $p < .005$. This is accounted for in Figure 10 by the inclusion of Test (d), which makes the choice between Method I and Method II depend on the size of the minimum term in missing addend problems, but there is no corresponding test in the procedure applied to subtraction problems with minus signs.

The relationship between the composite model of Figure 10 and the models of individual subjects' performance is fairly straightforward. Although the composite model omits some details that appear in one or more of the individual models, it is approximately a combination of the components of the various individual models. A simple combination of components would occur if for each feature of the problems, individual subjects either ignored the feature or used it in a way that was consistent with the other subjects who used the feature. If choice of methods depended on a feature such as gap size in one way for some subjects and in the opposite way for other subjects, then pooling across subjects could have the effect of cancelling out effects that were present for individuals and the composite model would appear simpler than the individual subject models. This apparently did not occur to any large extent. Instead, the combination of data across subjects appears to have reinforced most of the effects that were observed in the performance of individuals and when the combination of effects is taken into account, the composite model is more complex than any of the individual models.

The composite model can be considered as a generalization of the individual models in the sense that models of individuals can be obtained by selecting specific values of parameters of the composite model. For each of the features tested in the composite model, each individual has some probability of applying the test and if it is applied there is some criterion-value of the feature that determines which outcome the test will have. Each feature test can thus be characterized as a probability distribution of test outcomes (yes versus no) over the set of values that the feature has in the set of problems, and this probability distribution varies from subject to subject. Another source of variation among individuals would involve the sequence in which features are tested, perhaps reflecting salience of the features.

When individuals are compared with the composite model, most of the differences involve absence of features in the individual models that appear in the composite model. This could be represented in most cases as a probability of one for one of the outcomes. For example, the model in Figure 5 for Subject S6 after practice is considerably simpler than the composite model. Rather than Test (a), there is uniform choice of Method II; this would correspond to a probability of one of an outcome of yes at Test (a) in the composite model. Subject S6's model would also require that the tests for a small term in subtraction problems, Tests (c) and (f), should precede the tests for gap size, and that those tests, Tests (b) and (e), should have probability one of a yes outcome. This would produce the performance of Subject S6 in which Method III or IV was chosen when a small term was present, and that Method II was chosen uniformly otherwise.

Figure 3 provides another example that clarifies the relationship between the composite and individual models that we inferred. Test (a) in the composite model appears in the model for Subjects S5 and S8. The bias toward Method I could easily be represented as a property of the probability distribution corresponding to the test. For missing addend problems Test (b) in the composite model appears as Test (b) in the individual model, but if the outcome is no, Method I is uniformly chosen. This would require that Tests (c) and (d) in the composite model are nonfunctional for S5 and S8. For subtraction problems with minus signs Test (e) in the composite model would be nonfunctional in the model for S5 and S8 and Test (f) would be present and is shown as Test (c) in Figure 3.

The various models of individual subjects' performance all can be obtained from the composite model shown in Figure 10 in the manner illustrated by the two models described above. That is, the individual models that we inferred are obtained by deleting some tests and by performing some changes in the order in which tests are performed in the model shown in Figure 10.

Conclusions

A major objective of this study was the development of models representing procedures used by different individual subjects. The models that we developed provided a reasonable account of the performance of the various subjects. The data base for each model, consisting of performance on 54 problems, does not permit very firm conclusions that the features included in each model are an exact characterization of the subject's congitive processes. More data would probably show that some features affected performance on a very small class of problems and could be neglected, or that other features should be included to account for performance not observed with the problems we used. On the other hand, a fair range of performance was observed, and it may be reasonable to conclude that the kinds of differences between subjects that occurred in this experiment are representative of the nature of individual differences on this task.

The models that we developed differ noticeably, but are similar in important respects. In all the models, solution methods are chosen on the basis of some or all of a small set of features: whether the sign in the expression is plus or minus, the location of the unknown term in the expression, the size of the smaller term given in the problem, and the size of the gap between terms of the problem. We do not have strong evidence that these are the only features subjects use in selecting solution methods, but the models based on these features seem to give a reasonable account of the individual data. Furthermore, when the data of different subjects are pooled, choice of solution methods relates to the sizes of gaps and the minimum terms in systematic ways, suggesting that the inferences made about individual subjects were not completely fortuitous. A composite model developed to account for general trends in the pooled data may be considered as a kind of prototypic procedure. It seems reasonable to consider the individual subjects' procedures as variations on the same general theme, differing in detail but all based on the same general principles.

The results of this study are generally consistent with conclusions that have been taken from studies of numerical computation. Of course, in numerical computation the method of translating and adding or subtracting does not occur. However, the occurrence of Method II for addition problems is consistent with observations that in adding numbers, reaction time is approximately a linear function of the smaller addend (Groen & Parkman, 1972; Groen & Resnick, 1977).

The relation between our results and the previous findings concerning numerical subtraction are slightly more complex. The finding is that for subtraction problems of the form $\underline{s} - \underline{n} = ?$ (Woods, Resnick, & Groen, 1975) and for missing addend problems of the form $\underline{m} + ? = \underline{s}$ (Groen & Poll, 1973), reaction time is approximately linearly related to the minimum of two terms, the size of the term to be subtracted ($\underline{m}$ or $\underline{n}$), and the size of the gap between the two terms. It seems unlikely that subjects determine both the sizes of the minimum term and the gap before choosing a method of solution, especially when the size of the gap is the answer. However, tests of the kinds included in our models would also lead to the observed data. Method III and Method IV require an amount of processing roughly proportional to the size of the term being subtracted. In subtraction problems, Method II requires an amount of processing roughly proportional to the size of the gap between the term. If Method III or Method IV was chosen when a small term is to be subtracted, and if Method II was chosen when there is a small gap, then in a majority of problems the smaller of these two quantities would determine the time required for processing.

We have not discussed one problematic aspect of our models that relates closely to the process of subtracting numbers as well. The models include tests for the size of a gap between terms, and we assume that small gaps can be detected prior to counting their exact sizes. A possible mechanism for this is related to recent proposals (Banks, in press; Banks, Fujii, & Kayra-Stuart, 1976; Kosslyn, Murphy, Bemesderfer, & Feinstein, 1977) that we have information about ordered objects (presumably including letters) that is hierarchical in nature. For our experiment, suppose that the letters A - R are stored in five or six categories, perhaps with A, B, and C classed as very small, D, E, and F classed as small, G, H, I, and J classed as medium, and so on. A mechanism for identifying small gaps would be a test to see whether the two terms of a problem are in the same category. This would not lead to a perfectly consistent classification. A pair of terms could be in adjacent categories and be classed as far apart, yet be closer than some pairs that came from single categories. However, the judgment would be strongly correlated with actual gap size. In reality, it is unlikely that the boundaries between categories are firmly fixed, and additional information such as adjacency of letters probably is available regardless of the category boundaries. On the other hand, it seems quite likely that subjects in our experiment could have used knowledge about the general locations of letters in the alphabet, and that seems to provide a plausible mechanism that would permit small gaps to be identified.

A question on which we hoped to get information in this study is the process by which the syntactic structures of various problem formats are identified. The various forms of problems involving unknowns in different locations have been used in school arithmetic instruction, presumably to strengthen students' understanding of the meaning of arithmetic expressions. Our procedural models do not include components that represent parsing in any interesting way. Instead, the decision to add or subtract the terms is made by a simple pattern recognizer that identifies an addition problem if there is either of two specific configurations involving the sign and the location of the unknown, and performs a subtraction procedure in the other cases. It would be very hard to prove that procedures used by human problem solvers are as simple as those in our models, but the simple process that we have included is sufficient for the task and seems to agree well with our own introspections when we perform the task.

Another formal issue that is related to our findings involves the commutative property of addition and the inverse relation of addition and subtraction. Commutativity is presupposed by Method II for addition problems, since the larger term is used as the starting point of the count, regardless of its position. Our findings do not add significant information to that already available on this point (Groen & Parkman, 1972; Groen & Resnick, 1977), except perhaps that our failure to find differences between $\underline{m} + \underline{n} = ?$ and $? - \underline{n} = \underline{m}$ may indicate that the recognition of commutativity is independent of the format in which an addition question is given.

On the other hand, there is interesting information in our results concerning the inverse relation of addition and subtraction. A simple statement of the inverse property is that $\underline{m} + \underline{n} = ?$ and $? - \underline{n} = \underline{m}$ are alternative statements of the same question, and similarly, $\underline{s} - \underline{n} = ?$, $? + \underline{n} = \underline{s}$, $\underline{s} - ? = \underline{m}$, and $\underline{m} + ? = \underline{s}$ all ask the same question. It might be expected, then, that the same procedure would be used to find the answers in all cases involving the same question. Our results are consistent with that possibility in the case of addition problems $\underline{m} + \underline{n} = ?$ and $? - \underline{n} = \underline{m}$. However, there apparently were differences, at least in detail, between procedures used for solving subtraction problems and missing addend problems. The method of counting the size of the gap, Method II, was used more frequently for missing addend problems, suggesting that the gap between the terms might be apprehended more clearly or directly for missing addend than for subtraction sentences. A possibility that seems worth considering is that a valid reason for including the various forms of incomplete number sentences in school instruction might relate to increasing children's flexibility in retrieving arithmetic facts with varying retrieval cues. It seems unlikely that these tasks provide significant strengthening of children's understanding of the syntax of arithmetic, but a flexible retrieval system related to various forms of questions could be considered a significant operational form of understanding the inverse relationship between addition and subtraction.

References

Banks, W. P. Encoding and processing of semantic information in comparative judgments. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 11). New York: Academic Press, in press.

Banks, W. P., Fujii, M., & Kayra-Stuart, F. Semantic congruity effects in comparative judgments of magnitudes of digits. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 546-447.

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

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

Groen, G. J., & Resnick, L. B. Can preschool children invent addition algorithms? Journal of Educational Psychology, 1977, 69, 645-652.

Kosslyn, S. M., Murphy, G. L., Bemesderfer, M. E., & Feinstein, K. J. Category and continuum in mental comparisons. Journal of Experimental Psychology: General, 1977, 106, 341-375.

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