J.P. Hill (Ed.), Minnesota Symposia on Child Psychology. Minneapolis: University of Minnesota Press, 1967, pp. 160-200.

◈ PATRICK SUPPES, LESTER HYMAN, AND MAX JERMAN ◈

Linear Structural Models for Response and Latency Performance in Arithmetic on Computer-Controlled Terminals

IN THE cognitive domain mathematics provides one of the clearest examples of complex learning and performance, for the structure of the subject itself provides numerous constraints on any adequate theory. The learning and performance models derived from the main trends of contemporary mathematical learning theory have provided an excellent predictive analysis of a large variety of experimental situations. Unfortunately, however, most of these experimental situations are much simpler in structure than even the simplest corresponding parts of elementary mathematics. Because this claim is central to the motivations behind the present paper, we should like to expand on it in some detail.

NOTE: The research reported here has been supported by the U.S. Office of Education, the National Science Foundation, and the Carnegie Corporation of New York. Data analysis was supported by the U.S. Public Health Service, Grant MH-6154. We are indebted to Guy Groen for a number of useful comments, to Manuel Uy for extensive and much appreciated efforts in executing the computer programs used in the data analysis, and above all to Dow Brian and his staff of computer programers, particularly John Gwinn, in the Computer-Based Laboratory for Learning and Teaching of the Institute for Mathematical Studies in the Social Sciences, Stanford University. Experiments of the magnitude reported here would have been a practical impossibility without the hardware and software facilities of the Laboratory. We are also indebted to Miss Luanne D. Berkowitz, who assumed major responsibilities in running the experiments and in helping construct the curriculum materials, and to Mr. Victor Norton, the principal of Grant Elementary School of Cupertino Union School District in California, as well as to the teachers of Grant, all of whom contributed substantially to the rather hectic beginning stages of initiating a new technology in an elementary school setting.

The familiar and now classical linear model is a good starting point. We may take this model in its simplest form, as applied to a situation in which a given response is always reinforced, and all other responses are indicated as incorrect. For the formulation under this restriction let q_{n+1} be the probability of an incorrect response on trial $n + 1$. This probability is then the following simple linear function of the probability q_n of an incorrect response on trial n: $q_{n+1} = \alpha q_n$, where the learning parameter α is such that $0 \leqq \alpha < 1$. The formal properties of this simple model have now been investigated thoroughly and are well understood. It is apparent, however, that if the subject must learn a number of different items that vary in structure and therefore in learning difficulty, the simple linear model can accommodate this fact only by separately estimating a learning parameter α for each item. From the standpoint of classical psychological concerns with the character of learning and performance, this is far from satisfactory. What is desired, rather, is an analysis of the factors in the structure of the stimulus item which lead to varying difficulty. The estimation of a nonstructural parameter unique for each item is a way of handling data when no better resources are available, but it does not take us very deep into the psychological problems of learning complex items like those common in mathematics and other structured subjects. Above all, the estimation of a separate parameter for each item leads to a wasteful use of parameters. In general, if we take a collection of items from a given domain of mathematics, we should like to be able to attach weights to the various factors that may be objectively identified in the item, and then use estimates of a few such weights to predict the relative difficulty or the latency of response for a large number of items. The linear model itself cannot provide such mechanisms. This is not to denigrate the importance and significance of the linear model, for it will doubtless enter in many places to provide an analysis of particular mechanisms. But it cannot serve as the basis for a fundamental or general theory of complex learning.

At first glance, it might appear that we could use a learning theory with more structure, such as stimulus-sampling theory, to provide an adequate analysis of stimulus structure — adequate to make differential predictions of difficulty in cognitive domains like that of elementary mathematics. An examination of the explicit axiomatizations of stimulus-sampling theory, which may be found in Estes and Suppes (1959), Suppes and Atkinson (1960), or Atkinson and Estes (1963), shows, however, that the concept of stimulus used does not provide an adequate analysis of structure.

Roughly speaking, the situation is the following: The stimuli presented to a subject on a given trial are represented by a set of stimulus elements. In the concept of an arbitrary set of stimulus elements, there is the beginning of an adequate apparatus for the concept of structure, but the additional assumptions needed for a definite notion of structure have not yet been imposed on the concept of an arbitrary set. It is necessary to go beyond the current formulations of the theory in order to analyze even the simplest kind of stimulus items used in the teaching of elementary mathematics. Probably the most successful version of stimulus-sampling theory for a wide variety of experiments is the pattern conception of stimulus conditioning that originates with Estes (1959). In this theory, the individual stimulus elements are not conditioned as components to a correct response; rather, an entire pattern of stimulus elements is so conditioned, and in general, the number of patterns available for sampling in a given stimulus situation will be a parameter to be estimated from the data. But even these conceptions are far from providing an analysis sufficiently structured to yield differential predictions of difficulty in responding correctly to problems drawn from concepts and topics in elementary mathematics.

It might also be thought that the applications of stimulus-sampling theory or related kinds of theories to stimulus-discrimination problems during the past decade would yield theoretical ideas adequate to the analysis of complex structure. Again, however, an examination of the kinds of problems that have been handled shows quickly that a structural apparatus adequate to problems in simple addition, for example, is certainly not even implicitly inherent in theories of conditioning.

Both psychologists and educators interested in cognitive theories of learning would undoubtedly agree with the remarks we have just made about stimulus-response theories. However, we find that we must say much the same things about the current cognitive theories of learning and performance, which have attracted considerable interest in the last few years. As opposed to the stimulus-response theories that we have mentioned, perhaps the greatest defect of the cognitive theories is simply that they are not specific enough even to settle the question of whether or not specific predictions can be made. The kinds of cognitive considerations, for example, that enter into the studies reported in the well-known book by Bruner, Goodnow, and Austin (1956) simply do not provide a framework within which we can ask specific questions about the estimation of parameters for the prediction of differential difficulty over a selection of stimulus items drawn from some complex domain, whether it be elementary mathematics or elementary language learning. Again we would not want to be misunderstood on this point. The analysis of the types of strategies used in concept attainment is certainly a useful contribution to the psychology of concept formation and thinking, but it must be realistically asserted that no theory has yet been sufficiently developed to provide the kind of parametric predictions that are considered a minimum requirement in the area of mathematical models of learning and performance. The same remarks apply to the invaluable work of Piaget and his collaborators. Piaget has contributed much to our understanding of cognitive development in children and especially to our understanding of the kind of structures children find or, if you wish, create, in the stimulus environment. But again, Piaget's concepts have not been sufficiently articulated into a well-defined theory to provide parametric predictions of differential difficulty for items drawn from any cognitive domain. This is not particularly Piaget's task, as it was not Bruner's. Nevertheless, we do intend our remarks to be of a critical nature, for until parametric predictions can be produced from the theoretical proposals generated by various psychologists, these proposals cannot be accepted as a final analysis of what we hope to understand about cognitive processes.

The preceding remarks have mainly emphasized the inadequacy of current psychological theories to provide parametric predictions of differential difficulty as measured by the rate of correct responding. These theories are even more inadequate when we turn to response latencies. From the standpoint of the analysis of performance, latencies are in many respects more important as a source of information to the theorist than as response data. This is particularly true of any studies devoted to skill after a good deal of learning has taken place. As some of the data reported here show, and as one would expect anyway on a priori grounds, the range of latencies observed in a group shows systematic variation in a way that clearly reflects a measure of item difficulty. What is ultimately desired in this case is the kind of model that can predict from the structure of an item the process a subject must go through in finding the correct response. In the case of arithmetic, at least part of this process must undoubtedly be related to the standard algorithms taught as part of the curriculum; but even a casual glance at these algorithms will show that the conception of them used in teaching and in the curriculum does not provide a sufficent analysis of processing to make differential predictions of difficulty as reflected in response latencies.

What is also surprising about latency is that there have been so few studies that report detailed data on this measure. The only directly relevant studies that we have found in the literature on arithmetic are Batson and Combellick (1925), Helseth (1927), Knight and Behrens (1928), and Billington (1947). This absence of latency studies (even though there are undoubtedly several of which we are unaware) indicates how superficial has been the investigation of structural models adequate to predict differential difficulty in terms of either responses or response latencies.

The constructive aim of the present paper is to formulate and test some linear structural models that do lead to parametric predictions of the kind we have been discussing. These models are not linear in precisely the same sense that the linear learning model is, but they are in the context of linear regression models, a point that is made clear in the next section. The models and accompanying theory which we present and test in this paper are meant only as a beginning. We do believe that they provide a significant and promising foundation for further work.

The Theory

The learning models that arise in stimulus-sampling theory all exemplify a certain class of stochastic processes, and in general, a different class of such processes is exemplified by the linear models discussed at the beginning. In the same fashion, the linear structural models proposed in this paper all exemplify a general class of models that are classical in statistics. But simply to say that we are applying linear regression models to the study of arithmetic performance provides no more clue to the theories behind the analysis than does the assertion that we apply to a given body of learning phenomena a finite Markov chain as the primary mathematical tool of analysis. What is important and significant for psychology is the particular way in which the broad class of linear regression models is narrowed and made meaningful from the standpoint of response or latency performance in arithmetic.

It will perhaps be useful to begin with a class of problems that are simpler than those considered here in detail. The discussion of this first example follows Suppes (1967). Let us suppose that a set of problems consists only of simple addition problems of the following sort: $1 + 2 = n$, $1 + n = 3$, and $n + 2 = 3$. Let us restrict the sums to those not greater than

5. We postulate that five facts are held in memory: 1 = /, 2 = / /, 3 = / / /, 4 = / / / /, and 5 = / / / / /. Our algorithm is then: (1) Replace all Arabic numerals by their stroke definitions and delete all plus symbols. (2) If there are strokes on both sides of the equal sign, cancel one by one, starting from the left of each side until there remain no strokes on one side. Ignore n in canceling. (3) On the one side still having strokes, replace the strokes by an Arabic numeral, using the definitions in memory. The solution in the form n = c or c = n will result. To obtain a single factor f representing the number of steps, we simply count the number of steps required by the algorithm to solve a given problem. For example, the steps to solve 3 + n = 5 are five: (a) / / / n = / / / / / by rule (1); (b) / / n = / / / / by rule (2); (c) / n = / / / by rule (2); (d) n = / / by rule (2); (e) n = 2 by rule (3). Thus for this model and this problem, f = 5. A more realistic version of this algorithm, at least for many standard situations in which students are tested on their command of the simple facts about addition, is to postulate that the student counts the difference n by beginning at 3 and stopping at 5. A test of five variants of the latter counting algorithm is reported in Suppes and Groen (1967).

For the problems analyzed in this paper, the central difficulty is to identify the factors that contribute to the complexity of the problem. Typical factors that we shall examine are the magnitude of the largest number appearing in the problem, the magnitude of the smallest number, the form of the equation in which the problem is presented, and most important, the number of steps required to solve the problem. Exactly how the number of steps is to be defined is a matter that we take up in detail below. As a matter of notation we shall denote the j^{th} factor of problem i in a given set of problems or exercises by f_{ij}. The statistical parameters that must be estimated from the data are the weights to be attached to each factor. We shall denote the weight assigned to the j^{th} factor by α_j. We want to emphasize as explicitly as possible that the factors identified and used in the models presented in this paper are never factors in the sense of factor analysis — that is, the factors do not arise as abstract constructions from the data; rather, they are always objective factors identifiable by the experimenter in the problems themselves, independent of any data analysis. Which factors turn out to be important is a matter of the estimated weights α_j, but in no case does the decision about the numerical value of a factor for a given problem depend in any way on the response data themselves. In fact, it will be apparent that all of the factors used in the analyses presented here have an intuitive and direct relevance to commonsense ideas of difficulty, and their definitions are so straightforward and simple that there is little prospect of disagreement over their objective value in a given problem.

We may first consider the analysis of response data. Let p_i be the observed proportion of correct responses on problem i for a given group of subjects. The central task of a model is to predict the observed proportions p_i. The natural linear regression model in terms of the factors f_{ij} and the weights α_j is simply $p_i = \Sigma_j \alpha_j f_{ij} + \alpha_0$. However, there is a central difficulty with this particular model: there is no guarantee that probability will be preserved as the estimated weightings and identifiable factors are combined to predict the observed proportion of correct responses in new items. Consequently, in order to guarantee preservation of probability, that is, to ensure that the predicted p_i's will always lie between 0 and 1, it is natural to make the following transformation and to define a new variable z_i,*

$$z_i = \log \frac{1 - p_i}{p_i}. \qquad (1)$$

And then to use as the regression model

$$z_i = \Sigma_j \alpha_j f_{ij} + \alpha_0. \qquad (2)$$

It should be noted that the reason for putting $1 - p_i$ rather than p_i in the numerator of equation (1) is that it is desirable to make the variables z_i monotonically increasing in the magnitude of the factors f_{ij} rather than monotonically decreasing. For example, the magnitude of the largest number in a problem increases with the difficulty of the problem, and it is desirable that the model reflect this increase directly rather than reflect it inversely.

In the case of latencies a transformation like (1) is not required. Let t_i be the mean latency on problem i for a given group of subjects. We then apply the same model as (2), namely,

$$t_i = \Sigma_j \beta_j f_{ij} + \beta_0. \qquad (3)$$

It is also evident that no transformation is required to make latencies monotonically increasing in the expected difficulty of the items. We have shown different weights β_j for the latencies, because the empirical interpretation of the weights must necessarily be different for the variables z_i,

* To take care of the case when the observed p_i is either 0 or 1, we use the transformation $z_i = \log(2n_i - 1)$ for $p_i = 0$, and $z_i = \log[1/(2n_i - 1)]$ for $p_i = 1$, where n_i = the total number of subjects responding to item i. The exact form of this transformation is not important.

but as we should expect, there is a high positive correlation between the weights α_j and β_j. It is worth noting that, in the case of the analysis of the latencies, the individual factors and their weights may be identified as the direct contribution of a given factor to the total latency. Thus, for example, the contribution of factor j to the total latency is just $\beta_j f_{ij}$, which is scaled in seconds. The additive constant β_0 that arises in the linear regression model may be interpreted as the constant orientation and preparation time required in solving the problems of the class under investigation.

The variables we consider are of two sorts. The first is the kind of 0,1-variable standard in the analysis of variance. Such a variable would be appropriate, for example, in dealing with problem format. The second kind of variable is one that is in principle continuous, although in practice it assumes a finite set of values for the problems being considered here. For most of these variables the conception and formal definitions of the variables are quite straightforward within the context of elementary arithmetic itself. Typical variables have already been mentioned; however, the variable or factor dealing with the number of steps required to solve a problem is most important from the standpoint of the psychological analysis and also seems most promising for future developments of the models presented in this paper. We turn now to the appropriate formal definitions. As has already been emphasized, we believe that the greatest possibilities for subsequent theoretical analysis lie in this direction. What we report here is the result of only our first, relatively crude analysis, and we are already engaged in deepening this analysis, particularly by breaking up the single variable of number of steps into several components. Some preliminary results are reported at the end of the paper.

The steps postulated have been broken up into three classes: those required to transform the problem into canonical form, those corresponding to the number of operations performed, and those corresponding to the number of digits that must be held in memory. We refer to these three as the transformation, operation, and memory classes. As will be seen, there is a quite high correlation in most problems between the number of operation steps and the number of memory steps. An essential point for later work is to make these two processes more orthogonal in operational characterization. Another assumption that is surely too simple is reflected in the assignment of the same weight to addition and subtraction, in the analysis of operation steps. Other unrealistic simplifications have been made, but the general definitions required to characterize the number of steps to a solution are still relatively complex, and we think they constitute a reasonable beginning.

For simplicity we first consider just the transformation steps that convert any problem into canonical form. By canonical form we mean the equational form in which the blank or unknown stands by itself as the only term to the right of the equal sign. Thus for numbers m, n, and p, regardless of whether they are one digit or two, we have: (i) $m \pm n = \_\_$ is already in canonical form; (ii) $m + \_\_ = p$ is transformed to $\_\_ = p - m$, which is transformed to $p - m = \_\_$, requiring two steps; (iii) $\_\_ + n = p$ is identical to (ii); (iv) $m - \_\_ = p$ is transformed to $m - p = \_\_$, requiring one step; and finally (v) $\_\_ - n = p$ is transformed to $n + p = \_\_$, also requiring one step. The fact that $m + \_\_ = p$ requires one more transformation than $m - \_\_ = p$ agrees with the intuition that (ii) is really more difficult than (iv). We make explicit the number T of transformations in the following definitions that formalize (i)–(v).

$$T(m \pm n = \_\_) = 0$$
$$T(m + \_\_ = p) = 2$$
$$T(\_\_ + n = p) = 2$$
$$T(m - \_\_ = p) = 1$$
$$T(\_\_ - n = p) = 1$$
$$T(m + n = p + \_\_) = 1$$
$$T(m + n = \_\_ + p) = 1$$

The last two equations cover two additional cases that arise in the data we analyze.

Turning now to the operation and memory steps, we need to make explicit the number of digits involved, so we always use initial letters of the alphabet for single digits. Also, because we postulate that the operation and memory steps enter only after the transformation to canonical form has taken place, we may simplify the notation, writing, for example, $O(ab + cd)$ or $M(ab + cd)$ for the number of operation or memory steps respectively. For example, $O(5 + 0) = 0$, but $O(5 + 4) = 1$ because we postulate no operation is required for handling zero. Also, $O(15 + 12) = 2$ because one operation is $5 + 2$ and the second is $1 + 1$. On the other hand, in the form $ab + cd$, when $b + d > 9$, there are three operations. For example, $O(25 + 47) = 3$, because one operation is $5 + 7$; the second is the partial sum $1 + 2$ using the 1 that is "carried"; and the third is $3 + 4$, the partial sum plus 4, the other tens' digit.

In the case of memory, $M(15+12)=1$, because only 7, the sum of 5 and 2, must be held in memory while the tens are added and the correct tens' digit response is made (the problem format required input of the tens' digit before the ones' digit). On the other hand, $M(25+47)=3$ because (i) the 2 of 12, the sum of 7 and 5, must be held in memory for the ones' response, (ii) the 1 which is carried to the tens place must be held, and (iii) the partial sum $1+2$ must be held while it is added to 4. The definition for the more complicated format $ab+cd-ef$ is given recursively in terms of $ab+cd$, and thus does not need a separate treatment. In all cases considered, $a+c+1<10$. Formally the definitions of the number of operation and memory steps are as follows:

$$O(a+b)=0 \text{ if } a=0 \text{ or } b=0$$
$$O(a+b)=1 \text{ if } a\neq 0 \;\&\; b\neq 0$$

$$O(ab+d)=0(b+d) \text{ if } b+d\leqq 9$$
$$O(ab+d)=0(b+d)+1 \text{ if } b+d>9$$

$$O(ab+cd)=0(b+d)+1 \text{ if } b+d\leqq 9$$
$$O(ab+cd)=0(b+d)+2 \text{ if } b+d>9$$

$$O(a-b)=0 \text{ if } b=0$$
$$O(a-b)=1 \text{ if } b\neq 0$$

$$O(ab-c)=0(b-c) \text{ if } b\geqq c$$
$$O(ab-c)=0(b-c)+1 \text{ if } b<c$$

$$O(ab-cd)=1 \text{ if } d=0$$
$$O(ab-cd)=2 \text{ if } b\geqq d>0$$
$$O(ab-cd)=3 \text{ if } b<d$$

$$M(a\pm b)=0$$

$$M(ab+c)=1 \text{ if } b+c\leqq 9$$
$$M(ab+c)=2 \text{ if } b+c>9$$

$$M(ab+cd)=1 \text{ if } b+d\leqq 9$$
$$M(ab+cd)=3 \text{ if } b+d>9$$

$$M(ab-c)=1 \text{ if } b\geqq c$$
$$M(ab-c)=2 \text{ if } b<c$$

$$M(ab-cd)=1 \text{ if } b\geqq d$$
$$M(ab-cd)=3 \text{ if } b<d.$$

If $ab+cd=gh$, then $O(ab+cd-ef)=O(ab+cd)+O(gh-ef)$, and

$M(ab + cd - ef) = M(ab + cd) + M(gh - ef) + 1$. The additional step in the case of $M(ab + cd - ef)$ comes in from having to remember $a + c$, or $a + c + 1$, as the case may be, which is not part of $M(ab + cd)$ or $M(gh - ef)$.

Similarly, if $ab + c = fg$, then $O(ab + c - de) = O(ab + c) + O(fg - de)$, and $M(ab + c - de) = M(ab + c) + M(fg - de)$. A corresponding definition holds for $ab + cd - e$.

In evaluating problem structure, we determine the total number of transformation, operation, and memory steps. Thus, for example, $25 + 26 = 18 + \_\_$ has the maximum number of fourteen steps, because $T(25 + 26 - 18 + \_\_) = 1$, $O(25 + 26 - 18) = 6$, $M(25 + 26 - 18) = 7$; and on the other hand, the problem $5 + 0 = \_\_$ has the minimum of 0 steps. Of course, some students will solve many individual problems by a shorter method, and the present approach to counting steps does not incorporate any such special methods. This again is a matter for subsequent investigation.

In the analyses reported in this paper we have entered the total number N of steps as a single variable for most of the results reported, but in one case we have broken the steps up into classes, and further work in this direction is under way. In the linear regression models used for this purpose, we replace αN by $\alpha_1 T + \alpha_2 O + \alpha_3 M$. Generally, however, the single variable, total number of steps (NSTEPS), and two magnitude variables are the factors that enter in the regression analyses in this paper.

Method

The data reported and analyzed in this paper were collected as an integral part of an operational program in computer-assisted mathematics instruction lasting a full academic year. For this reason we shall describe in some detail this program.

Subjects. The approximately 270 subjects in this project consisted of the entire population of grades three, four, five, and six in the Grant Elementary School, except for those in the handicapped classes.

The children came from a middle-class, suburban community in California. All children lived within walking distance of the school.

Although there was some fluctuation in attendance figures during the year, school records show the following population at year's end: There were 32 boys and 30 girls in grade three, 41 boys and 35 girls in grade four, and 44 boys and 26 girls in grade five. The mean IQ of the fifth-grade group was 114, the range 72–145. Grade six had 35 boys and 27 girls. Mean IQ of the sixth-grade class was 117, range 88–156. There were no IQ scores for either grade three or grade four.

Equipment. The student terminals used in this project were commercially available teletype machines, connected by private, high-speed, telephone lines to a computer in the Institute for Mathematical Studies in the Social Sciences at Stanford University. A large book closet, which opened into the classroom, was modified by adding a ventilative fan, light, and electrical outlet. This provided privacy for the user and insulated the rest of the class from the operational noise of the teletype.

The control functions for the entire system were handled by a medium-sized computer. The PDP-1 has a 16,000-word core, and a 4,000-word core which can be interchanged with any of thirty-two bands of a magnetic drum. All input-output devices are processed through a time-sharing system. Two high-speed data channels permit simultaneous computation and servicing of peripheral devices. Additional backup in computational power, additional storage, and increased input-output speed are obtained through connections to disk storage of a larger computer (IBM 7090) located at the Stanford Computation Center.

Response time was measured from the instant (nearest .001 second) the type wheel was in position at the response area (or answer blank). When the student depressed one of the keys on the teletype, a signal was sent to the computer. The character was recognized by the computer approximately 1 millisecond (.001 second) after being initiated by the student. A reading was taken from a real-time clock, internal to the computer, and this information compared with the time recorded when the type wheel was positioned. Under optimal conditions latency measurements could be made with an accuracy of from 2 to 3 milliseconds. However, as mentioned above, the system was operating under a time-sharing arrangement, and this reduced the level of accuracy of latency measure to about one tenth of a second. Conversion from readings in thousandths to the nearest tenth was made by division and truncation.

Curriculum Materials. Daily lessons were prepared and organized by concepts or topics into blocks or units. The concept blocks were arranged sequentially and corresponded approximately to the order of topics in the textbook series *Sets and Numbers* (1966) by Suppes. The length of time needed to complete a concept block varied from three to twelve days when a single lesson was taken each day. The curriculum objective of the daily lessons was to provide an organized program of review, maintenance, and drill on basic skills and concepts of elementary mathematics, particularly arithmetic. Instruction in all concepts was given initially by the teacher, and consequently the drill-and-practice work at computer terminals did not include a detailed introduction to the concepts.

Teachers in the project could, subject to certain constraints, select any of the prepared blocks in order to correlate the drill-and-practice work with their daily instruction. Handbooks were furnished which described available concept blocks in detail. Also included in the handbooks were reprints of every lesson. Table 1 describes the concept blocks prepared for each grade level.

Each concept block was organized in the manner shown diagrammatically by Figure 1. Lessons were prepared at each of five levels of difficulty within each concept block. Among factors that determined intuitive estimates of relative difficulty are those discussed in this paper. They, and the exercises reflecting them, were chosen intuitively on the basis of teaching experience and previous project experience gained from preparation and testing of the textbook series cited above. Each class was restricted to a single concept block at a time. On the first day of a new block, every member of a class was given the same lesson. This lesson was of average difficulty (level 3). Those students who scored between 60 and 79 per cent were given a level-three lesson the following day; those who scored above

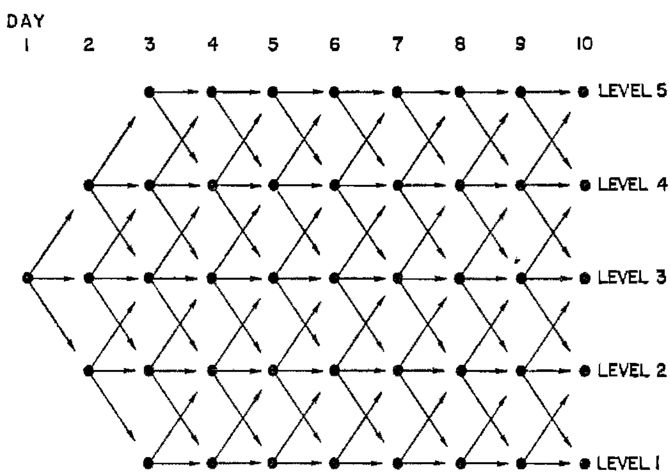

Figure 1. Diagram of branching structure followed in constructing sets of exercises for concept blocks.

Table 1. Concept Blocks for Grades 3–6, with Number of Days Spent on Each Block

Blocks	Grade 3		Grade 4		Grade 5		Grade 6	
	Days	Description	Days	Description	Days	Description	Days	Description
1....	8	sums 0–20	5	sums 0–40	5	sums 11–60	5	mixed drill, all operations
2....	5	differences 0–20	5	differences 0–40	5	differences 11–60	10	fractions, addition, subtraction, changing terms
3....	3	mixed addition & subtraction	5	sums 31–70	10	multiplication tables 3–12	5	multiplication tables 2–12
4....	5	multiplication tables 2 & 3	5	differences 31–70	5	multiplication tables, $c = a \times b$	4	factors, multiples, primes
5....	4	mixed addition & subtraction	7	multiplication tables 4–10	5	long division	5	fractions, simple equations
6....	5	word problems,[a] mixed review	5	division tables 6–12, $a/b = c$	7	mixed drill, all operations, word problems	10	achievement tests
7....	3	mixed review, addition, subtraction, multiplication	5	mixed review, all operations, inequalities	10	fractions	8	factors, multiples, fractions
8....	10	addition with carrying	8	CAD[b] laws for multiplication, addition, subtraction	5	units of measure	5	multiplication of large numbers
9....	10	subtraction, regrouping	5	multiplication tables 6–12	6	CAD laws	5	word problems, all operations
10....	3	money, equivalence[a]	5	mixed drill, all operations	5	CAD laws, giving reasons	3	long division, standard form, 1-digit divisor
11....	5	mixed review	5	word problems	3	word problems	8	decimal & common fractions, per cent[a]

[a] Blocks planned but not written.
[b] CAD stands for commutative, associative, and distributive.

Table 1 – continued

Blocks	Grade 3		Grade 4		Grade 5		Grade 6	
	Days	Description	Days	Description	Days	Description	Days	Description
12....			4	CAD laws [a] addition, subtraction, multiplication	5	CAD laws, giving reasons	4	metric units of measure [a]
13....			6	distribution law for division,[a] CAD laws	9	mixed drills, all operations	10	mixed drills
14....	5	multiplication tables 0–5	5	mixed drills, all operations	5	fractions,[a] word problems	4	logic
15....			5	subtraction, fractions	5	division	5	fractions,[a] addition, subtraction, multiplication, division
16....			5	fractions,[a] addition, subtraction	2	addition and subtraction of decimals	5	decimal operations [a]
17....			5	multiplication	5	addition, subtraction of integers	4	word problems [a]
18....			10	multiplication	6	exponents, word problems [a]	8	mixed review, all operations [a]
19....			15	mixed multiplication, subtraction [a]	2	metric measure	5	CAD laws
20....			15	mixed multiples of 10 [a]	10	mixed drill, all operations	10	long division, ladder form, 1- & 2-digit divisors
21....			2	fractions, addition, subtraction [a]	8	multiplication, exponents [a]	4	metric units of measure
22....			5	mixed drills,[a] all operations	3	coordinate systems	5	mixed review, negatives, inequalities [a]
23....			5	word problems	9	sets, review [a]	8	word problems giving reasons [a]
24....	5	multiplication tables	5	negatives, addition, subtraction [a]	5	word problems [a]	10	long division, ladder form, 1- & 2-digit divisors
25....	7	3- & 4-digit column addition	5	mixed addition, subtraction [a]	5	CAD laws, giving reasons	5	mixed review [a]
26....			5	CAD laws[a]	10	logic	5	vertical subtraction

Table 1 – continued

Blocks	Grade 3		Grade 4		Grade 5		Grade 6	
	Days	Description	Days	Description	Days	Description	Days	Description
27....			3	mixed multiplication, division [a]	5	per cent	2	special addition
28....			5	word problems, units of measure [a]	10	achievement tests	2	special multiplication
29....			4	mixed review, all operations [a]	5	vertical subtraction	12	long division, ladder form, 1-digit divisor
30....			2	CAD laws [a]	2	special addition	7	long division, standard form, 2-digit divisor
31....	10	multiplication to 12 × 12, vertical form	4	word problems, units of measure [a]	2	special multiplication	7	long division, ladder form, 2-digit divisor
32....	5	CAD laws for addition, subtraction, multiplication	10	achievement tests	5	long division, ladder form, 2-digit divisor	2	division tests, multiple-choice form, basic concepts
33....	5	division facts to 12 × 12 [a]	5	column subtraction				
34....	5	3- & 4-digit column addition with regrouping [a]	2	special addition				
35....	5	mixed review	2	special multiplication lessons				
36....	2	special addition	10	remedial multiplication tables 3–7				
37....	2	special multiplication	10	remedial multiplication tables 4–9				
38....		column multiplication [a]	5	division with variables, standard form				
39....			7	long division, ladder form, 1-digit divisor				
40....			7	long division, ladder form, 1-digit divisor				
41....			5	fractions				

79 per cent were given a lesson on the next higher level (level 4); those who failed to score at least 60 per cent were given a simpler lesson on a lower level (level 2). This procedure was followed throughout a concept block — that is, a score of above 79 per cent raised a student one level each day, whereas a score of below 60 per cent lowered a student one level each day, but of course a student could not move up beyond level 5 or down below level 1. Thus, by the third day, a student could have been at any one of five levels, with a different lesson at each level. It was intended that approximately 90 per cent of the students would alternate between levels 2, 3, and 4, and that those remaining on any level would be nearly homogeneous. Level 1 was mainly remedial, and level 5 was ordinarily meant to be difficult. Drills on all levels increased somewhat in difficulty from day to day within a block as successively more advanced aspects of each topic were reviewed.

Program Logic. Under computer control each problem was completely typed out, including a blank for the response. The type wheel of the teletype was then positioned at the blank so that the response would be properly placed. A correct response was reinforced by the appearance of the next exercise. When an incorrect first response was made, *wrong* was typed out, and the exercise itself was retyped. A second error on the same exercise was followed by the message "wrong, answer is __," with the correct answer being displayed. The exercise itself was then retyped once more to allow for a correction response. An error on the correction response caused the correct answer to be given again, but whether the third response was correct or incorrect, the next exercise was presented.

If a response was not given within a predetermined interval of time, usually ten seconds, the machine response followed the above pattern except that the words *time is up* were substituted for *wrong* at each step described above. A flow chart of the program logic is given in Figure 2.

Procedure. The two classes in each of grades four, five, and six began in October 1965, sharing one teletype between them. One class was scheduled to run in the morning, the other in the afternoon. However, this proved to be an unworkable arrangement. Beginning with the third week, the classes worked on the teletype on alternate days. The machines were operated daily between the hours of 8:30 A.M. and 3:00 P.M.

In late February 1966, the two third-grade classes began daily lessons with the addition of two more teletypes, which brought the total number of machines in operation daily to five. In early April 1966, three more machines were put in operation, bringing the total number of teletypes to eight. Each class in grades three, five, and six had its own teletype, but grade four had been divided into three classes to alleviate an overcrowded situation, so one of the fourth-grade classes had its own machine and the other two classes shared the remaining teletype.

The students took their lessons one at a time in the order prescribed by their teacher. The program began by asking the student to type his name. When the name had been typed correctly, the lesson began as described above in the section on program logic. If a student failed to spell his name correctly, or gave a fictitious name such as Batman, the program asked him to try again. An individual history was kept in computer memory for each student. When a student's name was typed correctly, the proper lesson was selected on the basis of the branching criteria and presented automatically. Students were free to sign on at any one of the machines in the school at any time during the day. It was also possible to take more than one lesson a day.

Lessons were designed to take from four to six minutes each, with an average of about five minutes, to allow each student in a class to take one lesson each day. The usual number of problems per lesson was twenty. Following the lesson, a summary of the student's work was given. A sample print-out of a lesson taken by a fifth-grade student, Mike O'Dell, is given

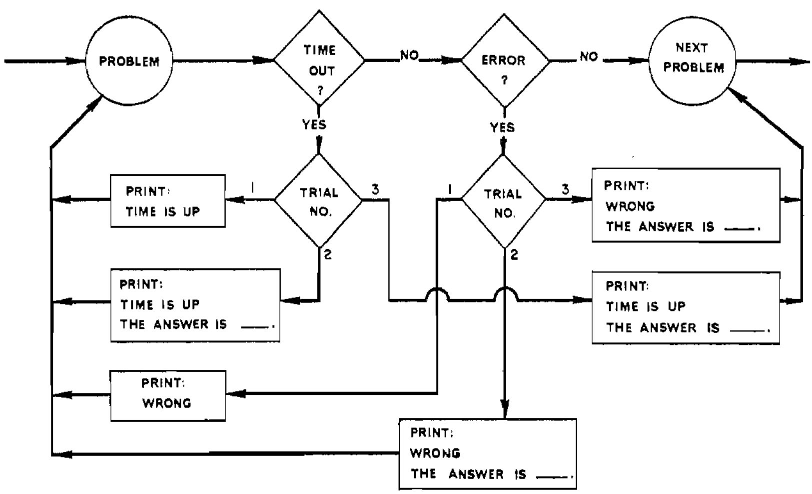

Figure 2. Flow chart of the program logic for presentation of problems and classification of responses.

Sample Print-out of a Lesson Taken by a Fifth-Grade Student

Please Type Your Name
Mike O'Dell
Drill Number 509013

$(42+63)/7 = (42/\underline{7}) + (63/\underline{7})$

$48-38 = 38-\underline{48}$

Wrong

$48-38 = 38-\underline{4}$

Wrong, Answer Is 28

$48-38 = 38-\underline{28}$

$76-(26-10) = (76-26)+\underline{10}$

$4\times(7+13) = (4\times\underline{7})+(4\times\underline{13})$

$(53-20)-11 = 53-(20+\underline{11})$

$32+(74+18) = (\underline{32}+74)+18$

$51\times(36\times 12) = (\underline{51}\times 36)\times 12$

$17\times(14+34) = (17\times 14)+(17\times\underline{34})$

$362+943 = 943+\underline{362}$

$(5+8)\times 7 = (\underline{5}\times 7)+(\underline{8}\times 7)$

$(90/10)/3 = \underline{90}/(10\times 3)$

$(72/9)/4 = 72/(\underline{9}\times 4)$

$(54+18)/6 = (54/6)+(18/\underline{\ \ })$

Time Is Up

$(54+18)/6 = (54/6)+(18/\underline{6})$

$60-(19-12) = (60-\underline{19})+12$

$72\times(43\times 11) = (72\times 43)\times\underline{11}$

$(63/7)+(56/7) = (\underline{63}+\underline{7})/7$

Wrong

$(63/7)+(56/7) = (\underline{63}+\underline{56})/7$

End of Drill Number 509013
13 May 1966
16 Problems

	Number	Percent
Correct	13	81
Wrong	2	12
Time Outs	1	6

Wrong
2
16
Time Outs
13
222.7 Seconds This Drill
Correct This Concept — 81 Percent, Correct to Date — 59 Percent
4 Hours, 46 Minutes, 59 Seconds Overall

Goodbye Mike

on page 178. The numbers in the summary for correct, wrong, and time outs are for first response only. The numbers following the word *wrong* are problem numbers. As shown, Mike had 81 per cent correct in this concept block and 59 per cent correct to date for the whole school year, which began October 18 on the teletype. The time given in hours, minutes, and seconds is the total time Mike had spent on all teletype lessons to date.

The students were not allowed to use pencil or paper when working on the teletype. Each exercise was worked on the machine so that all responses could be recorded and latencies measured. The response mode was limited to either numerical answers or simple single-letter answers for multiple-choice problems.

Initial instruction on the teletype and program operation consisted of explaining to each class the general procedure of taking turns on the machine and of showing that only the answer need be typed on the keyboard. The program logic was also explained. Staff members helped each student find the letters to type his name for the first two or three lessons. Students had little trouble learning how to type their names or answer the questions.

Following the summary and *goodbye* the student was told, "Please tear off on dotted line." A dotted line was printed, and the student then tore off his print-out and took it with him as a permanent record of his work.

Results

To begin with, it must be emphasized that we have not attempted a detailed model-theoretic analysis of data from all the concept blocks listed in Table 1. We have selected five topics on which we had considerable data and which were sufficiently simple to provide a good starting point. The first analysis deals with fourth-graders' and fifth-graders' performance on addition; the data are drawn from blocks 1 and 3 of grade four and block 1 of grade five listed in Table 1. The second analysis is concerned with subtraction at the same grade levels; the data are drawn from block 2 for each grade. The third analysis is of fourth-grade multiplication data, drawn from block 5. The fourth analysis deals with a relatively controlled experiment on the multiplication tables for grades three to six; the data are drawn from blocks 37, 35, 31, and 28, respectively, for each of grades three to six. The final analysis returns to the data of the first analysis and looks at the results of breaking up the regression analysis of the number of steps into several variables, as indicated in the theoretical discussion.

As remarked earlier, for each set of problems examined, success latency and error probability have been treated as separate dependent variables. Separate regression coefficients were obtained from the same independent variables to predict latency and error probability. This is justifiable by the intuitive assumption that success latency and error probability are different measures of common underlying processes, and is justified empirically by our finding that the correlation between the two dependent variables was consistently greater than 0.7 for the data we have collected.

To minimize the effects of subject variables such as IQ, the problems and data were usually treated separately by grade, concept block, and level, as is made explicit in Tables 2 and 5. It is assumed that children working within a given branching level form a more homogeneous group of subjects than children working on problems at different levels. We were unable to analyze data from some of the levels available because too few children entered those branches.

The first step in analysis was to obtain regression coefficients for each grade and level for the two dependent variables. A stepped, multiple linear regression analysis program, BIMD 02R, adapted for the IBM 7090 computer at Stanford University, was used to obtain regression coefficients, multiple correlation R and R^2. For a finer-grained analysis of the goodness of fit of the success latency predicted from the regression model and of the observed success latency, a computer program was written to calculate the predicted mean success latency for each problem and to give as a measure of fit $S^2 = \Sigma$ (obtained latency_i − predicted $\text{latency}_i)^2/(N-k)$, where N is the number of problems for which the latency was predicted and k is the number of estimated parameters. Similarly, for a finer analysis of the goodness of fit of the regression model to the error data, a program was written to calculate the predicted proportion of errors for each problem i from the obtained regression coefficients and to give as a measure of fit χ_i^2, where $\chi_i^2 = (f_i - p_iN)^2/[p_i(1-p_i)N]$ and f_i = observed frequency of correct responses, p_i = predicted probability of a correct response, N = number of students.

Addition — Grades Four and Five. The three independent variables used in the regression analyses for addition were the variable NSTEPS, which was described in detail earlier, and the two magnitude variables, magnitude of sum (MAGSUM) and magnitude of the smallest addend (MAGSMALL). It is obvious that the value of MAGSUM and MAGSMALL is independent of whether the problem for the student was to find the missing sum or a missing addend. For example, in the three related problems

7 + 9 = __, 7 + __ = 16, and __ + 9 = 16, MAGSUM = 16 and MAGSMALL = 7.

The coefficients obtained for the regression equations are shown in Table 2. This table indicates the level of problems analyzed, the number of children who worked on the problems in that level (Subjects),* the number of different problems analyzed,† the regression constant, and the

Table 2. Linear Regression Coefficients for Addition in the Fourth and Fifth Grades

Level	No. of Subjects	No. of Problems	Constant	NSTEPS	MAGSUM	MAGSMALL	R	R^2
		Grade 4, Set 1, Proportion of Errors						
2	6	19	−2.73	0.16	0.09	−0.03	0.61	0.37
3	21	38	−2.65	0.16	0.05	−0.03	0.56	0.32
4	24	38	−1.44	0.24	−0.01	0.05	0.86	0.74
5	9	19	−1.74	0.08	0.03	−0.09	0.40	0.16
		Grade 4, Set 1, Success Latency						
2	6	19	0.24	0.18	0.14	−0.07	0.64	0.40
3	21	38	−0.76	0.47	0.13	−0.09	0.69	0.48
4	24	38	2.32	0.57	−0.02	0.07	0.86	0.74
5	9	19	2.19	0.17	0.00	0.00	0.44	0.19
		Grade 4, Set 2, Proportion of Errors						
2	7	57	−1.69	0.17	0.02	−0.02	0.54	0.29
3	41	95	−0.73	0.21	−0.01	−0.01	0.64	0.41
4	34	76	−1.60	0.20	0.00	0.01	0.80	0.64
		Grade 4, Set 2, Success Latency						
2	7	57	0.95	0.56	0.06	−0.09	0.64	0.42
3	41	95	1.77	0.73	0.01	−0.06	0.82	0.68
4	34	76	1.55	0.47	0.01	0.02	0.75	0.56
		Grade 5, Proportion of Errors						
3 & 4 combined	12	57	−2.41	0.10	0.03	0.03	0.81	0.66
		Grade 5, Success Latency						
3 & 4 combined	12	57	−2.22	0.47	0.09	0.07	0.73	0.54

regression coefficients for the three independent variables. The absence of a value of a given coefficient indicates that the variable it applies to made no significant contribution to the regression equation, and the computer program therefore did not use that variable in obtaining a regression line. In reading the regression table it should be remembered that the transformation described previously was applied to the observed proportion of

* The number of subjects or students shown in the various tables is always an approximation, with the exact number varying slightly from day to day.

† For reasons mentioned below, the first problem was deleted from each drill, leaving 19 problems per drill. The number of different daily drills in an analysis can be calculated by dividing the number of problems by 19.

errors, and therefore when obtaining a prediction from the coefficients for proportion of errors, the numbers z_i calculated from the coefficients must be transformed to obtain the predicted proportion of errors.

It is clear from scanning the coefficients in Table 2 that NSTEPS is the most important of the three variables in predicting both errors and suc-

Table 3. Predicted and Observed Proportions of Errors and Success Latency in Fourth-Grade Addition, Concept Block 1, Level 4

Rank	Equations	Observed $(1-p_i)$	Predicted $(1-p_i)$	Observed Latency	Predicted Latency	χ^2
1......	0 + 29 = 0 + ___	0.01	0.03	2.40	2.23	0.40
2......	1 + 38 = ___ + 0	0.05	0.13	3.20	3.72	0.69
3......	1 + 27 = 0 + ___	0.06	0.16	2.60	3.97	2.84
4......	2 + 36 = 0 + ___	0.09	0.13	3.80	3.74	0.16
5......	0 + 34 = 0 + ___	0.09	0.03	3.40	2.12	1.45
6......	3 + 26 = 0 + ___	0.14	0.16	3.70	3.95	0.08
7......	1 + 25 = ___ + 0	0.17	0.16	2.90	4.02	0.00
8......	4 + 27 = 2 + ___	0.27	0.68	7.30	6.33	8.30
9......	33 + 3 = ___ + 29	0.27	0.79	5.70	6.86	17.51
10......	12 + 10 = ___ + 0	0.28	0.18	4.50	4.11	2.53
11......	10 + 16 = 6 + ___	0.39	0.41	3.90	5.00	0.04
12......	27 + 2 = 10 + ___	0.39	0.64	5.20	5.78	9.75
13......	7 + 18 = 0 + ___	0.39	0.38	4.50	5.19	0.01
14......	9 + 28 = 7 + ___	0.46	0.77	6.50	6.53	5.91
15......	10 + 29 = 8 + ___	0.46	0.53	5.00	5.41	0.26
16......	11 + 12 = ___ + 1	0.53	0.56	5.80	5.87	0.16
17......	24 + 3 = 5 + ___	0.58	0.46	5.00	5.35	2.37
18......	17 + 5 = ___ + 11	0.58	0.40	6.20	5.02	5.04
19......	9 + 14 = ___ + 2	0.58	0.72	7.50	6.51	3.11
20......	9 + 18 = ___ + 5	0.61	0.76	7.40	6.63	4.68
21......	34 + 5 = 11 + ___	0.64	0.59	5.90	5.78	0.11
22......	22 + 7 = ___ +14	0.72	0.69	7.60	6.15	0.19
23......	7 + 22 = 6 + ___	0.72	0.53	6.00	5.50	5.37
24......	11 + 28 = 8 + ___	0.73	0.67	6.00	5.98	0.19
25......	27 + 7 = ___ + 20	0.73	0.66	6.40	6.03	0.19
26......	17 + 22 = 5 + ___	0.73	0.59	7.00	5.78	0.90
27......	35 + 3 = ___ + 12	0.73	0.54			1.62
28......	30 + 2 = ___ + 5	0.73	0.54	4.60	5.74	1.54
29......	23 + 2 = ___ + 8	0.75	0.83	7.00	6.88	1.46
30......	25 + 4 = 11 + ___	0.75	0.61	5.70	5.94	2.90
31......	19 + 8 = ___ + 6	0.75	0.78	6.40	6.70	0.25
32......	32 + 5 = ___ + 9	0.91	0.72			1.91
33......	29 + 7 = ___ + 15	0.91	0.85	6.10	7.13	0.26
34......	22 + 12 = 16 + ___	0.91	0.95	8.10	8.09	0.39
35......	12 + 22 = ___ + 6	0.91	0.85			0.34
36......	33 + 1 = 7 + ___	0.91	0.64	7.60	6.19	3.49
37......	14 + 10 = 9 + ___	0.92	0.84	6.30	6.97	1.43
38......	29 + 3 = ___ + 17	0.96	0.93			0.13

$\chi^2 = 87.94$ (38 items); χ^2 (items <10) $= 70.43$ (37 items); $S^2 = 0.73$.

cess latencies. A rough indication of the goodness of fit of the regression lines is reflected by the multiple correlation coefficient R and its square R^2, which is an estimate of the amount of variance accounted for by the regression model. In only one case is less than 40 per cent of the variance in the success latency accounted for by the model. When one takes intc account the facts that the two magnitude variables account for a relatively small amount of the variance, and that in setting up the variable NSTEPS we have combined several potentially powerful and probably independent variables, the results are encouraging.

Table 3 presents an example of the individual contributions of the problems to χ^2 when a set of coefficients for response errors given in Table 2 was used to predict the proportions of errors. Included in Table 3 are the rank order of observed problem difficulty, the observed proportion of students making errors (Observed $[1-p_i]$), the proportion of errors predicted from the linear regression model (Predicted $[1-p_i]$), and the actual component of the χ^2 contributed by the problem. Also included in Table 3 are the observed and predicted success latencies for individual problems. The problems in Table 4 constitute one of the eight sets of drills for which variable weights were calculated and presented in Table 2.*

Table 4 summarizes the total χ^2 and S^2 for all eight sets of drills. Some of the χ^2 values obtained in Table 4 are extremely high and would usually be an indication of a poor fit, but a closer look at the components of the χ^2 showed that a few problems in each set or drill made extremely large contributions to the total χ^2. In Table 3, for instance, problem 9 contrib-

Table 4. Total χ^2 and S^2 Obtained as a Measure of Fit of Predicted versus Observed Proportion of Errors and Success Latency for Addition

Grade	Concept Block	Level	No. of Problems	S^2	χ^2	Reduced χ^2	No. of Items in Reduced χ^2
4	1	2	19	.65	18.03	18.03	
4	1	3	38	1.29	233.91	83.78	33
4	1	4	38	.73	87.94	70.43	37
4	1	5	19	1.17	24.69	24.69	
4	3	2	57	2.29	97.02	66.31	55
4	3	3	95	.62	391.87	156.75	83
4	3	4	76	1.04	225.08	134.90	70
5	3	3 & 4	57	2.32	64.00	64.00	

* Complete data on individual problems may be found in the Institute's Technical Report No. 100, *Linear Structural Models for Response and Latency Performance in Arithmetic.*

uted 20 per cent of the total χ^2 obtained. In several of the other sets of drills two or three particular problems contributed as much as 70 per cent of the total χ^2.

In some of the sets of drills, as in the one in Table 3, the reduction in χ^2 obtained by deleting a few extreme problems is still insufficient to yield a value of χ^2 such that the model would normally be accepted. When we do reduce the χ^2 values presented in Table 4 by removing the few extreme components whose individual contributions are equal to or greater than 10, we find that in four of the eight cases we obtain a χ^2 value whose probability is such that $.1<p<.9$ under the null hypothesis. The reduced χ^2 and the total number of problems contributing to the reduced χ^2 for each of the eight cases is also given in Table 4. Since calculation of the regression coefficients included the extreme problems, a recalculation of the regression coefficients omitting these problems from the data would yield better fits of the model to data than those obtained.

Perusal of the individual χ^2 components with particular attention paid to the items for which predictions are unsatisfactory, and thus the χ^2 contribution high, suggests immediate further analysis that incorporates variables designed to handle special algorithms. Moreover, in those cases for which $p<.1$, the actual individual predictions are mostly good. Our view of this matter is that we hardly expected to fit the data exactly with such a small number of variables. Above all, we emphasize that this procedure of dropping out individual problems with large χ^2 values is certainly not admissible as an inference procedure. We look upon the χ^2 values reported here as providing a useful descriptive statistic for summarizing the order of magnitude of deviations between the observed and predicted results for the bulk of the problems, and for identifying types of problems that require a more elaborate theory even when a rather lax goodness-of-fit concept is applied.

Figure 3 is a graph of the predicted and observed proportions of errors as a function of rank order of observed difficulty for fourth-grade addition, block 1, levels 2, 3, and 4. An inspection of the two curves shows a relatively good fit for the regression model, even in the heterogeneous case of problems drawn from different drills and levels and correspondingly different groups of children.

There are several qualitative observations about the data for individual problems that we want to make at this point. The first problem presented on each day was deleted from all analyses when the results showed a short but significant warm-up which rendered the initial problem more difficult, independent of structural variables. Happily, it was not necessary to include order of presentation as a variable, since there was no significant warm-up beyond the first problem of a drill. The sequential effects, if any, of errors on the second problem have not yet been analyzed systematically, but again this does not appear to be a very strong effect. Although we intend to go into this question in more detail subsequently, the assumption of statistical independence of problems seems to be correct to a first approximation.

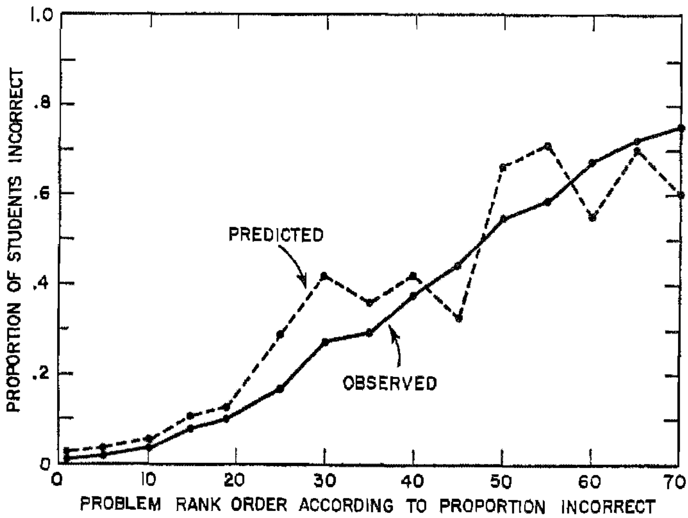

Figure 3. Predicted and observed proportions of errors in fourth-graders' addition.

Five sets of drills summarized in Table 2 (fourth-grade block 1, levels 1 and 2; fourth-grade block 2, levels 1 and 2; fifth-grade, levels 3 and 4) contribute data on problems of the general form m + n = p, where any of m, n, or p may be two-digit numbers. What is striking is that the hardest problems are to a very large extent of the form __ + n = p. The last problems (since the problems were ranked in order of difficulty from easiest to hardest) of all sets of problems of this type analyzed were always of this form. Moreover, if we look at the easiest problems in these sets of drills, the form __ + n = p is almost entirely excluded. All in all, these results suggest that the transformation steps defined in the theoretical section might well be broken into separately weighted classes to differentiate __ + n = p from m + __ = p. In some preliminary efforts aimed at refining and improving the predictive results reported here, we have had some success with this distinction.

Although the predictive results from the data corresponding to Table 3 for the other sets of drills are far from the best that a mature theory should be able to offer, we are not dissatisfied with them as a beginning because of the relative difficulty of intuitively rank-ordering the proportion of errors obtained in addition problems, particularly the form ab + cd = ef + gh, as shown in Table 3. The three variables that we consider bring a surprising amount of order to what appears at first glance to be quite a complex set of problems.

We turn now to the data on success latency for the same problems of fourth- and fifth-grade addition. An example of the predicted and observed latencies are also presented in Table 3, with the predicted values depending on the appropriate regression coefficients of Table 2. As is clear from Table 2, the multiple correlations obtained for the fit of the predicted latencies are comparable with those obtained for the predicted responses, and indicate that data on success latency are as regular in range of variation as the response data.

In the analysis of latencies we have restricted ourselves to the success latencies — that is, the latencies of correct responses — because of their direct relevance for the analysis of the structure of the algorithms students use. Although error latencies also contain much useful information, they include latencies of random guesses, false starts, and other heterogeneous factors that are not easily disentangled. In a few cases latency data were garbled in transmission from the school to the computer, and in such cases we have simply entered a blank in both the observed and predicted columns.

There are various ways of evaluating the overall fit of the latency predictions. The statistic S^2, already mentioned, is given in Table 4. Although this statistic may be used to find a level of significance for the fit of the structural models, at this stage of investigation it seems more useful to interpret S^2 directly in terms of the quantitative closeness of the predictions. When the structural variables f_{ij} are not random variables, then S^2 is a good estimator of the variance σ^2 of the errors in the prediction of the models. Taking the algebraic sign into account, the expectation of these errors is nearly zero, and the assumption that they are normally distributed with variance σ^2 is approximately satisfied also, and so we may evaluate the predictions of each set of drills by looking at the magnitude of S, the bulk of the errors being within one standard deviation of the observed values. The values of S for the eight sets of data are .81, 1.14, .85, 1.08, 1.51, .79, 1.02, and 1.52, respectively, which may be interpreted to mean that errors of prediction greater than 1 or 1.5 seconds do not occur very often. The observed values for the eight sets of drills have a range from about 3 seconds to more than 8.5 seconds, and consequently, predictions with an accuracy of 1 to 1.5 seconds are far from perfect, yet good enough to be practically useful.

Still another useful measure is the average percentage of error of the predictions. If there are n items in a table, if o_i is the mean observed success latency for item i, and if p_i is the predicted latency, then the average percentage error (A.E.) is defined by $A.E. = (100/n)\Sigma[(o_i - p_i)/p_i]$. This measure for the eight sets of drills has the values 16.4, 19.8, 12.3, 20.6, 26.7, 16.6, 12.8, and 31.4 per cent, respectively.

The qualitative remarks made about proportion of errors apply to success latencies, as would be expected because of the high positive correlation between the two variables. This is particularly true of latencies for problems of the form ___ + n = p.

Figure 4 is a graph of the predicted and observed success latencies for

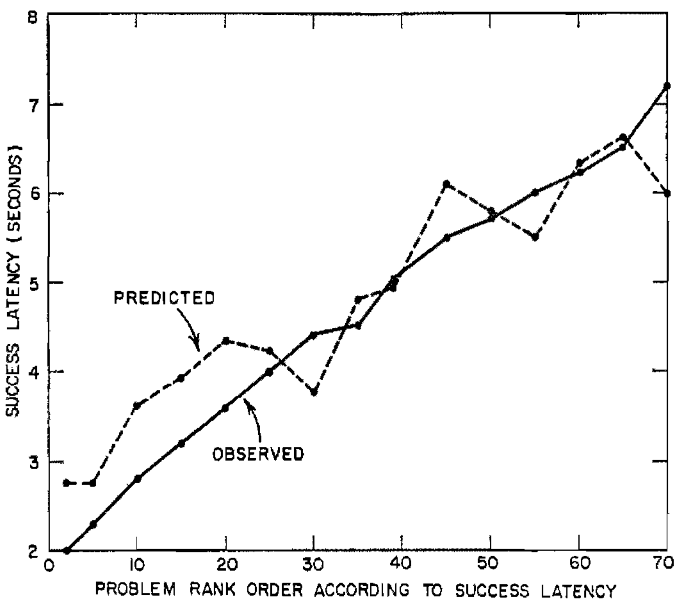

Figure 4. Predicted and observed success latencies for fourth-graders' addition.

the same problems for which the response predictions are shown in Figure 3. The predicted and observed latencies are plotted as a function of the rank of observed latency, and consequently, the curve of observed latencies is monotonically increasing and smoother than the predicted curve, but the fit is, qualitatively, a reasonably good one.

Subtraction — Grades Four and Five. The three independent variables used in the linear regression analyses of subtraction were NSTEPS as described previously and the two magnitude variables, magnitude of the difference (MAGDIF) and magnitude of the subtrahend (MAGSUB). The values of MAGDIF and MAGSUB are not affected by the problem format. For example, in all three problems, 31 − 16 = __, 31 − __ = 15 and __ − 16 = 15, MAGDIF has the value 15 and MAGSUB the value 16.

The coefficients obtained for the regression equations are shown in Table 5, which is laid out in a manner identical to that of Table 2. As in Table 2, it is clear that NSTEPS is the most important of the three variables in predicting errors or success latencies. The values obtained are also comparable with those given in Table 2. In the confines of the present paper it has not been possible to explore the possibility of a joint analysis of addition and subtraction, with a particular emphasis on process variables like NSTEPS, but this is a clearly indicated direction for future research.

Table 5. Regression Coefficients for Subtraction

Level	No. of Subjects	No. of Problems	Constant	NSTEPS	MAGDIF	MAGSUB	R	R^2
			Grade 4, Proportion of Errors					
1.........	5	19	−0.42	0.06	−0.03	0.09	0.73	0.54
2.........	11	38	−1.09	0.19	0.01	0.02	0.43	0.18
3.........	43	76	−1.63	0.12	0.02	0.09	0.61	0.38
			Grade 4, Success Latency					
1.........	5	19	6.82	0.58	−0.34	−0.27	0.62	0.38
2.........	11	38	1.42	0.49	0.05	0.11	0.48	0.23
3.........	43	76	1.49	0.32	0.06	0.20	0.64	0.41
			Grade 5, Proportion of Errors					
1.........	15	38	−1.50	0.15	0.00	0.08	0.70	0.49
2.........	27	57	−1.98	0.44	0.00	0.01	0.80	0.65
3.........	25	76	−1.65	0.40	−0.03	0.01	0.82	0.68
4.........	9	57	−1.14	0.20	−0.01	0.01	0.68	0.46
			Grade 5, Success Latency					
1.........	15	38	−1.77	0.65	0.12	0.32	0.73	0.54
2.........	27	57	0.70	0.99	0.03	0.01	0.80	0.64
3.........	25	76	−1.91	1.59	−0.02	0.04	0.80	0.64
4.........	9	57	2.58	0.71	−0.01	0.00	0.58	0.34

Again, as in the case of Table 2, the multiple correlation coefficients in Table 5 show that the three independent variables are accounting for much of the variation in the observed response proportions and success latencies.

Table 6. Total χ^2 and S^2 Obtained as a Measure of Fit of Predicted versus Observed Proportion of Errors and Success Latency for Subtraction

Grade	Concept Block	Level	No. of Problems	S^2	χ^2	Reduced χ^2	No. of Items in Reduced χ^2
4	2	1	19	3.84	16.65	16.65	
4	2	2	38	1.80	104.03	92.43	37
4	2	3	76	1.68	445.57	136.30	61
5	2	1	38	1.64	81.63	67.70	37
5	2	2	57	0.64	90.67	78.76	56
5	2	3	76	1.25	137.34	73.39	73
5	2	4	57	2.93	81.43	57.05	55

Table 6 presents total χ^2, reduced χ^2 and S^2, and other information for the seven sets of subtraction drills identical to that given in Table 3 for addition. The overall χ^2's for subtraction exhibit a pattern similar to those obtained for addition. Again we find that although the initial fits of the model to the seven sets of data are poor, the removal of a few extremely large χ^2 contributions gives a reasonable fit of the model to four of the seven sets of data with $.1 < p < .9$ for acceptance under the null hypothesis. Table 7 gives an example of individual subtraction problems and their contributions to χ^2 and S^2 similar to that given in Table 3 for some addition data. The statistic S^2 for each of the seven sets of drills is presented in Table 6. Applying the same interpretation as before to this statistic, we may look at the value of S for each set of drills as an estimate of the standard deviation of the approximately normal distribution of errors of prediction. For the seven sets of drills we find the values of S to be 1.96, 1.34, 1.30, 1.28, 0.80, 1.12, and 1.71, respectively, and it is reasonable to say that errors of prediction greater than about 1.50 seconds should not occur very often. The observed success latencies have a range from slightly more than 5 seconds to more than 8 seconds, and consequently, a model with errors that have an approximately normal distribution with a standard deviation of about 1.5 seconds yields meaningful and useful predictions. These results are comparable with those found for addition. The same is true of the measure of average percentage error, which is 25.7, 23.0, 24.6, 26.3, 10.9, 17.1, and 23.0 per cent for the seven sets of drills.

Table 7. Predicted and Observed Proportions of Errors and Success Latency in Fourth-Grade Subtraction, Level 1

Rank	Equations	Observed $(1-p_i)$	Predicted $(1-p_i)$	Observed Latency	Predicted Latency	χ^2
1	12 − 0 = ___	0.10	0.16	2.70	2.72	0.13
2	14 − 0 = ___	0.10	0.14	1.70	2.04	0.08
3	___ − 3 = 11	0.20	0.38	7.70	4.58	0.69
4	12 − 6 = ___	0.20	0.60	3.00	5.48	3.33
5	14 − 3 = ___	0.40	0.32	4.70	3.42	0.14
6	___ − 1 = 13	0.40	0.29	4.90	5.03	0.27
7	___ − 7 = 6	0.40	0.70	8.80	6.37	2.21
8	___ − 0 = 13	0.40	0.19	2.50	3.55	1.48
9	13 − ___ = 7	0.40	0.62	7.00	5.72	1.00
10	14 − ___ = 5	0.40	0.77	6.40	5.59	3.77
11	___ − 6 = 5	0.60	0.67	7.80	6.99	0.12
12	12 − ___ = 8	0.60	0.50	5.00	5.92	0.18
13	11 − 4 = ___	0.60	0.49	6.80	5.68	0.26
14	14 − ___ = 4	0.80	0.77	4.00	4.49	0.03
15	15 − ___ = 8	0.80	0.65	3.20	5.11	0.50
16	___ − 5 = 8	0.80	0.59	2.70	6.23	0.94
17	___ − 10 = 2	0.90	0.79			0.38
18	14 − ___ = 6	0.90	0.72			0.82
19	11 − ___ = 2	0.90	0.80			0.33

$\chi^2 = 16.65$ (19 items); χ^2 (items <10) $= 16.65$ (19 items); $S^2 = 3.84$.

Without making an exact statistical comparison, it still seems clear that the approximate measures of fit we have reported for the success latencies in subtraction reflect a better fit to the data than do the χ^2 measures for the predicted response proportions. The predictions of response proportions still leave a lot to be desired. The predictions of success latencies seem to reflect more regularly the observed rankings of latencies, even though this apparent difference in favor of latency predictions is not well reflected in the multiple correlation coefficients of Table 5.

Inspection of the data for drills in subtraction confirms the intuition that subtraction problems of the form ___ $- n = p$ are relatively not so difficult as addition problems of the same form. No doubt the reason for this is that a single simple transformation converts such subtraction problems into the easiest kind of addition problem, $p + n =$ ___.

It should be noted that one set of drills included problems with both letter variables and blanks. It is interesting that problems with letter variables are the six easiest problems in terms of response errors, although the same six problems do not have the shortest latencies. The format of these problems with letter variables was:

45		29		37		29	
−FG		− 5		− K		−24	
3	FG = ___	HK	HK = ___	32	K = ___	M	M = ___

The ease of handling algebraic notation is also confirmed by some other unpublished experiments conducted in the Institute for Mathematical Studies in the Social Sciences several years ago with first- and second-grade children.

Multiplication – Grade Four. The two sets of multiplication problems considered each contained twenty exercises, as in the case of addition and subtraction. The two sets concentrated on a review of multiples of 4 and 5, with the second factor ranging from 0 to 12. The problems occurred in the three forms m × n = ___, m × ___ = p, and ___ × n = p. Unlike the addition and subtraction analyses covered in the previous pages, NSTEPS was not considered as a variable because in all problems only one operation was involved. To see if transformations as described in the theory section defined a significant variable, we treated each of the three equational forms as an independent variable which took on the value 1 if the problem was in the given form and 0 if it was not. The other two independent variables used were the larger factor (LARGER) and smaller factor (SMALLER) that yielded the product. In the case of squares (4 × 4 and 5 × 5) the values of the two factors were equal. The regression coefficients for the five variables, with proportion of errors and success latency as dependent variables, are presented in the accompanying tabulation for forty problems worked by 47 subjects.

	Errors	*Success Latency*
Constant	−1.46	1.02
LARGER	0.10	0.18
SMALLER	0.01	0.33
a × b = ___	−0.29	−0.38
a × ___ = c	...	...
___ × b = c	...	...
R	0.70	0.78
R^2	0.50	0.62

Again we found that the linear regression model does well at predicting errors and success latency from a small number of variables. The only variable in equation form that significantly affected the regression line was the canonical form a × b = ___, and the negative coefficients of this variable indicate that problems of this form are easier than problems of the form ___ × b = c or a × ___ = c, a finding that is well in keeping with intuition.

When all forty problems were considered, the overall $\chi^2 = 113.29$, but the deletion of three problems, each with a χ^2 component greater than 10, dropped the total for the remaining thirty-seven problems to $\chi^2 = 41.43$. The three problems dropped from the analysis were 4 × 10 = ___, for which the predicted error proportion was much higher than observed, and 4 × ___ = 4 and ___ × 4 = 48, for both of which the predicted error proportions were much lower than observed. It is not difficult to analyze why these three problems probably deviated greatly from the predicted values. In general, 4 × 10 = ___ allows use of the simple algorithm a × 10 = a0. We would expect the same low finding of error for 5 × 10 = ___, but this problem was not in the two sets. The problem 4 × ___ = 4 turned out to be the first multiplication problem, and as mentioned in the discussion of addition, there is evidence of a warm-up which affects response to the first problem of the day. The problem ___ × 4 = 48 is the only problem of the set for which the initial factor is both 12 and also the answer to be found. The S^2 for comparing observed and predicted latency was quite low. The obtained value, $S^2 = .62$, indicates that most prediction errors were definitely less than 1.0 second.

Multiplication Tables — Grades Three, Four, Five, and Six. Toward the end of the school year we decided to run the one hundred one-digit multiplication problems of the form a × b = ___ to see how well a structural model would predict response behavior. Previous investigations of performance on these basic multiplication facts are not so numerous as we had expected, and the kind of regression model applied here has not been previously used so far as we know. The first point to note is that for all four grades the response performance was extremely good. The error rate was 8.0 per cent for third-graders and 3.2 per cent for sixth-graders, with the fourth- and fifth-graders falling between these two bounds. Consequently our analysis in this case is restricted entirely to success latencies. Because the form of the equations was constant in the hundred problems, we have restricted our regression to the two factors, SMALLER and LARGER, already used in analyzing fourth-grade multiplication. The regression coefficients, multiple correlation, and statistic S^2 for each grade are shown in Table 8. There are several observations to be made about this table. In the first place, for all four grades the multiple correlation R is extremely high, indicating that the two variables are giving a good account of the

Table 8. Linear Regression Coefficients for the Multiplication Tables

Grade	Subjects	Constant	LARGER	SMALLER	R	R^2	S^2
3	24	1.71	0.06	0.30	0.86	0.74	0.22
4	56	1.52	0.07	0.28	0.85	0.73	0.22
5	20	1.38	0.09	0.29	0.78	0.61	0.42
6	32	1.33	0.06	0.19	0.82	0.68	0.14

data. This inference is supported by the small values of S^2, which are the lowest values reported for any of the sets of data analyzed in this paper. It is also apparent from the values of the regression coefficients that the magnitude of the smaller factor is more important than that of the larger factor. Thus, for example, on the average it takes longer to say what 4×5 is than to say what 1×9 is. Finally, with analysis for four grades before us, it is natural to ask whether we can find evidence of development from one grade to another. Development is most apparent in the monotonically decreasing values of the constant, which reflect an increase in speed of response with age. In the regression model for latencies, the constant enters in a direct additive way. The decrease from 1.71 seconds in the third grade to 1.33 seconds in the sixth grade is not surprising. What is surprising is that the coefficients of the two factors do not also decrease monotonically with age – this complicates considerably the task of constructing a model of developmental processes and their effects on arithmetic performance.

Figures 5 and 6 show the predicted and observed success latency curves for the third and sixth grades, respectively. The hundred problems are rank-ordered according to success latency on the abscissa, and thus the

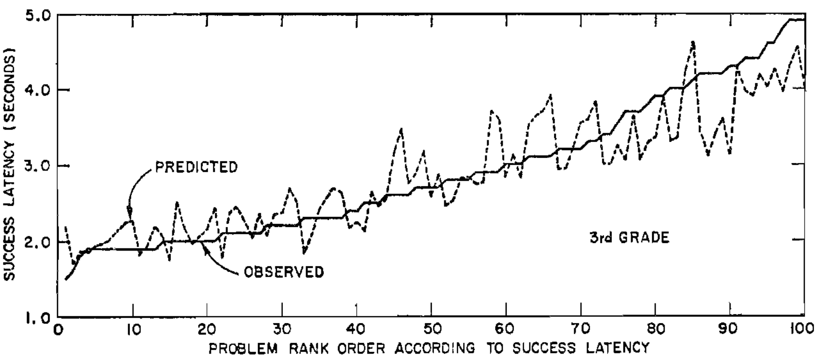

Figure 5. Predicted and observed success latencies of third-grade students on the multiplication tables.

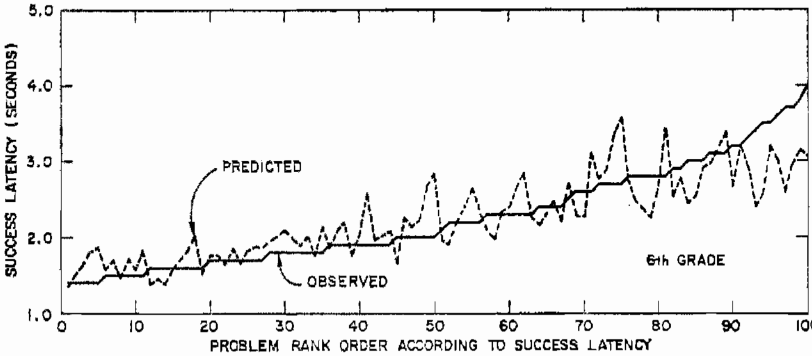

Figure 6. Predicted and observed success latencies of sixth-grade students on the multiplication tables.

observed data define a relatively smooth, monotonically increasing function. The predicted curve is determined for each grade level by the three estimated coefficients given in Table 8 and the two given factors of each multiplication problem. Considering the wide range of latencies found in each figure, running from 1.5 to 4.9 seconds in the third grade and from 1.4 to 4.0 seconds in the sixth grade, we believe that the predicted curves fit the observed data quite well. For those readers accustomed to looking at smooth predicted learning curves that are essentially exponential in form, we emphasize that the predictive task is different and rather more difficult, as we move not from like trial to like trial, but from a problem with one structure to another problem with a different structure.

Analysis of the Factors in NSTEPS. As we promised in the theoretical discussion of the second section, we now make a preliminary analysis of breaking up the single variable NSTEPS into its three components of transformation, operation, and memory. The analysis here differs slightly in one respect from the definition given in the second section: transformation steps always were either 0 or 1, never 2. With this exception, the analysis is based entirely on the earlier definitions. The data used in the first analysis are taken from those already summarized in Table 2, but without the

	Errors	*Success Latencies*
Constant	−1.29	2.94
Transformation	0.20	0.58
Operation	0.00	0.00
Memory	0.39	0.75
R	0.73	0.69
R^2	0.53	0.48

first item of each set of problems deleted. Thus this first analysis is of eighty fourth-grade addition problems. The results are shown in the accompanying tabulation.

In the case of both errors and success latencies it ought to be observed that memory is the most important variable, whereas operation plays no role. Moreover, in both cases we get nearly as good a fit simply by using memory as the single variable. In the case of errors, the difference in the multiple correlation R occurs only in the third decimal — .726 rather than .731 — and so does the difference in latency — .677 rather than .688. The χ^2 and S^2 values that come from using the coefficients of the tabulation are high, but are not out of line with those reported earlier. In particular $\chi^2 = 417.7$, and if we delete the twelve extreme items with individual χ^2's greater than 10, $\chi^2 = 170.3$ for the remaining sixty-eight items. The statistic S^2 equals 1.42, which yields an estimate of 1.19 for the standard deviation of the errors in prediction. What is particularly worth noting is that the correlation for fourth-grade addition (block 1, level 3) in the earlier analysis of the same data is lower than the correlation for the combined data of this tabulation.

A second, somewhat different analysis was performed on a set of 19 problems that, together with the initial problem omitted in the analysis, formed one day's exercises in fourth-grade addition, block 3, level 4. These 19 problems are among the 76 already analyzed. The departures from the earlier definitions of the components of NSTEPS were these: First, because the problems were all of the form ab + cd = ___ + ef or ab + cd = ef + ___, the number of transformations was the same for all problems and therefore was omitted as a variable. Second, the operations of addition and subtraction of single digits were treated as separate variables. Third, the number of digits in memory was expanded to include all digits used in obtaining a solution — including those given in the problem, those that occurred as partial solutions, and those present in the response. The three variables considered were, therefore, number of addition operations (O_1), number of subtractions operations (O_2), and number of digits processed (Memory). The accompanying tabulation presents the regression coefficients for the three variables found, with proportion of errors and success latency as dependent variables. The very high correlations for both errors and latencies warrant a closer look at the results. For the data entering this analysis, the mean number of addition operations was 2.4, the mean number of subtraction operations was 1.8, and the mean number of digits

	Errors	*Success Latencies*
Constant	–2.65	–0.42
O_1	0.06	0.00
O_2	0.25	0.00
Memory	0.25	0.77
R	0.89	0.86
R^2	0.79	0.73

processed was 8.7. It would appear that the number of addition operations has much less effect on errors than does the number of subtraction operations. Neither of these two variables has a significant effect on success latency. Figure 7 presents the observed and predicted proportion of errors as a function of ranked difficulty; with the exception of problems 6 and 8, the observed and predicted curves are quite similar. Figure 8 shows observed and predicted success latencies as a function of observed latency rank, and once more we find the general shapes of the observed and predicted curves quite similar. Figure 9 is a scatter plot of observed versus predicted errors, and Figure 10 is a similar scatter plot of observed versus predicted success latency. If all the points in the two scatter plots fell on the 45° straight line, the predictions would be perfect. The deviations of the points from this line are a measure of the goodness of fit of the model.

Discussion

In this final section, we shall not attempt to summarize systematically the results reported in the previous section. It is our belief that the results establish clearly enough the real possibility of analyzing and predicting in terms of meaningful variables the response and latency performance of children who are solving arithmetical problems. As we have already stated, the predictive results reported here have been good enough to be used practically, but they are incomplete enough to challenge anyone interested in systematic psychological theory.

From a psychological standpoint, the most suggestive single finding is probably the importance of the process variable NSTEPS or of its component variables, particularly memory, in all the relevant analyses. It marks a direction of major emphasis in our own future research as now planned. One way of putting the matter is this: If in Table 2, for example, the dominant variables had turned out to be magnitude variables, then our first step would have been less significant, because anyone would immediately ask what characteristics of the processing done internally by the students made these magnitude variables dominant. In postulating process vari-

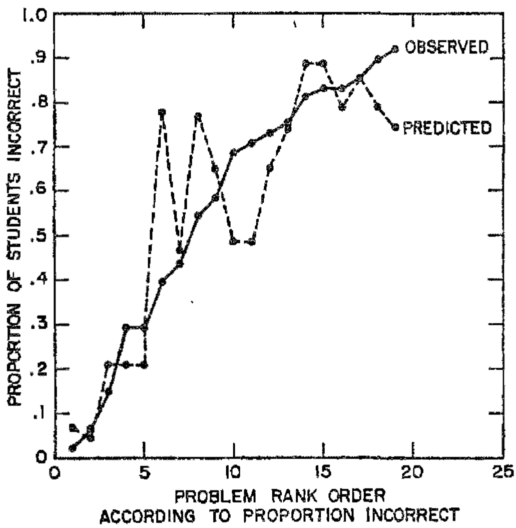

Figure 7. Predicted and observed proportions of errors in fourth-graders' addition. Proportions of errors were predicted from an analysis of the role played by transformation, operation, and memory steps in the solution of each problem.

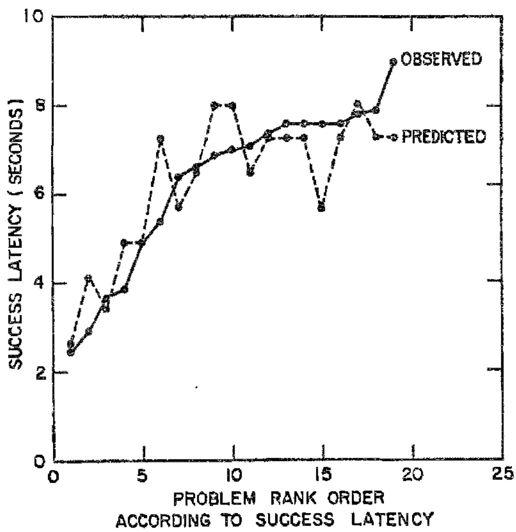

Figure 8. Predicted and observed success latencies for fourth-graders' addition. Success latencies were predicted from an analysis of the role played by transformation, operation, and memory steps in the solution of each problem.

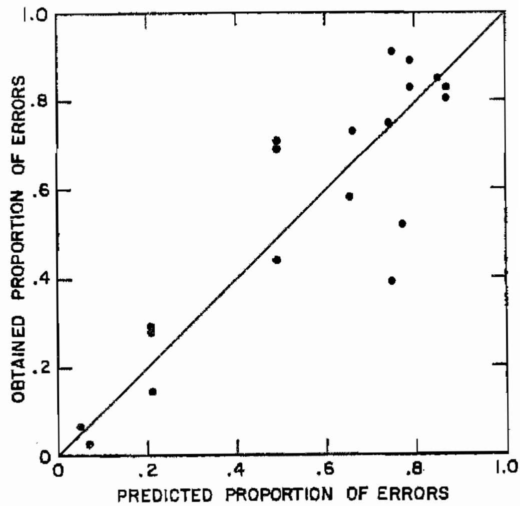

Figure 9. Scatter plot of observed versus predicted errors in fourth-graders' addition.

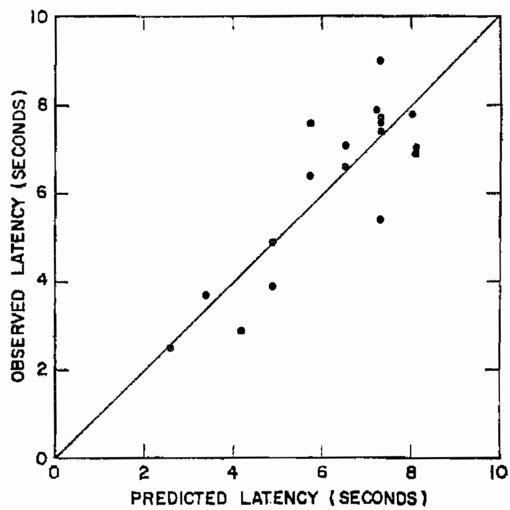

Figure 10. Scatter plot of observed versus predicted success latencies for fourth-graders' addition.

ables and being able to establish their direct importance, we have already been able to move past this first step. Now our central problem is to acquire a better understanding of these variables and to use this understanding to develop better predictive models.

All the analyses reported in this paper have been concerned with mean data averaged over individual students' performances. Moreover, when dealing with data from different age groups, no attempt has been made to estimate parameters that would reflect the course of developmental change in the performance of arithmetical tasks. Systematic amplification in both these directions — taking account of individual differences and of developmental processes — is relatively straightforward, although technically arduous, for all the models we have considered. A disadvantage of the data reported in this paper is that the number of students working at any given level and grade was small. Our main objective for the immediate future is to increase considerably the number of students in order to provide the quantity of data required for meaningful inferences about individual differences or developmental processes.

Finally, because the data reported here were actually collected in an ordinary classroom augmented by a computer-controlled terminal, and because the data are about performance on standard arithmetic problems, it is natural to ask about implications of our various predictive analyses for the teaching of arithmetic. Independent of making any positive remarks on this point, we want to underscore the preliminary value of our findings. Much additional, more refined analysis with data from larger numbers of students is needed to support any definitive pedagogic recommendations. Keeping in mind this explicit reservation, we do believe that the results that are most intriguing from a pedagogic standpoint are the ones reported at the end of the last section about the ability of the memory variable alone to offer a fairly adequate account of the observed data. From the way this variable was defined in the theoretical section, it should be evident that we can identify some specific points to emphasize in teaching multi-digit addition and subtraction. However, we leave for another time and place the taking of this explicit pedagogic step.

References

Atkinson, R. C., & W. K. Estes. Stimulus sampling theory, in R. R. Bush, E. Galanter, and R. D. Luce, eds., *Handbook of mathematical psychology*, Vol. II, pp. 121–268. New York: Wiley, 1963.

Batson, W. H., & O. E. Combellick. Relative difficulty of number combinations in addition and multiplication. *Journal of Educational Psychology*, 1925, 26, 467–481.

Billington, Lillian E. A study of the time interval between stimulus and response in arithmetic fundamentals. Unpublished Ph.D. thesis, Stanford University, 1947.

Bruner, J. S., Jacqueline J. Goodnow, & G. A. Austin. *A study of thinking*. New York: Wiley, 1956.

Estes, W. K. Component and pattern models with Markovian interpretations, in R. R. Bush and W. K. Estes, eds., pp. 9–52. *Studies in mathematical learning theory*. Stanford: Stanford University Press, 1959.

——— & P. Suppes. *Foundations of statistical learning theory*, II. *The stimulus sampling model*. Technical Report No. 26, Psychology Series, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1959.

Helseth, Inga Olla. The four fundamental arithmetical arithmetic processes in adults. *Mathematics Teacher*, 1927, 20, 265–273.

Knight, F., & Minnie S. Behrens. *The learning of the 100 addition combinations and the 100 subtraction combinations*. New York: Longmans, 1928.

Suppes, P. *Sets and numbers*. Books 1–6. Syracuse, N.Y.: L. W. Singer, 1966.

———. The psychological foundations of mathematics, in *Colloques Internationaux Centre National de la Recherche Scientifique*, pp. 213–242. Paris: Editions du Centre National de la Recherche Scientifique, 1967.

——— & R. C. Atkinson. *Markov learning models for multiperson interactions*. Stanford: Stanford University Press, 1960.

Suppes, P., & G. Groen. Some counting models for first-grade performance data on simple addition facts, in J. M. Scandura, ed., *Research in mathematics education*, pp. 35–43. Washington, D.C.: National Council of Teachers of Mathematics, 1967.