IV In J. M. Scandura (Ed.), Research in Mathematics Education. Washington, D.C.: NCTM, 1967. Pp. 35-43.

Some Counting Models for First-Grade Performance Data on Simple Addition Facts*

PATRICK SUPPES
GUY GROEN
Stanford University
Stanford, California

TO MATHEMATICIANS or educators who have not thought very much about the matter it usually comes as a surprise, and occasionally even as a shock, to find out how little we know about the learning of mathematics. It is not uncommon to hear mathematicians say that because mathematics is a systematic subject with an inherent order imposed on the development of topics, it should be relatively straightforward to give quite an adequate account of how students learn mathematics. Because students do learn mathematics and because many of the mathematicians who make this sort of statement have themselves been successful teachers, it is not always evident what is the best way to bring out the gross inadequacies in our present knowledge of mathematics learning.

Perhaps the most effective way—at least we have found that it sometimes works—is to rely heavily on computer analogies. First challenge: If you understand so well how mathematics is learned, please program my computer to learn it. It does not take much discussion to bring out the difficulties of this task, and one can then move on to a second challenge: Predict the points at which students will have learning difficulties, and make explicit the principles used to make the predictions. The requirement of explicitness is needed to make the challenge a scientific and theoretical one that cannot be answered by the nonverbalized and intui-

* This research was supported by the U.S. Office of Education, The National Science Foundation, and the Carnegie Corporation of New York. The authors would like to thank Mrs. Sue Matheson for running the experiments in the school.

tive experience of a good teacher. While one's opponent is struggling with this second challenge, a third challenge on performance data can be put ready at hand: Predict systematic variations in performance data involving mathematical concepts and skills already taught, and again make the principles of prediction explicit. The view that mathematical colleagues may have difficulty giving a serious constructive response to these three challenges is not meant as a criticism of their scientific prowess. The only criticism implied is of the opinion that we already know how to meet these three challenges in any serious way.

The present article is meant to be a small step toward a positive response to the third challenge. From a mathematical standpoint the performance task we have selected is ridiculously simple, that of handling correctly the simple addition facts, with the sums being no greater than 5. From a psychological standpoint, however, this task is not as simple as most of those that lie at the heart of the classical experiments in learning theory. Moreover, attempts to develop mathematically well-defined performance models for even this simple task do not seem to exist in the literature.

We reserve more detailed comments until after we have presented in the next section goodness-of-fit results, i.e., the extent of correspondence between the theoretical predictions and the experimental outcomes for five closely related models. A broader conceptual framework for the viewpoint expressed here is to be found in Suppes.[1]

An Experimental Test of Five Models

The results we will discuss are from an experiment in which a group of first-grade children in the first half of the school year were asked to solve a set of simple addition problems. Each problem was of the form

$$m + n = \_\_,$$

where $m + n \leq 5$. The line was colored red and the rest of the problem was printed in black. The task of each child was to provide the missing number.

Thirty subjects were used, randomly selected from two different homerooms. Each subject was run individually. Subjects were seated in front of a panel with six buttons marked 0, 1, 2, 3, 4, and 5. A sample problem was then projected on a screen in front of them. They were told that the red line meant that a number was missing and were instructed to

[1] Patrick Suppes, *The Psychological Foundation of Mathematics*, Technical Report No. 80 (Stanford, Calif.: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1965). Pp. 33.

push the correct button for the missing number. When a subject had responded, he was shown a new slide with the correct answer (printed in red) replacing the red line. Each child was then presented with a sequence of twenty-one problems consisting of all possible combinations of integers m and n, subject to the constraints

$$\begin{aligned} m + n &\leq 5, \\ m &\geq 0, \\ n &\geq 0. \end{aligned}$$

These problems were presented in a random order, the same sequence being used for each child. After each presentation of a problem, the child made a response and was shown the correct answer. Both the actual response and the response latency (the time between the onset [presentation] of the stimulus and the elicitation [occurrence] of the response) were recorded. This procedure was repeated for two more days. However, on the last two days, no preliminary instructions were given, and the child was asked to respond as quickly as possible. The order of presentation of items was different on each of the three days.

In this discussion, we will concentrate on the data obtained on the third day. It can be assumed that by then the children had become fully familiar with the experimental situation. The initial problem we proposed to consider was whether it is possible to formulate a simple model that will account in an approximate fashion for the children's responses.

Unfortunately, the error rate was too low for any systematic analysis to be based on this aspect of the response data. Although at least one subject made an error on each problem, seven subjects out of the thirty made errors on $1 + 3 =$ ___ and $1 + 2 =$ ___, and five subjects made errors on $4 + 1 =$ ___, $3 + 2 =$ ___ and $1 + 1 =$ ___. On most other problems one or two subjects made an error.

As a result of these low error rates, it seemed more promising to consider the response latencies. The most reasonable basic assumption to make is that the variations in response latencies between problems are the reflection of some kind of counting process that the child is using. For a problem of the form $m + n =$ ___, it is possible to distinguish between five different kinds of counting processes. In order to make this distinction, it is convenient to consider a counter on which two operations are possible: setting the value of the counter to a certain value (while clearing the previous value) and adding a number to the current value of the counter. The addition operation is performed by successively increasing the initial value of the counter by one until the second value has been added on. The operation of this counter is illustrated in Figure 1, as shown on the following page. Using this counter, an addition problem of the form $m + n =$ ___ can be solved in the following ways:

1. *The counter is initially set to 0, m is added and then n.*

2. *The counter is set to m (i.e., the left-most number) and n is then added.*

3. *The counter is set to n, and m is then added.*

4. *The counter is set to the minimum of m and n. The maximum is then added.*

5. *The counter is set to the maximum of m and n. The minimum is then added.*

The setting operation is assumed to take a constant time, independent of the value to which it is set. The addition time, on the other hand, is proportional to the number of times the counter must be increased. Suppose a counter takes time α to be set and time β to be increased by 1. If a counter is to be set to a certain value and then increased x times by 1 (which is equivalent to having x added to it) the total time T taken by the counter to perform these operations is

$$T = \alpha + \beta x. \tag{1}$$

Thus, Equation (1) gives the time taken to perform an addition problem of the form $m + n =$ ___. It will give differential predictions depending on the type of solution because, corresponding to the classification of solution types we have just proposed, x is determined as follows:

Type 1. $x = m + n.$
Type 2. $x = n.$

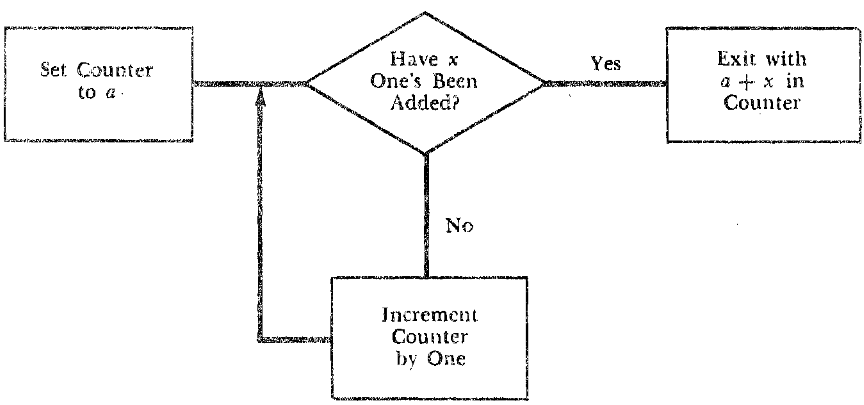

FIGURE 1.—EXAMPLE OF A DEVICE WHICH SETS A COUNTER TO *a* AND ADDS *x* TO *a*

Type 3. $x = m$.
Type 4. $x = \max(m, n)$.
Type 5. $x = \min.(m, n)$.

If we wish to apply this model to the latencies of our experimental subjects it cannot be assumed that the values of α and β are constant. Rather, it is correct to assume that α and β are random variables with two different distributions. However, we can eliminate this problem by taking the mean latencies, $E(\alpha)$ and $E(\beta)$, over all subjects. We then have, for a particular problem i,

$$E(T_i) = E(\alpha) + x_i E(\beta), \tag{2}$$

where x_i is computed according to the rules given above. For Equation (2) to hold, it is necessary, of course, that x_i be constant for all subjects on a given problem. In other words, it is necessary to assume that all subjects use the same type of solution. If this assumption is incorrect, then the goodness of fit of observed-to-predicted data will be affected.

In order to evaluate the goodness-of-fit of these five models, it is necessary to estimate the expected values $E(\alpha)$ and $E(\beta)$. These estimates will be denoted by $\hat{\alpha}$ and $\hat{\beta}$. For each problem, it is possible to compute a value of x_i under each of the five assumptions. Since Equation (2) is linear, $\hat{\alpha}$ and $\hat{\beta}$ can be computed for each model by means of a simple regression analysis, using x_i as the independent variable and the observed average-success latency on each problem as the dependent variable, with the index i ranging over all twenty-one problems. It is necessary to use the success latency rather than the overall latency for the dependent variable because it is reasonable to assume only that Equation (2) holds for correct solutions.

An analysis of this type was performed on the data obtained on the third day of the experiment. Two problems ($3 + 0 =$ __ and $2 + 3 =$ __) were omitted from the analysis. On both these problems, many individual response latencies were excessively high. The former was always the first problem to be presented. The high latencies on the latter can also be accounted for on the basis of sequential ordering effects. From the data obtained from the remaining nineteen problems, $\hat{\alpha}$ and $\hat{\beta}$ were evaluated for each of the five models, and two indexes of goodness of fit were computed. The first was the mean squared deviation between predicted and observed values:

$$s^2 = \frac{1}{17} \sum_{i=1}^{19} (T_i - \hat{\alpha} - \hat{\beta} x_i)^2,$$

where T_i denotes the observed success latency for problem i. Also computed was the ratio of $\hat{\beta}$ to the standard error of $\hat{\beta}$. If T_i is normally

TABLE 1
REGRESSION ESTIMATES FOR THE DIFFERENT SOLUTION TYPES
($\hat{\alpha}$ AND $\hat{\beta}$ MEASURED IN SECONDS)

MODEL	$\hat{\alpha}$	$\hat{\beta}$	s^2
1. $x = m + n$.	2.96	.216	.369
2. $x = n$.	3.50	.098	.465
3. $x = m$.	3.48	.119	.404
4. $x = \max(m, n)$.	3.43	.092	.471
5. $x = \min(m, n)$.	3.26	.710	.233

TABLE 2

MODEL 1: $x = m + n$.				MODEL 5: $x = \min(m, n)$.			
Problem	x	Mean Success Latency (in seconds)		Problem	x	Mean Success Latency (in seconds)	
		Pred.	Obs.			Pred.	Obs.
0 + 0	0	2.96	2.98	0 + 0	0	3.26	2.98
0 + 1	1	3.18	3.36	0 + 1	0	3.26	3.36
1 + 0	1	3.18	3.27	1 + 0	0	3.26	3.27
0 + 2	2	3.40	3.57	0 + 2	0	3.26	3.57
1 + 1	2	3.40	2.67	2 + 0	0	3.26	2.88
2 + 0	2	3.40	2.88	0 + 3	0	3.26	3.45
0 + 3	3	3.61	3.45	0 + 4	0	3.26	3.48
1 + 2	3	3.61	4.20	4 + 0	0	3.26	3.40
2 + 1	3	3.61	4.28	0 + 5	0	3.26	2.85
0 + 4	4	3.83	3.48	5 + 0	0	3.26	3.03
1 + 3	4	3.83	4.18	1 + 1	1	3.97	2.67
2 + 2	4	3.83	3.90	1 + 2	1	3.97	4.20
3 + 1	4	3.83	4.04	2 + 1	1	3.97	4.28
4 + 0	4	3.83	3.40	1 + 3	1	3.97	4.18
0 + 5	5	4.05	2.85	3 + 1	1	3.97	4.04
1 + 4	5	4.05	4.49	1 + 4	1	3.97	4.49
3 + 2	5	4.05	5.15	4 + 1	1	3.97	4.53
4 + 1	5	4.05	4.53	2 + 2	2	4.68	3.90
5 + 0	5	4.05	3.03	3 + 2	2	4.68	5.15

distributed, then this has a t distribution with $n - 2$ degrees of freedom. (In the present case, $n = 19$ [problems], so that $n - 2 = 17$. Although the summation is over subjects, as well as over problems, the details have been omitted so as not to obscure the basic ideas.) While it is not entirely clear whether or not the assumptions of the test are satisfied in the present experiment, its application does provide a rough index of whether or not the fit is satisfactory. The values of $\hat{\alpha}$, $\hat{\beta}$, and s^2 resulting from the analyses of the various models are shown in Table 1. Model 1 and Model 5 provided the best fits (i.e., s^2 was smallest for these models). The second goodness-of-fit computation resulted in levels of significance beyond .05 for all but these models. The predicted success latencies obtained on the basis of Model 1 and Model 5, together with the corresponding observed mean success latencies, are shown in Table 2. Notice that, especially in Model 5, each value of x involves a number of data points. As a result, a clearer notion of the fit can be obtained by comparing the predicted latency with the observed latency averaged over all problems that contribute to a given value of x. This is done in Figure 2. It is clear that the best fit is provided by Model 5. Although there are

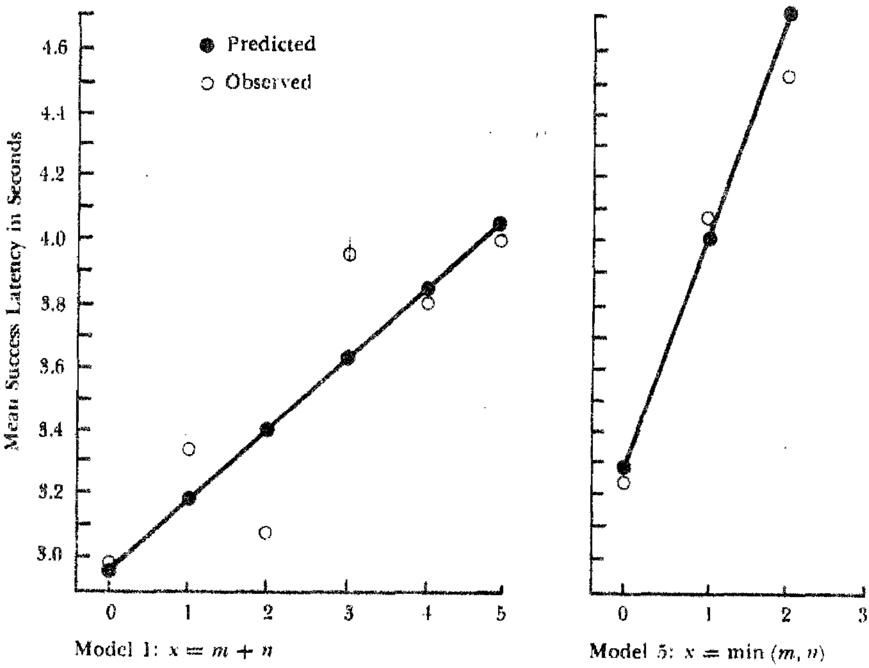

FIGURE 2.—PREDICTED AND OBSERVED MEAN SUCCESS LATENCIES PLOTTED AS A FUNCTION OF x FOR THE TWO BEST FITTING MODELS

less values of x in Model 5, the better fit cannot be ascribed to this circumstance, since the value of s^2 is lower for Model 5, despite the fact that s^2 was computed for each model on the basis of deviations between predicted and observed latencies for individual problems. Further evidence in favor of Model 5 is the fact that, with the exception of $1 + 1$, *all* problems with $x = 1$ have larger latencies than those with $x = 0$.

While it can safely be concluded that Model 5 fits better than Model 1, this result can only be considered to be a first step. There is no guarantee that no other model exists that would fit the data in a more satisfactory fashion. Moreover, it cannot be inferred that the good fit of Model 5 implies that subjects tend to add two numbers according to the mechanism suggested by the model. For this model, x ranges from 0 to 2. It is only when $x = 2$ that neither a 0 nor a 1 appears in the problem. Hence the data might be accounted for by a model which assumes specific algorithms for solving problems involving a 0 or 1 rather than the general algorithms used by the models we have proposed in this paper. Finally, there is, of course, the possibility that different individuals use different algorithms. Subsequent research that deals with these matters is now planned.

Some Concluding Remarks

It would be good if we could report that the algorithm represented by Model 5 was the one explicitly taught the children by their teachers. This does not seem to have been the case. At the present time most first-grade teachers do not teach their students an explicit counting algorithm for handling the simple addition facts ordinarily taught in the first grade. As would be expected there is usually some mention, and often even a fair amount of discussion, of counting and its relation to the first introduction of addition. But—and this is the important point—an explicit algorithm is not developed and taught as is done later for addition of multi-digit numbers.

The results of the present paper suggest that more attention might profitably be devoted to these first algorithms, and that the algorithm of Model 5, which seems more sophisticated than that of Model 1, might well receive more explicit emphasis in the teaching of first-grade arithmetic.

It has not been our intention in this short paper to present any definitive research, but only to illustrate how even so simple a thing as learning the addition facts presents an interesting challenge to learning theorists and affords an opportunity to test some alternative mathematical models, each of which rests on a clear intuition of how a simple addition problem may be solved. The central idea of a counting model seems so natural that it seems difficult to think of other possible approaches, but this is not really the case—for example, a table-look-up model with parameters appropriately introduced for scanning the table can be formulated in such a way that it is identical in all behavioral predictions with Model 5. Moreover, simple counting ideas are not sufficient to account for all the significant variations in the observed data of Table 2, and as a larger body of data is accumulated, more complex and subtle ideas will be needed in constructing an adequate model of the observed phenomena. On the other hand, it seems to us that the learning of elementary mathematics affords a natural testing ground for mathematical models of learning or performance, and there is some reason to hope that in a first approximation, at least, models of a reasonable degree of simplicity will suffice.

It should be apparent that as such models are developed and the range and depth of their success is increased they will have increasing significance in suggesting and guiding curriculum modifications, particularly as regards the fundamental problem of finding out how students can on the *average* best learn mathematical concepts and skills.