Reprinted from AMERICAN EDUCATIONAL RESEARCH JOURNAL, Vol. 1, No. 2, March 1964

MODERN LEARNING THEORY AND THE ELEMENTARY-SCHOOL CURRICULUM*

PATRICK SUPPES

Institute for Mathematical Studies in the Social Sciences
Stanford University

INTRODUCTION

What I want to do this evening is to sketch for you some of the implications of recent research in learning theory for the elementary-school curriculum and to indicate how I think these implications can best be realized by appropriate application of modern computer technology. I shall say a number of things about computers, but I want to emphasize at the very beginning that I do not intend to let the fascinating problems of computer technology distort the primary focus on learning and the curriculum. The research and curriculum work done thus far as a result of the current interest in teaching machines and programed instruction show clearly that technology alone is not going to produce any fundamental or long-lasting changes in the curriculum. Only if its applications are guided by appropriate psychological principles and subject-matter insight will such changes be brought about.

To show that I mean what I say about keeping technology at heel, I will turn directly to certain results derived from recent work in learning theory and indicate some of their pedagogical implications. Only after that has been done, will I discuss how computers can be used to implement these results.

INDIVIDUAL DIFFERENCES IN LEARNING

One of the most firmly established generalizations of behavioral psychology is the existence of significant differences in individual rates of learning. In spite of the obeisance paid to this tenet in discussions of curriculum, I consider it the most important principle of learning as yet *unaccepted* in the day-to-day practice of subject-matter teaching in the classroom. In many large elementary schools, homogeneous grouping of

*The substance of this paper was given as an invited address to the annual AERA meeting in Chicago on February 20, 1964. The research described here has been supported by several sources: contracts OE 3-14-020 and OE 3-10-009 between Stanford University and the U. S. Office of Education, the National Science Foundation (Course Content Improvement Section), and the Carnegie Corporation of New York.

children at the same grade level is done routinely. The recent movement toward the ungraded elementary school constitutes a significant development for the accommodation of individual differences in learning. However, I venture to suggest that the staggering implications of these differences have scarcely been realized in any school in the country.

Let me give an example from some recent pedagogical experience of my own. Beginning with the current academic year we have undertaken at Stanford to give an intensive accelerated program in mathematics to a group of gifted first graders for the 6 years of their elementary-school attendance. The initial group of 40 students was selected by giving the *New York Test of Arithmetical Meanings* to all the children entering the first grade in four schools. Those above the 70th percentile on this test of first-grade achievement were then tested more intensively, most of them being given a form of the Stanford-Binet. The 40 children selected by this procedure have IQ's ranging from 122 to 166 with a mean of 137. They are being taught mathematics in the four schools in small groups of 9 to 12 pupils each. For each child, we are collecting subject-matter learning data, consisting of a daily list of the problems worked and the problems answered incorrectly. Those of you who have attempted detailed behavioral analysis of subject-matter learning will know that this is no small task, particularly when the pupils are bright and therefore able to move rapidly through the course work. Because the children are being handled in small groups, it is possible to give them something close to individual treatment. From our standpoint, the most significant aspect of this individual treatment is the fantastic differences in rate of learning. At the end of the first four weeks, the fastest child in the group covered approximately half again as much material as the slowest (and I hasten to add that the relative rates of progress are not strongly correlated with the IQ scores). To give you a sense of numbers, during the first 7 weeks (the classes meet 4 days a week with each session lasting not more than 35 minutes) the fastest child worked approximately 3,400 problems with an error rate less than 2 per cent. During this same period the slowest child in this highly selected group worked approximately 2,200 problems. Learning curves for these two students are shown in the figure. The number of weeks of participation is shown on the abscissa and the cumulative number of problems worked on the ordinate. A calibration of the number of problems in terms of the *Sets and Numbers* elementary-school mathematics series I have been writing is also shown on the ordinate. Books 1A and 1B constitute a full first-grade curriculum, and Book 2A the first half of a full second-grade curriculum. At the end of the seventh week, these two bright first graders are already separated by almost a third of the second-grade curriculum. For understandable reasons, we are reluctant to forecast the maximum separation after this class has continued for another couple of years. The rate at which these bright children are working problems and proceeding through the curriculum is most surpris-

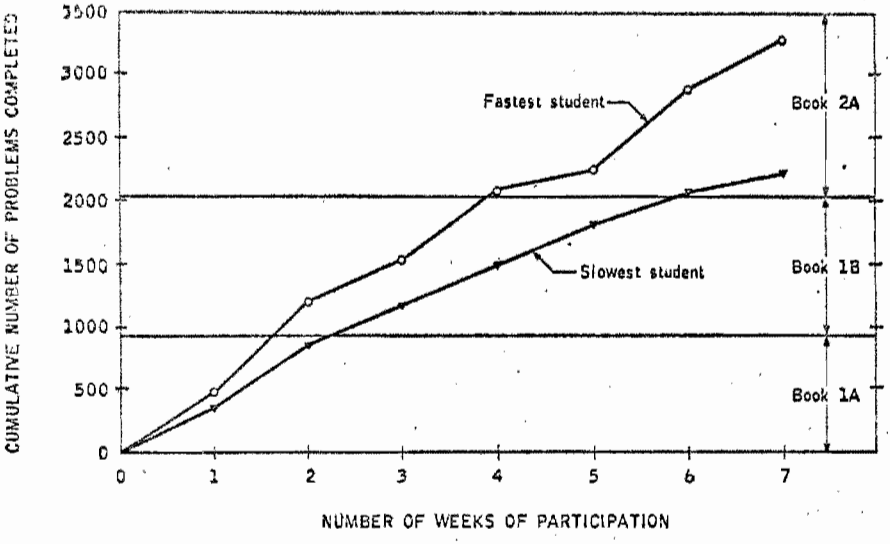

Fig. 1. Progress of Fastest and Slowest of Forty Gifted First-Grade Pupils in Mathematics.

ing; it has certainly exceeded our initial expectations. We are not encouraging a speed contest; in fact, the teachers are making a decided effort to discourage intensive competition. These comparisons are, I think, a fair representation of the very large differences to be expected among even very bright children when the rate of progress of the individual child is not tied to that of the group.

Because we wanted to know what learning rates to anticipate, we attempted to search the literature on mathematics learning in the elementary school. It is disappointing to find how few hard data are reported on the relative rates of learning of bright or slow children under teaching conditions approximating individual treatment.

To underscore what is probably obvious to all of you anyway, let me describe a recent experiment I performed jointly with Edward J. Crothers on the beginning stages of learning to read by kindergarten children. In contrast to the mathemathics project just described, this was a controlled experiment with all subjects being run by an individual investigator in an isolated experimental room. The subjects were not selected for ability, but were all kindergartners in an elementary school near Stanford. To build up a working vocabulary without requiring many months of training, we selected 14 letters of the alphabet from which a reasonable number of familiar words could be constructed. I won't attempt to describe all aspects of the experiment, which in total design was fairly complicated. For the purpose of my present remarks the essential facts are these. The subjects were trained to a criterion on 14 single words built up from the 14 letters, the criterion being 28 successively correct responses. After this

criterion was satisfied, they were given a list of 14 two-word phrases, then three-word phrases, then four-word phrases and five-word phrases, all built up from the initial vocabulary of 14 words. To obtain quick and accurate error data, the subjects were trained in the following way: on each trial the experimenter said a word or phrase and the child responded by pointing to one of three possible graphemic representations of the auditory stimulus. (I emphasize that this experimental procedure is not recommended as a method of teaching reading. It was an efficient and objective procedure for studying the acquisition of the initial phases.)

Thirty-eight kindergarten children completed the experiment. The point I want to emphasize is the wide range in the total number of trials required. The fastest child needed only 196 learning trials while the slowest child needed 2,506. This range is not misleading, for the mean was 967.4 with a standard deviation greater than 400.

Data of this sort, which may be multiplied tenfold by references to the literature, argue that by far the greatest improvement in subject-matter learning will result from an almost single-minded concentration on individual differences. I am sure that the optimum sequencing of curriculum materials, the analysis of the subject matter so as to present it in steps of the proper size, etc., are not nearly so important elements in learning as the single one of individual accommodation. I shall have more to say on this matter when I discuss computer technology.

Immediate Reinforcement and Overt Correction Procedures

A variety of experimental studies has shown the importance of both immediate reinforcement and overt correction in children's learning. As a typical example, let me mention one experiment we conducted three years ago at Stanford (Suppes and Ginsberg, 1962). There are good studies (e.g., Burke, Estes, and Hellyer, 1954) to show that requiring an adult subject to make an overt correction response after reinforcement neither increases his learning rate nor influences appreciably his asymptotic behavior. Our problem was to determine whether these findings would hold for primary-grade children. The experiment concerned recognizing the abstract patterns involved in notation for any number system—in this case the numbers four and five in binary notation. On each trial a stimulus was presented to the subject to which he could make one of two responses: he could place the stimulus on top of an Arabic numeral 4 or on top of an Arabic numeral 5. Six different stimuli were presented, three of them (*e.g.*, αββ, λππ, δεε) representing four in the binary number system and the remaining three (*e.g.*, αβα, λπλ, δεδ) representing five. There were 48 subjects in the study, half of them coming from grade one and half from a kindergarten class. Twenty-four of the subjects, 12 kindergartners and 12 first graders, were asked to make an overt correction response after a wrong choice, that is, after placing a stimulus card on the wrong Arabic numeral. The other 24 subjects were

not required to make such an overt correction response, although they were told on each trial whether or not their responses were correct. The learning curves for the two groups showed a significantly faster rate of learning for the overt correction group throughout the entire experiment, which consisted of 96 trials. A *t*-test computed on the over-all responses of the two groups was significant at the .001 level.

Other studies demonstrating the desirability of overt correction procedures to facilitate children's learning may easily be cited. On the other hand, those of you who have done much observing of the environment of elementary-school classrooms realize fully the difficulties facing the teacher in attempting to provide immediate reinforcement and overt correction. If we take a typical lesson in arithmetic, a little simple arithmetic itself will show these difficulties. As a conservative estimate, it is not uncommon in the latter part of the first grade for a workbook page to contain ten problems. Let us suppose, again conservatively, that the class consists of only 25 children. If the class is working this particular page together, the teacher cannot hope to monitor the 250 problems being done and, on the spot, have each child make appropriate corrections. Some of the elementary-school mathematics teachers with whom I have worked have told me that in order to do a good job in the first or second grade it is necessary to spend at least one hour after each school day correcting the children's workbooks. Some of them insisted that a more realistic estimate would be two hours. I have watched teachers who have gone through this laborious process of marking each child's work on each problem and requiring him to correct his errors the next day. I wish I were able to report some experimental data on the efficacy of overt correction delayed by at least one day. The impression of the teachers who have tried this procedure has been positive, but I cannot cite any controlled data. Again, however, I want to emphasize that technological methods are almost necessary in order to provide overt correction procedures, particularly with immediate reinforcement, on a sustained basis. This is all the more true if we do not restrict ourselves to mathematics but think about the other parts of the curriculum, such as foreign language, elementary science, and social studies.

Specific Nature of Transfer

One of the most important and fundamental measures of learning is the extent of transfer to new situations. The classical paradigm for evaluating the benefits of a new curriculum is in terms of transfer, with the experimental group receiving training of Type A, the control group training of Type B, and both groups then being tested on transfer items of Type C. A comparison of performance on the transfer items is the primary method of judging the efficacy of the new treatment given the experimental group.

Transfer and generalization have been a major focus of investigations

by learning theorists. In almost all the fundamental work on stimulus generalization and transfer in the literature of mathematical learning theory during the past ten years (e.g., that by Estes, Bush and Mosteller, Restle), the similarity between stimulus patterns, or displays, is represented by the amount of overlap between two sets of stimulus elements. In the simplest experimental paradigm, only two stimulus situations are treated. Training takes place on one, and generalization or transfer to the other is tested. The amount of generalization or transfer is taken to be a measure of the amount of stimulus overlap in terms of the set of stimulus elements common to the two situations. A number of ingenious and important experiments of this general type have been conducted within the domain of learning theory in the past fifteen years, but it is not appropriate to review them here.

There is, however, a common failure in the reports of these experiments, which is reflected in the paradigm of educational research sketched above. This common failure may be characterized as a lack of structural depth in theory. It is a well-known mathematical fact that sets without structure provide little basis for the elaboration of a theory.

Experiments on mathematical concept formation provide an excellent opportunity to dig deeper into the structure of the *set* of stimuli, because in most instances of such concept formation a natural and clearly defined structure is imposed by the mathematical character of the concept itself. Let me give an example that I have studied experimentally in collaboration with Rose Ginsberg (Suppes and Ginsberg, 1963). Given any pair of sets or classes of objects, we may classify the pair of sets in one of the four following mutually exclusive and exhaustive ways. The first possibility is that the two sets are identical in terms of their members and in terms of the order assigned to these members. The second possibility is that two sets are identical in the sense of having the same members, but the members are not given in the same order. For example, the set consisting of the numbers 1 and 2 is identical in this sense to the set consisting of the numbers 2 and 1. The third possibility is that the two sets are *equivalent* but not identical; that is, they have the same number of elements but the elements are different. For example, the set consisting of 1 and 2 is equivalent but not identical to the set consisting of 3 and 4. The fourth and final possibility is that the two sets are not equivalent; that is, they do not have the same number of elements. These four possibilities, or categories, may be used as four subconcepts for analyzing the learning of more general concepts.

In one of our experiments concerned with these subconcepts, first-grade subjects were given 56 trials on identity of ordered sets and then 56 trials on identity of sets without concern as to the order of the members. Similar subjects were given 56 initial trials on identity of unordered sets. The sets pictured by the stimulus displays consisted of one, two, or three elements. On each trial, two of these sets were displayed. Subjects

were told to press one of two buttons—one button when the two stimulus displays presented were "the same" and the other button when they were "not the same." When the correct response was made, the appropriate one of two reinforcing lights flashed. Using simple principles of interference, we might expect negative transfer in going from identity of ordered sets to identity of sets. That this rather unsurprising result did occur is easily seen by comparing the mean learning curve for the children who began the experiment on identity of sets with the curve for the children who learned identity of sets *after* identity of ordered sets. The evidence of negative transfer from comparison of these learning curves is clear. The interesting question of greater depth is: can we identify the point of negative transfer in terms of the structure of the four subconcepts described above? The answer is clearly affirmative. In terms of the four categories, the pairs that are identical but not in the sense of order and *only these* require a different response in passing from the subconcept of identity of ordered sets to that of identity of unordered sets. And all the negative transfer is indeed isolated in the pairs that are instances of this subconcept. For the other three subconcepts—identity of ordered sets, equivalence of sets, and nonequivalence of sets—there is, in fact, slight evidence of positive transfer. Detailed data are presented in Suppes (1963b). The highly specific nature of the transfer results obtained here is undoubtedly present in other concept learning.

The identification of the structure of subconcepts determining the nature of transfer is a central problem for our future work in mathematical concept formation, and I hope soon to have more results to report.

Optimal Block Size in Learning a List of Simple Items

The learning task I want to take up under this heading is, on the face of it, a very simple one. We want subjects to learn a list of items. The list is rather long, so we may consider the possibility of breaking up the list into blocks of varying size.

A number of applications of the results of studies of this kind can be mentioned. Among the most interesting are those arising in learning a second language. The items may be foreign words, and the task is to learn their approximate meanings in English. In psychological terms, this means that to each word of the foreign language in the list the subject is asked to associate a unique English word. We all know that the words and phrases of a second language vary in difficulty. In graphemic representation some words are almost identical in the two languages although at the phoneme level they may differ considerably; e.g., "application" is graphemically the same but phonetically different in English and French.

In the present discussion, I want to ignore all these differences in difficulty. Suppose we want subjects to learn an initial Russian vocabulary of approximately 100 words. How should we work our way through this list of words in order to maximize learning? We may fix the total number of

trials, with the terminal trials used as test items. The problem is to decide how the remaining trials, apart from the test, should be allocated for training purposes. At the end of the training trials, each item will have a mean error probability and in order to maximize learning we want to minimize this mean error. Items are to be given in blocks of a fixed size, and our problem is to determine what this size should be.

Fortunately, under quite general assumptions about learning and forgetting, some recommendations about block size can readily be reached. Assuming that learning is faster than forgetting, we can show that the block size should be as large as possible, which means that in the case of our 100 Russian words the subjects should be taken through the entire list once before a second presentation of any item. On the other hand, when learning is slower than forgetting, the block size should be as small as possible; that is, all exposures of a given item should be contiguous. In most situations, learning seems to be faster than forgetting and thus, according to the theory, the largest possible block size should be used.

The formal basis of these results is easily described in mathematical terms. Let $p_{i,n}$ be the probability of a correct response to item i on trial n, and let $q_{i,n}$ be the probability of an incorrect response to item i on trial n. Then

$$q_{i,n} = 1 - p_{i,n}.$$

When an item is presented and the correct response reinforced,

$$q_{i,n+1} = \alpha q_{i,n};$$

i.e., there is learning and thus a decrease in the error probability. (We may interpret $q_{i,n}$ as the *average* probability of an error if we wish to hold an all-or-none view of the learning of individual items by individual subjects.)

On trials when other items are presented, there is forgetting and thus a decrease in $p_{i,n}$:

$$p_{i,n+1} = \beta p_{i,n}.$$

The learning parameter is α and the forgetting parameter is β, and $0 \leq \alpha, \beta \leq 1$. When $\beta > \alpha$, learning is faster than forgetting; learning is fastest when $\alpha = 0$, and therefore the largest possible block size should be used. Mathematical details of this derivation are given in Suppes (1963a). From a psychological standpoint, an analysis of this sort is closely related to the classical experimental literature on part vs. whole learning. See, for example, Woodworth and Schlosberg (1954).

When I first obtained these simple results about block size based on a purely theoretical argument, I was skeptical that they would hold for a very wide range of learning phenomena. However, Edward J. Crothers, Ruth Weir, and I have recently been testing them in some experimental work with Russian and they have held up well. In the first experiment completed, the task presented to subjects was learning 108 Russian words. On each trial the subject heard the Russian word spoken in a short phrase.

At the same time he was shown three written translations of the phrase, with the word in question being the source of variation. Four groups of subjects were run. One group received the vocabulary items in blocks of 6, the second group in blocks of 18, the third group in blocks of 36, and the fourth group in blocks of 108 (that is, the total list). Each item was given 21 reinforcements and thus there were about 2,200 trials for each group. It required 13 days of one-hour sessions for subjects to complete the experiment. The group of subjects that had blocks of 108 performed slightly better than the other groups. As expected theoretically when learning is faster than forgetting, the 6-block group was the worst.

We are now testing the theory with a second experiment using a basic vocabulary of 216 words. This experiment is not yet complete, but at present the 216-block group is doing better than the 108-block group.

This work on Russian has been conducted with college and junior-high-school students. We are now planning some work on the learning of Russian by elementary-school pupils, and our planning of this program has been very much influenced by the experiments done thus far with older subjects. I expect our main findings to hold for this younger group as well.

Response Latency as Criterion of Learning

Although interest in response latencies, or reaction times, is a very old one in psychology, it is surprising how little systematic work has been done during the past decade in this area, either on a theoretical level or in experimental studies. Only during the last year or two has there been a serious revival of interest in response latency as a variable that may serve as a central criterion of learning. Some of this recent experimentation is showing clearly what all of us probably know in a muddle-headed way: namely, that, for many kinds of learning, latencies are a much more sensitive index than are response errors themselves. For at least three areas of the elementary-school curriculum—reading, mathematics, and foreign language, there is no doubt that response latencies are more sensitive measures of skill mastery and depth of learning than the responses themselves. A certain amount of work on latencies has been done with respect to reading, but not of the sort that is very satisfactory from an experimental standpoint. In the case of elementary mathematics, I have recently had an assistant searching the literature. He has found very little that is directly relevant. Since what happens to response latencies under continued practice of the algorithms of arithmetic is such a natural subject for investigation, I can hardly believe that detailed studies in this area do not exist. I have been particularly surprised not to find a substantial body of literature from the twenties on this topic. In any case, an extensive series of studies beginning with first-grade children is now under way in our laboratories, and by this time next year I may be able to report some interesting results. Interest in response latencies in the doing of simple arithmetical calculations is not a matter of sheer empiricism. A few ele-

mentary computations show clearly that even the facts of arithmetic involved in sums and differences of numbers not greater than 10 are far too numerous to be simply stored in memory. Simple algorithms of some sort are almost surely learned in working problems like "4 plus what number is equal to 7?" Structural hypotheses about the nature of these algorithms are not easily verified. One of the few possibilities for getting evidence in this area lies in the quantitative analysis of latency data. It is fair to claim that as yet we know very little about how algorithms are learned, and until a fundamental theory of some validity is developed, we shall continue to stumble around in much of our discussion of how arithmetic should be taught.

There have been a few isolated studies of response latencies in second-language learning; for example, Lambert (1955) has done some interesting work on the measurement of "fluency" in bilinguals. However, in view of the obvious relevance of response latencies to measures of second-language mastery, the paucity of serious experimental studies in the literature is almost shocking.

To indicate some of the ways in which response latencies are sensitive measures of learning, let me describe a recent study of ours dealing with the recognition of spoken Russian words. On each trial the stimuli were spoken by a native speaker. The Cyrillic response alternatives were shown in multiple-choice format. When the subject heard a word, he pressed the key corresponding to one of the alternatives and was immediately informed of the correct answer. As an independent variable, the Russian items were classified according to the similarity of their constituent printed characters to one another and to the letters of the Roman alphabet. For example, we put in one category Russian words whose graphemic representation is composed entirely of letters that also occur in the Roman alphabet. On every trial the response latency as well as the response of the subject was recorded. We found that latencies strongly depended on the type of item. For example, longer latencies were associated with items having Cyrillic characters similar to each other and distinct from Roman letters. It was possible to divide the items into five categories and have for all sessions a clear separation of the latencies. Most important, even after the response-latency curves had flattened out to an asymptotic level by essentially the tenth session, the clear separation in response latencies for the five types of items remained. A foremost problem for language teaching is to devise methods for breaking through these asymptotic levels of response latency by finding methods that will, with sufficient practice, drive them down to the asymptotic latency levels of the subject's native language.

Computer Technology

I said that I would have a fair amount to say about computer technology and the possibilities it opens up for application of learning theory

to the elementary-school curriculum. Let me begin by describing the computer-based laboratory for learning and teaching we are now constructing at Stanford. (The Executive Committee of the laboratory consists of Richard C. Atkinson, William K. Estes, and me.) Then I shall indicate how the laboratory may be used to implement some of the research findings already discussed.

The core of the laboratory is a small fast computer, which controls a variety of terminal equipment located in six student booths. Because the computer is used in a time-sharing mode, each student works independently of the others. He faces a 16-inch cathode tube that looks like a television screen. This cathode tube—"scope" as we call it—is connected directly to the computer, and visual stimulus displays are generated directly from programs in the computer. One of the great advantages of the scope is that the student's responses to a stimulus display may be immediately made an integral part of the display itself. The student has two methods of response available: one is to use a light probe to make selections or indicate choices on the face of the scope; the other is to use a standard keyboard that is placed just below the scope.

The second important device for the visual display of stimuli is a microfilm unit. The equivalent of a 512-page book may be encoded on microfilm and any quarter of a page may be reached under random access within one second for display on a visual frame slightly larger than the standard 8½" by 11" page. The student may respond to a frame presenting a problem by using a light probe on the face of the display itself. Multiple-choice responses are not the only kind available with the microfilm equipment; the possibilities of constructed responses are quite substantial. As part of the visual display of a page, we may show a keyboard at the bottom of the page and the student may respond on the keyboard by indicating the appropriate sequence of letters with his light probe.

Since visual stimuli are not sufficient for young students, we can also present auditory stimuli—either through individual speakers in each student's booth or through high-fidelity earphones. The auditory system will be of sufficient fidelity to permit work with foreign languages. Finally, we plan to have a closed-circuit-television setup, so that we will be able to show the students a televised instructor. I emphasize that both the auditory stimuli and the instructional material on videotape will be under computer control, which means that the material will not have to be worked through in a simple linear order, but can be presented according to the needs and learning rate of the individual student.

What we hope to obtain with this computer-based laboratory is a reasonably good simulation of the teaching environment created by a tutor working with an individual student. In terms of the five phases of learning research I have mentioned, let us see how this computerized environment will facilitate presentation, organization, and behavioral analysis of curriculum material.

The application to individual differences, which I mentioned first, is apparent. Since each student will proceed at his own pace (independent of the progress of the other students), individual differences can easily be accommodated. The computer-based terminal equipment can also provide immediate reinforcement and overt correction when necessary. In different parts of the curriculum, different approaches to immediate reinforcement and overt correction may be adopted, but it should be clear how easily and directly a variety of procedures may be instituted. Let me give just one example, which concerns the teaching of place value in the first grade. The child is asked to decompose a three-digit number into the number of hundreds, the number of tens, and the number of ones. He has been shown on the scope the sentence,

456 is _____ hundreds _____ tens _____ones.

Above the hundreds place is a small star indicating, like the bouncing ball of singing movies and television, that this is the first blank he is to fill. If he makes the correct response, the numeral 4 is immediately placed in the blank and the star then moves to the next blank. If in making the next response he makes an error, a large X appears and he knows that he must continue to respond until the correct number is given.

For the past several years we have been conducting an experimental pedagogical program in the teaching of mathematical logic to able fifth and sixth graders. We have just recently begun to experiment with how this program may be adapted to the special environment of our new laboratory. Let me describe how we can apply immediate reinforcement and correction even to the variety of individualized mathematical proofs students construct very early in their training. Also I would like to indicate how we have begun to use the computer to reduce, rather than increase, the amount of tedious work on the part of the student. A good portion of the course consists of formally deriving conclusions from given premises. When the premises are symbolized in mathematical form, the student in the usual teaching situation is asked to write out each line of derivation and to justify it according to an appropriate logical rule of inference. With the facilities available in the new laboratory we have been able to shorten this procedure and to require the student to think only of which rule of inference to apply and which lines he wants to use to get the new line of derivation. He then types in on the keyboard the abbreviation for the rule and the numbers of the lines he is going to use. The computer does the rest as part of its teaching program and displays on the scope the new line of proof. If an incorrect response is made, the reason why the rule typed in is inappropriate is displayed in place of the line requested by the student. I emphasize that in general the correct response to be made by the student is not unique. Any appropriate response is accepted, i.e., any response that satisfies the criterion of being a logically valid application of the rules of inference. It is most impressive to

watch the speed with which able fifth graders accommodate to this new environment and rapidly proceed to work a large volume of problems.

In relation to my earlier remarks about transfer, the most immediate application of the new laboratory is to permit an efficient analysis of transfer among concepts over a substantial period of time. If we want to follow 30 students, let us say, through the first two years of the elementary mathematics curriculum, the task of collecting, organizing, and analyzing response data is almost overwhelming without a computerized facility. One of our first aims will be to obtain this sort of data in order to make a deeper analysis of transfer phenomena from one mathematical concept to another as the child progresses through the early phases of his mathematics education. We are particularly interested in seeing what the effects will be of weak and strong criteria of attainment on each concept before passing to another concept. How will the variation in criterion of performance affect transfer to a new concept and subsequent learning?

The application of results of optimization to the organization of stimulus material scarcely needs detailed comments. In students' work with the elementary parts of a foreign language, for example, it is apparent that a properly prepared computer program can be sensitive to points of difficulty in a student's performance, certainly at the level of language recognition, in at least as effective a way as all but the best language teachers. The major criticism I would make, from a learning standpoint, of the vast majority of language laboratories is this. The behavioral aspect of language learning has not been made an integral part of the work. Too often the student is *passively* listening to linguistic material, when it is a simple matter with any sort of computer facility to arrange for recognition responses to be made and evaluated. In learning to recognize spoken words and phrases in a foreign language, the student can be given repeated opportunities to make the appropriate identification. Those items he finds easy can be dropped from the list of items presented and new items inserted as the occasion demands. Even the best teacher would have difficulty keeping up with the rapid tabulation and decision to insert new items that can be made with a computer program.

The same comparison in favor of the computer-based laboratory can be made when we attempt to apply response latency as a criterion of learning. Even the best individual tutor working with an individual student finds it difficult to make an accurate appraisal of response times. An approximation that is not too bad can be made in the case of arithmetical skills—for example, speed skills in doing the longer algorithms of arithmetic—because the times involved are relatively long. The situation is quite different in terms of language recognition and production. The response times involved are short, on the order of one second, and it is not an easy matter for the experienced tutor to make an accurate discrimination about the performance of the learner. I daresay that, in the latency

experiment I described, none of us here would have made a very successful discrimination in the response times for the five different kinds of items in terms of their relation to English sounds and graphemes. Moreover, the imposition of a sophisticated latency criterion of performance can be made with great ease in a computer-based laboratory but is extremely difficult even in work with an individual child by an experienced tutor. It is our own conjecture that the use of latency criteria will be one of the most important new aspects in which we shall be able to deepen our understanding of the relation between learning and the curriculum.

I have emphasized various detailed ways in which the computer-based laboratory can facilitate the applications of learning theory to the school curriculum. Let me conclude by mentioning the over-all point of greatest importance for the immediate future. A computer-based laboratory such as I have described is one of the few ways of building a satisfactory bridge between research in learning and curriculum work. The reasons are really not scientific, but technological. The difficulties of collecting an adequate amount of behavioral data on subject-matter learning are so great and the problem is so complex that it is difficult to conceive of doing an adequate job with simpler apparatus. The existence of large bodies of organized data will surely promote a more intimate interaction between systematic theories of learning and their application to education. The relation of psychology to education will come closer to what Dewey forecast it should be many years ago (McLellan and Dewey, 1895), "In a word, education itself is precisely the work of supplying the conditions which will enable the psychical functions, as they successively arise, to mature and pass into higher functions in the freest and fullest manner, and this result can be secured only by knowledge of the process—that is, only by a knowledge of psychology."

References

Burke, Cletus J.; Estes, William K.; and Hellyer, Sidney. "Rate of Verbal Conditioning in Relation to Stimulus Variability." *Journal of Experimental Psychology* 48: 153–61; September 1954.

Lambert, Wallace E. "Measurement of the Linguistic Dominance of Bilinguals." *Journal of Abnormal and Social Psychology* 50: 197–200; March 1955.

McLellan, James A., and Dewey, John. *The Psychology of Number and Its Applications to Methods of Teaching Arithmetic.* New York: D. Appleton, 1895. 309 pp.

Suppes, Patrick. *Problems of Optimization in Learning a List of Simple Items.* Psychology Series, Technical Report No. 57. Stanford: Institute for Mathematical Studies in the Social Sciences, Stanford University, July 22, 1963a.

Suppes, Patrick. "Mathematical Concept Formation in Children." Un-

published invited address to Division 15 of the American Psychological Association, August 31, 1963b.

SUPPES, PATRICK, and GINSBERG, ROSE. "Application of a Stimulus Sampling Model to Children's Concept Formation With and Without Overt Correction Responses." *Journal of Experimental Psychology* 63: 330–36; April 1962.

SUPPES, PATRICK, and GINSBERG, ROSE. "A Fundamental Property of All-or-None Models, Binomial Distribution of Responses Prior to Conditioning, With Application to Concept Formation in Children." *Psychological Review* 70: 139–61; March 1963.

WOODWORTH, ROBERT S., and SCHLOSBERG, HAROLD. *Experimental Psychology*. Revised edition. New York: Henry Holt, 1954. 948 pp.