DOCUMENT RESUME

ED 076 411 SE 015 992

AUTHOR Suppes, Patrick
TITLE Facts and Fantasies of Education.
INSTITUTION Stanford Univ., Calif. Inst. for Mathematical Studies in Social Science.
REPORT NO TR-193
PUB DATE 18 Oct 72
NOTE 72p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Curriculum; Education; *Educational Theories; Elementary School Mathematics; Instruction; Mathematical Models; *Mathematics Education; Reading; *Research; *Research Methodology

ABSTRACT

The relevance of research to educational ideas is discussed in this paper. General ways of organizing education in matters of theory, policy, or practice, and methodology are the "fantasies" analyzed; the need for research to produce the evidence that will convert these fantasies to fact is emphasized. A research example from elementary mathematics is briefly discussed, with technical details included in the appendix. (DT)

FACTS AND FANTASIES OF EDUCATION

BY

PATRICK SUPPES

TECHNICAL REPORT NO. 193

OCTOBER 18, 1972

PSYCHOLOGY & EDUCATION SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

TECHNICAL REPORTS

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

(Place of publication shown in parentheses; if published title is different from title of Technical Report, this is also shown in parentheses.)

(For reports no. 1 - 44, see Technical Report no. 125.)

50 R. C. Atkinson and R. C. Calfee. Mathematical learning theory. January 2, 1963. (In B. B. Wolman (Ed.), Scientific Psychology. New York: Basic Books, Inc., 1965. Pp. 254-275)

51 P. Suppes, E. Crothers, and R. Weir. Application of mathematical learning theory and linguistic analysis to vowel phoneme matching in Russian words. December 28, 1962.

52 R. C. Atkinson, R. Calfee, G. Sommer, W. Jeffrey and R. Shoemaker. A test of three models for stimulus compounding with children. January 29, 1963. (J. exp. Psychol., 1964, 67, 52-58)

53 E. Crothers. General Markov models for learning with inter-trial forgetting. April 8, 1963.

54 J. L. Myers and R. C. Atkinson. Choice behavior and reward structure. May 24, 1963. (Journal math. Psychol., 1964, 1, 170-203)

55 R. E. Robinson. A set-theoretical approach to empirical meaningfulness of measurement statements. June 10, 1963.

56 E. Crothers, R. Weir and P. Palmer. The role of transcription in the learning of the orthographic representations of Russian sounds. June 17, 1963.

57 P. Suppes. Problems of optimization in learning a list of simple items. July 22, 1963. (In Maynard W. Shelly, II and Glenn L. Bryan (Eds.), Human Judgments and Optimality. New York: Wiley. 1964. Pp. 116-126)

58 R. C. Atkinson and E. J. Crothers. Theoretical note: all-or-none learning and intertrial forgetting. July 24, 1963.

59 R. C. Calfee. Long-term behavior of rats under probabilistic reinforcement schedules. October 1, 1963.

60 R. C. Atkinson and E. J. Crothers. Tests of acquisition and retention, axioms for paired-associate learning. October 25, 1963. (A comparison of paired-associate learning models having different acquisition and retention axioms, J. math. Psychol., 1964, 1, 285-315)

61 W. J. McGill and J. Gibbon. The general-gamma distribution and reaction times. November 20, 1963. (J. math. Psychol., 1965, 2, 1-18)

62 M. F. Norman. Incremental learning on random trials. December 9, 1963. (J. math. Psychol., 1964, 1, 336-351)

63 P. Suppes. The development of mathematical concepts in children. February 25, 1964. (On the behavioral foundations of mathematical concepts. Monographs of the Society for Research in Child Development, 1965, 30, 60-96)

64 P. Suppes. Mathematical concept formation in children. April 10, 1964. (Amer. Psychologist, 1966, 21, 139-150)

65 R. C. Calfee, R. C. Atkinson, and T. Shelton, Jr. Mathematical models for verbal learning. August 21, 1964. (In N. Wiener and J. P. Schoda (Eds.), Cybernetics of the Nervous System: Progress in Brain Research. Amsterdam, The Netherlands: Elsevier Publishing Co., 1965. Pp. 333-349)

66 L. Keller, M. Cole, C. J. Burke, and W. K. Estes. Paired associate learning with differential rewards. August 20, 1964. (Reward and information values of trial outcomes in paired associate learning. (Psychol. Monogr., 1965, 79, 1-21)

67 M. F. Norman. A probabilistic model for free-responding. December 14, 1964.

68 W. K. Estes and H. A. Taylor. Visual detection in relation to display size and redundancy of critical elements. January 25, 1965, Revised 7-1-65. (Perception and Psychophysics, 1966, 1, 9-16)

69 P. Suppes and J. Donio. Foundations of stimulus-sampling theory for continuous-time processes. February 9, 1965. (J. math. Psychol., 1967, 4, 202-225)

70 R. C. Atkinson and R. A. Kinchla. A learning model for forced-choice detection experiments. February 10, 1965. (Br. J. math stat. Psychol., 1965, 18, 184-206)

71 E. J. Crothers. Presentation orders for items from different categories. March 10, 1965.

72 P. Suppes, G. Groen, and M. Schlag-Rey. Some models for response latency in paired-associates learning. May 5, 1965. (J. math. Psychol., 1966, 3, 99-128)

73 M. V. Levine. The generalization function in the probability learning experiment. June 3, 1965.

74 D. Hansen and T. S. Rodgers. An exploration of psycholinguistic units in initial reading. July 6, 1965.

75 B. C. Arnold. A correlated urn-scheme for a continuum of responses. July 20, 1965.

76 C. Izawa and W. K. Estes. Reinforcement-test sequences in paired-associate learning. August 1, 1965. (Psychol. Reports, 1966, 18, 879-919)

77 S. L. Blehart. Pattern discrimination learning with Rhesus monkeys. September 1, 1965. (Psychol. Reports, 1966, 19, 311-324)

78 J. L. Phillips and R. C. Atkinson. The effects of display size on short-term memory. August 31, 1965.

79 R. C. Atkinson and R. M. Shiffrin. Mathematical models for memory and learning. September 20, 1965.

80 P. Suppes. The psychological foundations of mathematics. October 25, 1965. (Colloques Internationaux du Centre National de la Recherche Scientifique. Editions du Centre National de la Recherche Scientifique. Paris: 1967. Pp. 213-242)

81 P. Suppes. Computer-assisted instruction in the schools: potentialities, problems, prospects. October 29, 1965.

82 R. A. Kinchla, J. Townsend, J. Yellott, Jr., and R. C. Atkinson. Influence of correlated visual cues on auditory signal detection. November 2, 1965. (Perception and Psychophysics, 1966, 1, 67-73)

83 P. Suppes, M. Jerman, and G. Groen. Arithmetic drills and review on a computer-based teletype. November 5, 1965. (Arithmetic Teacher, April 1966, 303-309.

84 P. Suppes and L. Hyman. Concept learning with non-verbal geometrical stimuli. November 15, 1968.

85 P. Holland. A variation on the minimum chi-square test. (J. math. Psychol., 1967, 3, 377-413).

86 P. Suppes. Accelerated program in elementary-school mathematics -- the second year. November 22, 1965. (Psychology in the Schools, 1966, 3, 294-307)

87 P. Lorenzen and F. Binford. Logic as a dialogical game. November 29, 1965.

88 L. Keller, W. J. Thomson, J. R. Tweedy, and R. C. Atkinson. The effects of reinforcement interval on the acquisition of paired-associate responses. December 10, 1965. (J. exp. Psychol., 1967, 73, 268-277)

89 J. I. Yellott, Jr. Some effects on noncontingent success in human probability learning. December 15, 1965.

90 P. Suppes and G. Groen. Some counting models for first-grade performance data on simple addition facts. January 14, 1966. (In J. M. Scandura (Ed.), Research in Mathematics Education. Washington, D. C.: NCTM, 1967. Pp. 35-43.

91 P. Suppes. Information processing and choice behavior. January 31, 1966.

92 G. Groen and R. C. Atkinson. Models for optimizing the learning process. February 11, 1966. (Psychol. Bulletin, 1966, 66, 309-320)

93 R. C. Atkinson and D. Hansen. Computer-assisted instruction in initial reading: Stanford project. March 17, 1966. (Reading Research Quarterly, 1966, 2, 5-25)

94 P. Suppes. Probabilistic inference and the concept of total evidence. March 23, 1966. (In J. Hintikka and P. Suppes (Eds.), Aspects of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1966. Pp. 49-65.

95 P. Suppes. The axiomatic method in high-school mathematics. April 12, 1966. (The Role of Axiomatics and Problem Solving in Mathematics. The Conference Board of the Mathematical Sciences, Washington, D. C. Ginn and Co., 1966. Pp. 69-76.

(Continued on inside back cover)

FACTS AND FANTASIES OF EDUCATION

by

Patrick Suppes

TECHNICAL REPORT NO. 193

October 18, 1972

PSYCHOLOGY AND EDUCATION SERIES

This paper is the author's address to the American Educational Research Association as recipient of the 1971 Meritorious Researcher Award conferred jointly by Phi Delta Kappa and the American Educational Research Association. It will be published separately by Phi Delta Kappa.

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Facts and Fantasies of Education

Patrick Suppes

Stanford University

1. Introduction

The text for my sermon today is the closing paragraph of Hume's *Enquiries Concerning Human Understanding.*

> When we run over libraries, persuaded of these principles (Hume's principles of the understanding), what havoc must we make? If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.

Hume would be the first to admit that we are all entitled to our fantasies, but he would also insist that we recognize them as the fantasies they are. To reformulate his text, general ideas about educational policy and practice contain little but sophistry and illusion, unless they can be defended by abstract reasoning from some other accepted general principles or be inferred in a definite manner from particular matters of fact.

This sounds like a hard line, and it is. It is too hard to be used at all times and places in discussions of educational policy and practice, but it is not too hard for a reflective occasion like this one, which is aimed at appraising the relevance of research to educational ideas.

Rather than begin with any general remarks or general propositions, I shall first report some of the results of perusing my own library and applying Hume's tests. The initial examples that I critically examine will be those that most of us can sagely and benignly agree about. I shall move on to others that are more controversial, and for that reason, probably more important.

Rewording Hume's text still once again, the thesis of this paper may be expressed succinctly in the following way: Without proper evidence, alleged facts on which educational policy or practice is based can only be classed as fantasies. It is the task of research to convert the 'right' fantasies into facts and to show the others to be the unsubstantial fantasies they are.

Before turning to particular examples, I want to say a preliminary word about evidence. I have a somewhat skeptical and Bayesian attitude. I do not think it is possible to state in mathematically precise terms what is to count as evidence and what is not. Evidence also need not be collected by systematic experimentation. The most glorious quantitative science of them all, namely, astronomy, has scarcely ever been able to include experiments. Evidence is also not just a matter of quantitative data, organized in obvious quantitative fashion. We do not need to perform an experiment or take systematic observations in order to hold the firm factual belief that the sun will rise tomorrow. On the other hand, when we turn to the formulation of general principles or general ideas about human conduct and how that conduct might be changed by the process of education, we must forever be wary and skeptical of those who promise much in general terms and give us principles unsupported by evidence.

I fully recognize also that over large periods of time most people are indeed taken in by some unsupported principles. One of the most sobering facts I know is that one of the earliest English charities was organized to collect money to buy wood to burn witches. The record of human folly committed in the name of morality or truth is too long and dismal to survey here.

At this point, I turn to some examples that give a more vivid sense of the continued need and the vital place of research in education. I consider initially what I call first-order fantasies. These are fantasies about general ways of organizing education in matters of theory, policy or practice. Afterward I turn to second-order fantasies, which are fantasies about methodology or about how we should determine the truth or falsity of first-order fantasies. Remember that in the sense I am using the terms here, fantasies of either the first or second order can be good fantasies in the sense that they can be true. It is the task of research to produce the evidence that will convert them from fantasies to fact.

2. First-order Fantasies

I begin with a classic example of applied linguistics.

Linguistic fantasies about reading. The importance and significance of the work of the American linguist Leonard Bloomfield is widely recognized and not challenged by me. The very quality of Bloomfield as a linguist makes all the more striking the simplicity of his ideas and his apparent total unawareness of the need for data in recommending how reading should be taught. His ideas are set forth in the book *Let's Read*, A Linguistic Approach (1961), jointly authored with Clarence L. Barnhart and published some years after Bloomfield's death.

So as not to enter into too elaborate an analysis, I shall restrict myself to Bloomfield's recommendations about the first materials of reading. He enunciates in a few paragraphs (pp. 39-42) the following.

1. "Our first material must show each letter in only one phonetic value."

2. "Our first material should contain no words with silent letters (such as knit or gnat) and none with double letters, and none with combinations of letters having a special value (as th in then or be in bean)."

3. "The letter x cannot be used, because it represents two phonemes (ks or gz)."

4. "The letter q cannot be used, because it occurs only in connection with an unusual value of the letter u."

5. "The best selection of values of letters to be used in the first materials for reading is the following," and here follows a large set of recommendations.

6. "Our first reading material will consist of two-letter and three-letter words in which the letters have the sound values given in the above list."

7. "We should not, at this stage, ask the child to write or print the words: that comes much later."

8. "The early reading lessons should not be very long, for they demand a severe intellectual effort. It may be well to take up only two words in the first lesson."

9. "In the second lesson, after review, add two or three more words of the same group."

10. "The drill should continue until the child can read correctly any one of the words when the parent or teacher points to it."

11. "If the child has learned the pattern in the list of actual words, he should be able to read nonsense syllables using the same pattern. . . . The nonsense syllables are a test of the child's mastery of the phoneme. Tell the child the nonsense syllables are parts of real words which he will find in the books that he reads."

12. "The acquisition of nonsense syllables is an important part of the task of mastering the reading process. The child will learn the patterns of the language more rapidly if you use the nonsense syllables in teaching."

A number of additional principles (pp. 19-42) are stated, but I have given a large enough sample to indicate in an explicit way why I label these principles <u>Bloomfield's fantasies</u>. They represent one of the purest examples of an analysis of one kind being extrapolated and applied to a different kind of problem without recognition of the need for data and for evaluation of the correctness of principles in their new application. Bloomfield applies some fairly obvious phonetic principles and generalizations, but shows no recognition at any point of the need for data to check on the correctness of these principles as principles of reading.

As I use <u>fantasy</u> in this paper, a fantasy can be correct or true, but it remains a fantasy unless proper evidence is offered, and this is certainly the case for what Bloomfield offers us for principles about reading. It is almost breathtaking to have him assert, for example, the principle that children learn language more rapidly if nonsense syllables are used in teaching. To fail to recognize the need for data and to state such a learning principle without any systematic concept of learning seems, in our currently skeptical time, almost incredible.

Piaget's stages. The influence of Jean Piaget on developmental psychology is recognized throughout the world. His very status, however, as an imaginative creator of new concepts and theories about children's behavior and development has led to an often uncritical acceptance of his ideas. I select for emphasis in the present discussion his concept of stages of development, which has played such a central part in many of his works and has also been taken over into developmental psycholinguistics. At first glance one might think that the concept of stages of development is a methodologically innocent one and scarcely a subject of controversy. A perusal of Piaget's own writings and the large derivative corpus soon leads one to another conclusion. As an example I shall discuss the analysis of three stages of multiple seriation in The Early Growth of Logic in the Child by Bärbel Inhelder and Piaget (1964).

To indicate the unequivocal adoption of the idea of stages, I quote from page 270:

> We shall distinguish three stages, corresponding to the usual three levels. During stage I, there are no seriations in the strict sense. The child's constructions are intermediate between classification and seriation. . . . During stage II, there is seriation, but only according to one of the criteria, or else the child switches from one criterion to the other. . . . Finally, during stage III (starting at

7-8 years), the child reaches a multiplicative arrangement based on the twofold seriation of the set of elements.

There is in this discussion, as elsewhere in Piaget, no suggestion that matters could be otherwise--that development could be incremental and continuous and that no stages could be identified in nonconventional fashion. To adopt the idea of stages as a convenient, conventional way of talking in certain restricted contexts is, of course, quite natural.[1] It is another thing to talk as if they were real abstractions with a verifiable and unequivocal empirical content.

It is a truism that children develop new capacities and new skills as they get older. The problem in evaluating the existence of stages is not one of affirming this truism, but rather one of differentiating the concept of stages from the equally natural concept of continuous development.

A second related problem that needs detailed study is the extent to which the mastery of different concepts follows the same order in different children. Again, it is important that the experimental design be as meticulous as possible in order not to prejudge the issue. It can scarcely be said that Piaget's design in the study I am considering satisfies this criterion, and very few others of like nature do either.

Moreover, if nonconventional plateaus were discovered in the behavior of individual children, we would also want to know whether these plateaus run across the same set of concepts or occur in a manner that is much more randomly related to the concepts themselves and that might reflect quite different sorts of processes of maturation in the child. Again, little evidence is to be found on this point.

My first draft of these remarks on the concept of stages received an excellent critique from Harry Beilin, and he has provided me references to his own work and that of Piaget where the question of the actual existence of stages is discussed. Piaget writes:

> I now come to the big problem: the problem of the very existence of stages; do there exist steps in development or is complete continuity observed? . . . when we are faced macroscopically with a certain discontinuity we never know whether there do not exist small transformations which would be continuous but which we do not manage to measure on our scale of approximation. In other words, continuity would depend fundamentally on a question of scale; for a certain scale of measurement we obtain discontinuity when with a finer scale we should get continuity. Of course this argument is quite valid, because the very manner of defining continuity and discontinuity implies that these ideas remain fundamentally relative to the scale of measurement or observation. This, then, is the alternative which confronts us: either a basic continuity or else development by steps, which would allow us to speak of stages at least to our scale of approximation [1960, p. 121].

A more detailed and careful analytical discussion of the concept of stage is to be found in Beilin (1971), and he raises a number of issues that are not pertinent to the main point I wish to make here. In the printed discussion following Beilin's paper, there are remarks by Piaget that seem to me incorrect, and in general, incorrect for the same reasons that the above quotation about the existence of stages is incorrect. In the context of Beilin's paper Piaget (1971, p. 194) is discussing the problem of novelty, and he has this to say:

> Thus, to my regret, I did not find in Beilin's paper any reference to this problem of novelty--of the formation of novelties in general and the conditions necessary for the development of new structures. . . . If there are no novelties, then the concept of stages is artificial. There lies the whole problem.

Before commenting directly on these remarks of Piaget's, I want to compare the almost total absence of serious critical analysis of the concept of stages with the controversy in learning theory that existed about ten years ago between all-or-none and incremental learning. In the latter case the battle was joined with intensity and fruitful result because there were strong protagonists on both sides of the argument, and each was determined to establish the incorrectness of the other's view, and if possible, the correctness of his own view. A large number of experiments were performed, and quite detailed analyses were made of the data to test whether or not the learning of simple concepts or simple associations satisfied all-or-none or incremental properties. Analyses with special relevance to concept formation in children are to be found in Suppes (1965) and Suppes and Ginsburg (1963). For the purpose of this discussion, the all-or-none model would correspond to Piaget's idea of stages, and the tests of incremental learning models to the continuous learning that is the natural alternative to stages. We need the kind of sharp exchange and critical examination of experiments and concepts characteristic of that learning-theoretic controversy of a decade ago in the analysis of the concept of stages in developmental psychology.

It should be apparent that the attitudes expressed in the quotations from Piaget do not provide the basis for this kind of sharp exchange and critical examination. Piaget raises what is essentially an irrelevant question of scale. The problem is to find out for the given scale at which experimentation is conducted whether the process is all-or-none or incremental, and not whether there are microscales, for example, at which the process is continuous even if the data indicate all-or-none learning at the ordinary scale of experimentation. The second remark about novelty also seems to me to miss the point, especially as reflected in the extensive work on all-or-none learning in concept identification as opposed to concept formation experiments. It certainly is perfectly possible that learning is all-or-none or in terms of stages even when no questions of novelty are involved. There is also no reason to think that when concept formation and mastery of novel concepts are evident that learning is necessarily to be characterized in terms of stages rather than incrementally. I say once again that the element missing in this discussion by Piaget, and even by Beilin in his otherwise excellent article, is the concept of precise and detailed experimentation with quantitative analysis of data to test for the existence of stages.

Finding out the true state of affairs about stages is important not simply for theoretical purposes in developmental psychology, but because continual use of these concepts is found in the talk of educators in their organization of curriculum for young children, in their discussion of the skills of young children and in other related ways. It would be easy to document the continual casual reference to Piaget in a variety of educational publications in which teachers are once again being taught dogma without data that developmental stages are the way to think about the development of children.

I cannot resist one passing remark on this matter of stages and concepts like that of groupement and seriation. The very language used by Piaget and his more ardent followers is itself a kind of fantasy of mathematics. Those of good faith can believe that back of such talk is a real body of concepts that can be put into reasonable order. Those of us who are more skeptical face the beginning of the fantasies here and wait for a new round of theory and experiments to clear the air.

A comparative point of intellectual history is perhaps in order. There is much about Piaget's ideas that resembles the kind of suggestive web woven by Descartes in his principles of philosophy. Descartes, of course, was dealing with the physical world and Piaget with the psychological world of child development. Both operate in large theoretical terms and with little regard for detailed experimental investigation. Descartes' tale proved to be enormously seductive in the seventeenth century, and even Newton found it difficult to throw off the Cartesian ideas. Leibniz, however, put it correctly when he characterized Descartes' physics as a _roman de physique_, and I shall be bold enough to say that we may very likely in the future characterize Piaget's work as a _roman de psychologie_. To say this is not to deny that Descartes has occupied an important place in the history of physics or that Piaget has occupied an important place in the history of psychology. It is rather to put in proper perspective large-scale theories that are as close to fantasies as they are to facts.

Skinner on arithmetic. As an example of a different sort, but at the same first-order level of fantasy, I next would like to consider what Skinner has to say about teaching arithmetic in his book The Technology of Teaching (1968, pp. 14-15). Here is the opening passage on arithmetic.

> From this exciting prospect of an advancing science of learning, it is a great shock to turn to that branch of technology which is most directly concerned with the learning process--education. Let us consider, for example, the teaching of arithmetic in the lower grades. The school is concerned with imparting to the child a large number of responses of a special sort. The responses are all verbal. They consist of speaking and writing certain words, figures, and signs which, to put it roughly, refer to numbers and to arithmetic operations. The first task is to shape these responses--to get the child to pronounce and to write responses correctly, but the principal task is to bring this behavior under many sorts of stimulus control. This is what happens when the child learns to count, to recite tables, to count while ticking off the items in an assemblage of objects, to respond to spoken or written numbers by saying "odd," "even," or "prime." Over and above this elaborate repertoire of numerical behavior, most of which is often dismissed as the product of rote learning, the teaching of arithmetic looks forward to those complex serial arrangements of responses involved in original mathematical thinking.

> The child must acquire responses of transposing, clearing fractions, and so on, which modify the order or pattern of the original material so that the response called a solution is eventually made possible.

The crudeness of this talk about responses and shaping them without serious reference to how arithmetical concepts should be built up is typical of this strange and undocumented proposal of how arithmetic ought to be taught. The naive and impressionistic character of the remarks is attested to by the juxtaposition of the words 'odd', 'even', and 'prime' in the middle of the passage. The very special role, for example, of students' ever responding prime to spoken or written numbers is to be emphasized. The casual way of talking about moving from arithmetic to "those complex serial arrangements of responses involved in original mathematical thinking" is a reflection of how vague and unsubstantial Skinner's ideas about the teaching of arithmetic or other parts of mathematics are. It would be interesting indeed to have those complex serial arrangements of responses made to match any serious piece of mathematical instruction, let alone original mathematical thinking. The casual talk about acquiring "responses of transposing, clearing fractions, and so on" is again indicative of the unthought-out and undocumented character of the remarks.

No evidence is offered about the effectiveness of these ideas for the teaching of arithmetic. What is more important--it would not even be clear from this passage or the passages that follow how any teacher would begin to arrange the complex material of arithmetic in proper order for learning by children. It would be interesting to see what

Skinner would have to say about the detailed sequence of materials in arithmetic, and how the proper arrangement of materials should be made, according to which principles and on the basis of what data. It is especially ironic to have such a broad and unsubstantiated sketch of how arithmetic should be taught without reference to any of the extensive literature on the learning of arithmetic.

I cannot think of a better challenge to Skinnerians than to produce a genuine psychological theory of mathematical learning and thinking. So far as I know, there is not yet a serious contribution from either Skinner or his followers on this important educational topic. In some quarters at least, I am sure the fantasy will remain that somehow operant conditioning is the key to successful mathematical learning.

3. Second-order Fantasies

By a second-order fantasy I mean a belief about the efficacy or lack of it of a certain methodology, which is unsupported by evidence or systematic argument. I begin with an example much closer to home than any I have yet considered and refer to the writings of two authors with whom I am in general intellectually sympathetic.

Campbell and Stanley on experimentation. The fantasy I have in mind is the unsupported and yet wholly enthusiastic support of experimentation by Campbell and Stanley in their well-known chapter on this subject in the Handbook of Research on Teaching (1963). As most of you will remember, the handbook was itself a product of this association. Let me begin with two quotations from the second and third pages of the chapter.

> This chapter is committed to the experiment: as the only means for settling disputes regarding educational practice, as the only way of verifying educational improvements, and as the only way of establishing a cumulative tradition in which improvements can be introduced without the danger of a faddish discard of old wisdom in favor of inferior novelties. . . . even though we recognize experimentation as the basic language of proof, as the only decision court for disagreement between rival theories, we should not expect that "crucial experiments" which pit opposing theories will be likely to have clear-cut outcomes. When one finds, for example, that competent observers advocate strongly divergent points of view, it seems likely on a priori grounds that both have observed something valid about the natural situation, and that both represent a part of the truth. The stronger the controversy, the more likely this is. Thus we might expect in such cases an experimental outcome with mixed results, or with the balance of truth varying subtly from experiment to experiment.

As matters of personal belief, I accept with certain reservations what Campbell and Stanley have to say in the quoted passages. What I am criticizing is the lack of argument for the position, and for this reason I have labeled the passage an example of a second-order fantasy. The chapter contains no systematic examination of alternatives to experimentation, no review of sciences like astronomy which do not engage in experimentation in any serious way and yet achieve remarkable results, no attempt to formulate general principles to make it clear why experimentation is so important; in fact, there is no scientifically serious attempt to define the concept of an experiment.

I emphasize that I do not have in mind a rigorous formal treatment of the concept of an experiment, but rather a densely argued informal consideration of the principles of evidence that offer a systematic defense of the use of experimental procedures. For example, within a Bayesian framework (I do not mean to suggest that they necessarily should adopt such a framework), one can argue that the likelihoods, as opposed to the prior distributions of opinion or belief, can be agreed upon by different investigators of different theoretical persuasions. It is the practical possibility of agreement on likelihood functions that makes experimentation attractive. We can of course go on to ask the deeper question, why is it that different individuals of quite different orientations can agree on likelihood functions and the conceptual scheme of experimentation when they are far apart theoretically? It is not always true that they can so agree, but it is true often enough that an analysis can be given of the reasons for agreement in a wide range of circumstances.

My own view would be that the defense should be built on the basis of the different status of different kinds of knowledge. We can, for instance, agree on how a given group of students answered the items on a test if the test was multiple choice, but we may not be able to agree on how to interpret the results, or if we gave an essay test how to evaluate even narrowly the essay responses. It seems to me that the defense of experimentation depends heavily on the drawing of such distinctions between the kinds of knowledge we have.

The second major aspect of classical psychological and educational experimentation centers around the difficult and elusive concept of randomization. Here too, it seems to me that Campbell and Stanley do not give the research worker in search of help a detailed and closely argued defense of the reasons for randomizing in experiments. If the authors felt that the subtle topic of randomization was too difficult a one to enter into, clear warnings should have been given the readers that they were not attempting any defense of the concept and that it was being taken on faith as a wonderful thing.

Later in the chapter there is a section entitled "Some preliminary comments on the theory of experimentation," and once again wise remarks are made about statistical lore and experimental practice. What is missing, however, is that sense of intellectual openness on the one hand and precision of argument on the other so very much required in the theory of experimentation, or more generally, in applied statistics. Applying Hume's dichotomy of having either reasoning about abstract matters or evidence about matters of fact, we find that both the quotation above and the longer section on the theory of experimentation are neither organized around abstract principles from which more principles of experimentation are derived, nor validated by a systematic collection of empirical evidence bearing on the theory of experimentation.

It is appropriate to add weight to these general statements by some more detailed examples. There are at least three respects in which I think the innocent reader might be misled by Campbell and Stanley's generally excellent article. To begin with, the deeper and more varied the contact a person has with applied statistics, the more evident it is that some experience in seeing the statistical procedures and tests of significance derived from first principles is of importance. It is too easy for the innocent researcher to divorce in his mind the simple algebraic formulation of particular tests or procedures from the probabilistic background that justifies their derivation and interpretation. I am not suggesting something that I think is easy to do within the restrictions Campbell and Stanley set for themselves; however, some sense of derivation from first principles in at least one example would deepen considerably the basis the reader would have for accepting the kinds of distinctions introduced.

My second remark is a more serious and important one. Already at the beginning of the nineteenth century, in his treatise on the theory of probability, Laplace (1820) emphasized the importance of not simply establishing the existence of an effect, but establishing a method for estimating the magnitude of an effect. From a broad methodological standpoint, perhaps the single most important criticism one might make of the statistical procedures used and exemplified in the Journal of Experimental Psychology, methods of which Campbell and Stanley in general approve, is the overwhelming use of tests of significance establishing the existence of effects, in contrast to the almost total absence of tests that estimate magnitudes of effects.

A simple, but powerful, analysis of such an example is provided by Laplace's attempt to estimate the benefits of inoculation for smallpox by variolar virus before vaccine was available. (Laplace concluded that the mean increased longevity from inoculation was about three years, provided that there was no food shortage or other violent disruption of the environment.) Such estimates of the magnitude of causal effects are of the first importance in both pure and applied science, and it is especially important to bring them more to the fore in educational research. We may leave to the psychologists aloft in the pages in the *Journal of Experimental Psychology* the design of experiments that test for *existence* of effects. In education we are much more concerned with estimating *magnitudes* of effect. If, for example, a new curriculum that costs twice as much as an old curriculum produces a measurable effect, but that measurable effect is very small in magnitude, then the practical use of this curriculum is questionable.

Mentioning the problem of estimating magnitudes of effects suggests immediately broadening the framework of statistical analysis to that of statistical decision theory. For many educational experiments, a three-fold decision procedure: accept the new procedure of instruction, reject it, or continue further experimentation where the current verdict of nothing yet proved would lead to a new look at experimental procedures, and especially their interpretation. But I shall not attempt to explore these matters further in the present context.

My third and final comment on the "interior analysis" of Campbell and Stanley's chapter concerns some remarks they make about linear models. In discussing tests of significance for time series designs, on page 43 they assert "Statistical tests would probably involve, in all but the most extended time series, linear fits to the data, both for convenience and because more exact fitting would exhaust the degrees of freedom, leaving no opportunity to test the hypothesis of change."

It seems to me that here is an example of simplifying too drasticall, and therefore introducing a small-scale kind of fantasy too easily adopted by educational researchers. It is a fantasy that we must always test for linear relations, because we have no ability to handle nonlinear ones. Especially with the use of modern computers, it is almost as easy to deal with simple nonlinear models as linear ones. Exploring the alternatives to linearity provides excellent insight into the nature of the relations between the variables and does not require necessarily the use of more degrees of freedom. Let us consider, for example, just the simple case of two variables, with x the independent variable and y the dependent variable. We may express the linear model by the following equation:

$$y = a + bx .$$

This model has two parameters to be estimated from the data and thus two degrees of freedom are lost. If we think of the effects of increase in x on y proceeding at a faster than linear fashion, we can estimate the same number of parameters for the quadratic model:

$$y = a + bx^2 .$$

On the other hand, if we think of the nonlinear increase in y with increases in x as less than linear, we can easily test the logarithmic model:

$$y = a + b \log x .$$

There is much more to be said about these matters, and I am not pretending to give a detailed analysis to complement these brief remarks. It is just that in my search for fantasies I have tried to look everywhere, even among some of the best established and generally most sensible sources.

Chomsky's theory of competence. As a second example of a second-order fantasy, I select Chomsky's theory of competence. If the ideas that he seems to be putting forth were correct, they would have some fairly far-reaching implications for educational research and educational practice. I classify his remarks quoted below as second order, because they recommend an approach to the study of behavior that is at considerable variance with current emphases. The following passage (Language and Mind, 1972, pp. 72-73) states Chomsky's methodological point in succinct form.

> The theory of generative grammar, both particular and universal, points to a conceptual lacuna in psychological theory that I believe is worth mentioning. Psychology conceived as "behavioral science" has been concerned with behavior and acquisition or control of behavior. It has no concept corresponding to "competence," in the sense in which competence is characterized by a generative grammar. The theory of learning has limited itself to a narrow and surely inadequate concept of what is learned--namely a system of stimulus-response connections, a network of associations, a repertoire of behavioral items, a habit hierarchy, or a system of dispositions to respond in a particular way under specifiable stimulus conditions. Insofar as behavioral psychology has been applied to education or therapy, it has correspondingly limited itself to this concept of "what is learned." But a generative grammar cannot be characterized in these terms. What is necessary, in addition to the concept of behavior and learning, is a concept of what is learned--a notion of competence--that lies beyond the

conceptual limits of behaviorist psychological theory. Like much of modern linguistics and modern philosophy of language, behaviorist psychology has quite consciously accepted methodological restrictions that do not permit the study of systems of the necessary complexity and abstractness. One important future contribution of the study of language to general psychology may be to focus attention on this conceptual gap and to demonstrate how it may be filled by the elaboration of a system of underlying competence in one domain of human intelligence.

As in the case of Skinner, the thesis set forth by Chomsky is breathtaking in its dogmatic simplicity. It could be said that it seems dogmatically simple only because I am quoting the introduction of a long and complex empirical or theoretical argument. Substantial formal arguments and substantial empirical data are offered subsequently, and I have distorted the analysis by restricting myself to the quotation just given. Although in the pages that follow, Chomsky amplifies the views about competence set forth in this paragraph, he does not amplify them in a way that satisfies the Humean standards stated at the beginning of this lecture. Because a number of psychologists who have influence in education have been much impressed by Chomsky's notion of competence, it will be useful to examine what he has said and the concept itself in somewhat more detail. It is the most elegant of the fantasies I have evoked and therefore the appropriate one for final consideration.

Let me begin with a key sentence of Chomsky's remarks that is characteristic of conceptual fantasies. After describing the nature of behavioral psychology, he says, "but a generative grammar cannot be characterized in these terms." He goes on to say that behavioral psychology has accepted methodological restrictions that do not permit the study of systems of the appropriate complexity. The fantasy consists in this negative claim that a generative grammar cannot be characterized within the framework of behavioral psychology. I have on another occasion (Suppes, 1968) criticized a similar claim by Bever, Fodor and Garrett (1968), who attempted to offer what they consider a formal proof of the limitations of associationism as a basis for language learning. In criticizing their work I characterized it as an example of negative dogma as contrasted to negative proof.

The fantasy claim is especially appropriate in matters of this kind, because of the long and classical tradition in mathematics of converting negative dogma into negative arguments and establishing thereby a subject of much intellectual richness. To transpose the situation slightly, I can imagine without difficulty the sardonic grin with which a mathematician at Alexandria in, let us say, 100 B.C. would have greeted the unsupported claim that it is obvious that the trisection of an angle cannot be characterized in terms of operations performable by a straightedge and compass alone. A two-word response would have been sufficient: Prove it. The austerity and precision of negative mathematical arguments are too restraining and perhaps puritanical in their methods for Chomsky and his ardent followers.

To give a negative proof, we must first have a much clearer idea of what is meant by the theory of competence than the characterizations given by Chomsky or his cohorts. If we are talking about language, for example, it is strange and wonderful to find only grammar and not semantics mentioned in the discussion of competence. By example we are told that generative grammars provide a model for theories of competence, but what is the model of semantic competence? On the one hand, we are urged not to consider arbitrary grammars and permit thereby the generation of any recursively enumerable set; rather, we should pick grammars with appropriate restrictions. On the other hand, we are told that it is no part of a theory of competence to build in a model of human memory and perception and to deal with it in terms of competence ideas. Reflection on the passages cited and similar writings by linguists in the Chomsky tradition does not give one confidence that a serious intellectual body of ideas is being developed under the heading of the theory of competence.

As my final remark on this, let us even assume that there is such a body of serious ideas to be developed. While there are certain mathematical areas in which one can conceive of formulating what would seem to be a theory of competence, one is struck by how irrelevant it is to any educational or psychological problems.

The mathematical example I want to deal with is that of mathematical proofs. In principle, it is quite straightforward to give a simpleminded theory of competence for mathematical proofs; namely, we know that we can formulate within first-order logic almost all current mathematical ideas, and we can then enumerate the theorems of the subject by enumerating the proofs. The enumeration of the proofs will constitute a kind of theory of competence. Any proof that exists will eventually turn up in the list after only a finite number of predecessors. We have thereby a simple algorithm for the production of any proof, and we can show that abstractly, simply as an algorithm, we can do no better than this.

No one thinks that this formal theory of competence has anything serious to do with the psychology of students' discovering elementary mathematical proofs in elementary mathematical courses or in mathematicians at work in unknown territory discovering new and complex proofs. On the one hand, we give a clear and simple theory of competence, one that we can state much more about in a sharp mathematical fashion than we can in the case of the relation of generative grammars to language; yet on the other hand, we can all recognize at once the essential irrelevance of this theory of competence to the psychological problems of understanding how someone finds a proof or to the educational problem of providing instruction to students in giving proofs.

It seems to me that there is some reason to conjecture that the relation of this theory of competence for proofs that I have given may bear as close a relation to the proper performance theory of proofs as does current work on generative grammars, especially with a complex transformational component, to correct performance models of language usage. In any case, we certainly need something much more definite and intellectually precise than Chomsky's historical ruminations on the decline and fall of rationalism and its new resurrection under a linguistic flag.

With some regret, I terminate my remarks on fantasies about competence at this point and return to my general theme.

4. Research and the Belief Structures of Education

I could easily have seized upon a host of lesser targets to provide further case material. Because education is of such universal concern in our society, everyone feels free and often competent to speak about it in general terms. The body of literature full of unsubstantiated general ideas and principles is now overwhelming. Its authors run from the new romantics like Friedenberg and Holt to a bevy of journalists turned sometime scholars. Characteristic of this literature is the lack of intellectual discipline, either in terms of rigorous analysis of general principles or in the presentation of detailed factual evidence to support the principles stated. Unfortunately, this kind of literature represents nothing new in education. The history of educational change is awash with firm prejudices and soggy arguments. I am not, however, an advocate of pessimism or skepticism. I think that it is possible to improve education, and that research can make an increasingly important contribution to this improvement. Let me try to sketch some of the ways I think this can happen.

First of all, it is important to recognize that the belief structure of education, the basis on which decisions about policy and practice are taken, represents an accretion of many years of experience and fantasy. Many of the beliefs are interwoven with other strongly held beliefs about how individual, family and societal life should be organized. If nothing else, the data of the Coleman report have shown us how difficult it is to isolate any particular effects of education from the broad spectrum of family and cultural influences. A central problem of research is to attack that belief structure where it is unsupported by data or systematic is due to the closeness of the subject matter to the layman's own experience. The man in the street does not expect to be able to give a serious opinion about how one should build a better television set or a nuclear fuel plant that will reduce pollution. He recognizes, of course, that both of these things are worth having, but he seldom has opinions about how they can be accomplished. In contrast, ask the mother in the street who has a first grader about reading and you are likely to hear some definite views on the teaching of reading.

The nearness of the subject matter is one aspect of the problem, but the other aspect, I think, is a problem about the research itself. Even a casual scientific inspection of the process the child goes through in learning to read quickly demonstrates its complexity. The perceptual, cognitive, linguistic and motivational aspects of the process are each enormously complicated, and a detailed conception of how the visual perception of what the child sees is related to the spoken language he already knows is far from available. If we compare this situation to the task of improving television sets, the picture is rather dismal. The fundamental physics of the processes involved in projecting a televised image on a screen are well understood; many of the fundamental concepts go back to classical electromagnetic theory of the nineteenth century.

We have in psychology no comparable fundamental theory of perception, nor do we have a comparable theory of spoken language comprehension or production. At the present time, in solving problems of learning to read, we are more in the position of bridge builders before the theory of statics was developed than we are in the position of designers of is due to the closeness of the subject matter to the layman's own experience. The man in the street does not expect to be able to give a serious opinion about how one should build a better television set or a nuclear fuel plant that will reduce pollution. He recognizes, of course, that both of these things are worth having, but he seldom has opinions about how they can be accomplished. In contrast, ask the mother in the street who has a first grader about reading and you are likely to hear some definite views on the teaching of reading.

The nearness of the subject matter is one aspect of the problem, but the other aspect, I think, is a problem about the research itself. Even a casual scientific inspection of the process the child goes through in learning to read quickly demonstrates its complexity. The perceptual, cognitive, linguistic and motivational aspects of the process are each enormously complicated, and a detailed conception of how the visual perception of what the child sees is related to the spoken language he already knows is far from available. If we compare this situation to the task of improving television sets, the picture is rather dismal. The fundamental physics of the processes involved in projecting a televised image on a screen are well understood; many of the fundamental concepts go back to classical electromagnetic theory of the nineteenth century.

We have in psychology no comparable fundamental theory of perception, nor do we have a comparable theory of spoken language comprehension or production. At the present time, in solving problems of learning to read, we are more in the position of bridge builders before the theory of statics was developed than we are in the position of designers of television sets. It is my own view that no matter how beautiful the Latin squares of experimental design, purely empirical studies of different methods of teaching reading will not solve the problem of giving us the best possible methods, any more than a similarly purely empirical approach would ever have led us from the nonelectronic world of 1870 to the electronic marvels of the 1970s. By this remark I do not mean to denigrate the many good empirical studies that have been made of reading, but I do wish to put in perspective the severe limitations we face in practice in the absence of a deeper running theory of the processes involved.

To build such a theory is a good example of a major relevant problem for research in education. Like most research problems in education, the solution cannot depend upon the work solely of persons working in education, but rather it must draw upon scientific results from many disciplines, in this case ranging from neurophysiology through psychology to linguistics. What I consider important as a first step is the recognition that we do not have a fundamental theory of the reading process, and in all likelihood we shall not for some time to come. Let me be a little more explicit about what I mean by a <u>fundamental theory of the reading process</u>. I have in mind a theory that not only can predict errors or difficulties of students, but a theory that postulates structures rich enough to process information in the same sequence of steps a student does. Put another way, the models of the fundamental theory should be complete models of the student, and the sense of completeness I use can be given precision by using concepts from logic and computer science.

What I have said about reading applies to most other skills and subjects taught in our schools and colleges. I emphasize that I mean to sound a note of honesty, not of pessimism. Above all, I think the time has come to call for a much deeper theoretical orientation of research in education in order thereby to increase its relevance. In many areas, ranging from the teaching of reading to the teaching of civics, the greatest limitation on research is not the absence of hard-data studies, but the absence of serious and sophisticated theory. Of course, we cannot hope to build a mathematical and quantitative theory of educational processes over night. We can begin, however, to recognize clearly the absence of fundamental theory and to insist on the kind of intellectual discipline in the training of our graduate students that will give them the tools not merely to make well-designed experimental studies, but to construct well-put-together theories that have definite and precise assumptions and deductive consequences that bear on behavior and the way students learn.

In important ways a good beginning already exists. I would mention especially the statistical theory of tests, the theory of measurement, some parts of learning theory, and recent economic work on productivity in education. Most pressingly needed are mathematical and quantitative theories applicable to major areas of curriculum. In certain areas I see the possibility of rapid advance once a cadre of sufficiently sophisticated research workers is available. In elementary-mathematics education the well-defined structure of the subject and the long tradition of good empirical studies, as well as the modern theory of algorithms and abstract machine processes, make available a welter of concepts and intellectual tools for the development of a fundamental theory of mathematics learning and performance at the elementary-school level. To some extent, the same is true of second-language learning, although there is not the same tradition of fifty years of careful studies as there is in the case of elementary mathematics. Other areas that involve complex perceptual or cognitive processes are less amenable to any direct theoretical attack as yet, and it will undoubtedly be some time before even reasonable looking theories, let alone correct ones, are formulated.

5. A Research Example from Elementary Mathematics

I recognize, as does everyone else, that it is much easier to criticize than to produce definite constructive results in any area of scientific investigation. My original intention was to give as a final example of a fantasy some excerpts from my own past writings, because the sins of fantasy I have charged others with I have also committed myself in the past. Even worse, I forecast that I shall commit them again in the future.

After further reflection, I decided it would be more useful, and in a deeper sense, expose better my own biases and weaknesses, to sketch in a constructive fashion how a precise theoretical attack on problems of educational relevance can be made. The curriculum I consider is standard and elementary, namely, the algorithms children are taught for performing the basic arithmetical operations of addition, subtraction, multiplication and division. Also, I first consider performance data and only later say something about learning. Since the detailed theory of these matters is relatively technical, I have put the formal developments in the Appendix.

The psychological study of arithmetic skills, like most other parts of psychology, has a relatively recent history--only a few systematic studies were made before 1890. The real impetus was provided by E. L. Thorndike's analysis of the learning of arithmetic in his Educational Psychology (1913, 1914) and later in his The Psychology of Arithmetic (1922). In an attempt to account for the acquisition of arithmetic skills in terms of his three psychological laws--the law of readiness, the law of exercise and the law of effect--he tried to justify and analyze the reason for the traditional importance attached to drill and practice in arithmetic; for him the psychological purpose of drill is to strengthen the bonds between stimuli and appropriate responses. He moved on from such fundamental questions to the more practical ones of amount and distribution of practice. He emphasized the advantages of distributed practice and criticized the actual distribution of practice in textbooks of his time. Some effects of his work on the revisions of textbooks in the 1920s and later are documented in Cronbach and Suppes (1969, pp. 103-110).

In the twenties and thirties there were a large number of good empirical studies of arithmetic skills, many of which were concerned with detailed questions that had to be answered in any complete psychological theory of arithmetic. For example, Buckingham (1925) studied student preferences and aptitudes for adding up or down in column addition problems. An extensive review of this literature may be found in Suppes, Jerman and Brian (1968).

Empirical studies like those of Buckingham were not designed to develop an overall theory of arithmetic skills; nor, it is probably fair to say, was Thorndike completely sensitive to the gap that existed between his theoretical ideas and the actual algorithms students were taught to solve problems. There are many stages to work through in developing an adequate theory, and so far as I can see, there is no one point at which one can say the theory is now complete in all respects. If, for example, the theory is adequate at some conceptual level of information processing, then it is possible to move on to additional perceptual questions. Moreover, once a perceptual theory of a certain level of abstraction is successfully developed, it is possible to go on to still more detailed perceptual questions, such as requiring the theory to include eye movements of students as well as their numerical responses. It is for me an important methodological precept that at no foreseeable point shall we reach a fixed and firm bottom beyond which we cannot probe for further details and a more refined theory.

I would like to briefly sketch the history of some work of mine and my younger collaborators over the past six or seven years. Rather than attempt a general coverage, I have decided to select a singular example--the simple one of column addition--to illustrate how we have tried continually to deepen the theory, and then to discuss what I see as yet undone, but practically possible in the near future.

The data referred to are all taken from our work in computer-assisted instruction, but I shall not enter into any of the details. The kinds of models discussed can be applied to students using pencil and paper.

The first question we tried to answer was how can one predict the relative difficulty of different exercises of column addition? If, for example, we consider problems up to the size of three columns and three rows, we are confronted with approximately one billion problems. A meaningful theory must drastically reduce this large number of exercises to a small number of classes in which all members of a class are essentially the same in difficulty.

Our first approach (Suppes, Hyman and Jerman, 1967) was to identify a small number of structural features that would permit us to apply linear regression models to predict either probability of correct response or expected latency of response. Additional applications of such regression models may be found in Suppes, Jerman and Brian (1968) and Suppes and Morningstar (1972). The application of such regression models is exemplified in equation (3) of the Appendix. As can be seen from the information given there, the fit of the regression model to mean student-response data on column addition exercises is not bad. Conceptually, however, there are obvious lacunae. The regression model that predicts response probabilities does not really postulate a specific process by which students apply an algorithm to solve an exercise.

The next level of theory developed is aimed precisely at offering such process models, models that satisfy the information-processing requirements laid down for reading models in the earlier discussion. Without doubt, providing an adequate information-processing model for column addition is a much simpler affair than providing one for reading, and I have no illusions about the difference in complexity. The natural theoretical tools for providing process models of algorithmic tasks are automata, and for most of elementary arithmetic, simple finite automata are satisfactory. There is, however, one weakness in finite automata as ordinarily defined, namely, they have no place for a probabilistic theory of error, so the natural step is to move from finite deterministic automata to probabilistic automata.

An automaton becomes probabilistic by making the transition function from state to state probabilistic in character. Thus, from a given input and a given internal state there is a probability of going to any one of several different states. In general one wants to make the output function probabilistic also. This means that given an internal state and an input there is a probability distribution over the next output. (These ideas are made formally definite in Definitions 1 and 2 of the Appendix.) By drastically reducing the source of error to a small number of parameters, we can develop and apply manageable probabilistic automata to student-response data. A detailed example including maximum-likelihood estimates of the three parameters of the automaton are given in the Appendix.

Such a probabilistic automaton model takes a definite step beyond a regression model in providing in an abstract sense an adequate information-processing model. From a psychological standpoint, on the other hand, the automaton models described in the Appendix are unsatisfactory in that they lack any perceptual components, and therefore they do not deal directly with how the student actually processes the format of written symbols in front of him.

Our current work is very much directed at this point. In principle, it would be possible to continue the development of automaton models with an abstract concept of state to represent the student's perceptual processing. A weakness of this extension of the automaton models is that when the states are left in a general abstract formulation it is natural to end up designing a different automaton for each of the different tasks in elementary mathematics, and a plethora of models results. Closer examination of the algorithmic tasks of arithmetic facing the student in solving exercises indicates that the various tasks have much in common. This commonality suggests a somewhat different approach, an approach via register machines with perceptual instructions.

Register machines were first introduced by Shepherdson and Sturgis (1963) to give a natural representation of computable functions in terms that are closer to the idea of a computer accepting instructions than to a Turing machine. In the case of the representation of computable functions, a rather simple set of arithmetic instructions is sufficient. In particular, an unlimited register machine has a denumerable sequence of registers, but any given program only uses a finite number of these registers and the machine accepts six basic instructions: add one to a register, subtract one, clear a register, copy from one register to another, and two jump instructions, one conditional and one not. (This set of six instructions is not minimal, but it is convenient.) Obviously, for the perceptual processing that a student does we want a different register machine and a radically different set of instructions. In addition, it is natural to postulate only a finite fixed number of registers that the student can use.

The basic idea of this approach is to drastically simplify the perceptual situation by conceiving each exercise as being presented on a grid. The student is represented by a model that has instructions for attending to a given square on the grid; for example, in the standard algorithms of addition, subtraction and multiplication we begin in the upper right-hand corner and then have instructions to move downward through each column and from right to left across columns. Additional instructions for storing the results of an operation, for outputting the last digit of a stored numeral, etc., are needed. Some further details are given in the Appendix, but the discussion is not as complete as that for automaton models.

The basic idea of register machines is that the different algorithms are represented by subroutines. One subroutine may be called in another, as complex routines are built up. The procedure is familiar to most of us, even if the language I am using is not. For example, in performing column multiplication we use the algorithm of addition, which in this case means calling the subroutine for addition; in long division we call the subroutines for subtraction and multiplication, as well as for addition. Each basic subroutine is represented by a program in terms of the primitive instructions. The problem from a psychological standpoint is to find instructions that provide not only a realistic description of what the student does, a description that can be fitted to data in the same way that the automaton models have been applied to data, but also a fuller account of how the student processes the exercise.

At the first stage of analyzing register-machine models we can get results similar to those for the automaton models by postulating error parameters for execution of main subroutines of the routine for a given algorithm. More is said about this in the Appendix. However, the real purpose of the register machines, in addition to providing some explicit analysis of perceptual processing, is to provide a natural method for analyzing learning.

The approach we have adopted is this. At each given stage, the student has command of a certain set of subroutines or procedures. To master more complex exercises and concepts the student must expand these subroutines or imbed them in more complex ones. A plausible approach is that the student builds up these more complex routines by verbal instruction received from the teacher and by interpretation, especially perceptual interpretation, of examples. When the verbal instruction by the teacher, or say, a computer-assisted instruction program, is explicit, and the link to the necessary internal instructions is close, a surprisingly simple theory of learning within a classical framework can be given. For example, the kind of determinate reinforcement for obtaining finite automata from stimulus-response models, as developed theoretically in Suppes (1969), can without much modification provide the theory for the buildup of the appropriate subroutines.

I emphasize, however, that we are only beginning the detailed analysis of learning in this complex setting, and I am describing the conceptual situation. I shall have to wait until later to report on the actual empirical accuracy of the learning models we have developed. The empirical results obtained with automaton models of performance have been good enough to encourage us to push on as rapidly as possible to the deeper problems of learning.

In this section I have tried to sketch an example of how one can pursue a systematic theory of relevance to education. It should be apparent to everyone that the example I have chosen is exceedingly limited, and from many people's standpoint it is an almost trivial part of the curriculum. On the other hand, it should be equally apparent that the psychological theory of learning and performance in a subject matter as simple even as elementary arithmetic is not in itself simple. In fact, a detailed learning theory of elementary arithmetic is far more complex than the usual kind of theory psychologists consider. One of the problems we have to face in education is the too great willingness of psychologists and others to generalize from quite simple tasks to complex ones.

What I hope to have brought out in the present discussion, which is developed technically in the Appendix, is that the problems of subject-matter learning require conceptual developments in their own right that do not fall naturally out of general ideas of current psychological theories. Yet, with proper use of the variety of conceptual tools now available, it does seem possible to provide an increasingly adequate theory of learning for at least the basic skills, for instance, the basic skills of mathematics and language, that constitute a fair portion of school curriculum everywhere in the world.

6. Conclusion

I would like to conclude with a final remark about theory construction relevant to education. The times have probably never been so propitious for luring some of the ablest young minds into the problems of educational research. There exists already a body of methods and results of which we can be proud; but it is also clear, especially when we turn to the construction of systematic theories of learning or instruction, that we have as yet scarcely scratched the surface. While we are scratching that surface those of us in educational research must impose exacting standards not only on ourselves, but also on our neighbors, be they linguists or psychologists. We must demand of them, as well as of ourselves, the best possible effort in theory construction. We must above all reject the attitude that has in the past sometimes been prevalent that second-rate theories and second-rate efforts in the development of theory will suffice for education, and that we are lucky to get small crumbs from the occasional psychologist or linguist or economist who happens to become interested in education. We do not need ill-worked-out theories from other disciplines. We do not need fantasies of abstractions and platitudes unsupported by serious and rigorous development. What we need for relevance in education are theories of intellectual power and rigor, and we should not rest until we get them.

Appendix[2]

In this appendix I give some (but by no means all) of the technical details of our research in the psychology of arithmetic. The first three sections deal with performance models and the last section deals with a learning model. Each section attempts to dig a step deeper than its predecessor into the skills of arithmetic. For simplicity I have restricted the analysis in this appendix to the simple case of column addition, but the methods either already have been or in principle can be extended to essentially the entire domain of elementary-school mathematics (in addition to the references in the main text, see Groen & Parkman, 1972; Suppes & Groen, 1967). On the other hand, a good many additional developments will be needed to extend this work even to routine parts of the undergraduate college mathematics curriculum. (Some very empirical first steps at this college level are to be found in Goldberg & Suppes, 1972; Kane, 1972; Moloney, 1972.)

Linear regression models. As mentioned in the main text I begin with regression models that use as independent variables structural features of individual arithmetic exercises. I denote the j^{th} structural feature of exercise i in a given set of exercises by f_{ij}. The parameters estimated from the data are the values attached to each structural feature. (In previous publications we have referred to these structural features as factors, but this can lead to confusion with the concept of factor as used in factor analysis.) I denote the coefficient assigned to the j^{th} structural feature by α_j, and I emphasize that the structural features themselves, as opposed to their coefficients, are objectively identifiable by the experimenter in terms of the exercises themselves, independent of the response data.

Let p_i be the observed proportion of correct responses on exercise i for a given group of students. The natural linear regression in terms of the structural features f_{ij} and the coefficients α_j is simply

$$p_i = \sum_j \alpha_j f_{ij} + \alpha_0 .$$

Unfortunately, when the regression is put in this form, there is no guarantee that probability will be preserved as the structural features are combined to predict the observed proportion of correct responses. To guarantee conservation of probability, it is natural to make the following transformation and to define a new variable z_i.

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

and then to use as the regression model

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

The numerator of equation (1) contains $1 - p_i$ rather than p_i, so that the variable z_i increases monotonically rather than decreases monotonically with the magnitude of the structural features f_{ij}.

In Chapter 3 of Suppes and Morningstar (1972), the following structural features were defined for column-addition exercises.

The feature SUMR is the number of columns in the largest addend. For three-row exercises SUMR is defined as 1.5 times the number of columns, plus .5 if a column sum is 20 or more. For example,

$$\mathrm{SUMR}\left(\begin{array}{r} a \\ +\ b \\ \hline c \end{array}\right) = 1$$

$$\mathrm{SUMR}\left(\begin{array}{r} a \\ b \\ +\ c \\ \hline de \end{array}\right) = \begin{cases} 1.5 & \text{if } de < 20 \\ 2 & \text{if } de \geq 20 \end{cases}$$

$$\mathrm{SUMR}\ (\underline{ab} + c = de) = 2\ .$$

This structural feature reflects the number of columns of addition, with greater weight being given to columns in three-row exercises than in two-row exercises.

The second structural feature is CAR, which represents the number of times the sum of a column, including any numbers carried to it, exceeds nine. For example,

$$\mathrm{CAR}\left(\begin{array}{r} a \\ +\ b \\ \hline c \end{array}\right) = 0$$

$$\mathrm{CAR}\ (a + b = \underline{cd}) = 1$$

$$\mathrm{CAR}\left(\begin{array}{r} ab \\ +\ cd \\ \hline ef \end{array}\right) = \begin{cases} 0 & \text{if } b + d \leq 9 \\ 1 & \text{if } b + d > 9 \end{cases}$$

$$\mathrm{CAR}\left(\begin{array}{r} ab \\ cd \\ +\ ef \\ \hline ghi \end{array}\right) = \begin{cases} 1 & \text{if } b + d + f \leq 9,\ \ a + c + e > 9\ . \\ 2 & \text{if } b + d + f > 9,\ \ a + c + e \geq 9\ . \end{cases}$$

The third structural feature VF reflected the vertical format of the exercise. The vertical exercises with one-digit responses were given the value 0. Multicolumn exercises with multidigit exercises and one-column addition exercises with a response of 11 were given the value 1. One-column addition exercises with a multidigit response other than 11 were given the value 3. For example,

$$VF\left(\begin{array}{r} ab \\ -\ cd \\ \hline e \end{array}\right) = 0$$

$$VF\left(\begin{array}{r} abc \\ +\ def \\ \hline ghi \end{array}\right) = 1$$

$$VF\left(\begin{array}{r} a \\ +\ b \\ \hline cd \end{array}\right) = 3 \ .$$

This structural feature is meant to reflect the likelihood of the mistake of reversing the digits of the correct response, especially in a one-column addition exercise. In the computer-assisted instruction environment where students were responding at teletype terminals, responses to vertical exercises were typed from right to left, while responses to horizontal exercises were typed from left to right. Thus, it was possible for a student to have in mind the correct answer, but to err by typing the digits in the reverse order. It is fair to say that this structural feature is of more importance in working at a computer-based terminal than when using paper and pencil.

Table 1 shows a pretest on column addition given to third graders. The following regression equation was obtained for the mean response

----
Insert Table 1 about here
----

data of 63 students taking the test.

(3) $$p_i = .53\ SUMR_i + .93\ CAR_i + .31\ VF - 4.06\ .$$

The multiple R was .74 and R^2 was .54, which reflects a reasonable fit to the data. I shall not enter into further details of the regression model, but shall move on to the next level of analysis of these same response data. As should be obvious, I am not attempting anything

TABLE 1

Pretest Exercises in Column Addition

1) 17 + 2

2) 6 + 6 + 5

3) 14 + 15

4) 6 + 13

5) 363 + 214

6) 416 + 212

7) 12 + 31 + 10

8) 11 + 22 + 14

9) 27 + 4

10) 8 + 32

11) 639 + 212

12) 66 + 14

13) 378 + 125

14) 557 + 256

15) 5267 + 283

16) 46 + 75 + 23

17) 3986 + 4735

18) 27 + 46 + 88

19) 7657 + 1875

20) 69 + 36 + 48

like a systematic presentation of data, but only enough to give a sense of how some of the models do fit.

Three-state automaton model. The central weakness of the regression models is that they are not process models. They do not provide even a schematic analysis of the algorithmic steps the student uses to find an answer. Automaton models are process models and therefore their use represents a natural extension of the regression analysis. For the exercises in column addition we may restrict ourselves to finite automata, but as ordinarily defined they have no place for errors. However, this is easily introduced by moving from deterministic state transitions to probabilistic ones.

I begin with the definition of a finite deterministic automaton, and then generalize. These developments follow Suppes (1969).

Definition 1. A structure $\mathfrak{A} = \langle A, V_I, V_O, M, Q, s_0 \rangle$ is a finite (deterministic) automaton with output if and only if

(i) A is a finite, nonempty set,

(ii) V_I and V_O are finite nonempty sets (the input and output vocabularies, respectively),

(iii) M is a function from the Cartesian product $A \times V_I$ to A (M defines the transition table),

(iv) Q is a function from the Cartesian product $A \times V_I$ to V_O (Q is the output function),

(v) s_0 is in A (s_0 is the initial state).

As an example of a finite automaton with output, that is, a finite automaton in the sense of this definition, we may characterize an automaton that will perform two-row column addition.

$$A = \{0,1\} ,$$

$$V_I = \{(m,n) : 0 \leq m,\ n \leq 9\} ,$$

$$V_O = \{0,1,\ldots,9\} ,$$

$$M(k,(m,n)) = \begin{cases} 0 & \text{if } m + n + k \leq 9 , \\ 1 & \text{if } m + n + k > 9 , \quad \text{for } k = 0,1 , \end{cases}$$

$$Q(k,(m,n)) = (k + m + n) \bmod 10 ,$$

$$s_0 = 0 .$$

Thus the automaton operates by adding first the ones' column, storing as internal state 0 if there is no carry, 1 if there is a carry, outputting the sum of the ones' column modulus 10, and then moving on to the input of the two tens' column digits, etc. The initial internal state s_0 is 0, because at the beginning of the exercise there is no 'carry'.

Definition 2. A structure $\mathfrak{A} = \langle A, V_I, V_O, p, q, s_0\rangle$ is a (finite) probabilistic automaton if and only if

(i) A is a finite, nonempty set,

(ii) V_I and V_O are finite, nonempty sets,

(iii) p is a function on $A \times V \times A$ to the interval $[0,1]$ such that for each s in A and σ in V, $p_{s,\sigma}$ is a probability density over A, i.e.,

(a) for each s' in A, $p_{s,\sigma}(s') \geq 0$,

(b) $\sum_{s' \in A} p_{s,\sigma}(s') = 1$,

(iv) q is a function on $A \times V_I \times V_O$ to $[0,1]$ such that for each s in A and σ in V, $q_{s,\sigma}$ is a probability density over V_O,

(v) s_0 is in A.

In the probabilistic generalization of the automaton for column addition, the number of possible parameters that can be introduced is uninterestingly large. Each transition $M(k,(m,n))$ may be replaced by a probabilistic transition $1 - \epsilon_{k,m,n}$ and $\epsilon_{k,m,n}$, and each output $Q(k(m,n))$, by 10 probabilities for a total of 2200 parameters.

A three-parameter automaton model structurally rather close to the regression model is easily defined. First, two parameters, ϵ and η, are introduced according to whether there is a 'carry' to the next column.

$$P(M(k,(m,n)) = 0 \mid k + m + n \leq 9) = 1 - \epsilon$$

and

$$P(M(k,(m,n)) = 1 \mid k + m + n > 9) = 1 - \eta \ .$$

In other words, if there is no 'carry', the probability of a correct transition is $1 - \epsilon$ and if there is a 'carry' the probability of such a transition is $1 - \eta$. The third parameter, γ, is simply the probability of an output error. Conversely, the probability of a correct output is:

$$P(Q(k,(m,n)) = (k + m + n) \bmod 10) = 1 - \gamma \ .$$

Consider now exercise i with C_i carrys and D_i digits. If we ignore the probability of two errors leading to a correct response (e.g., a transition error followed by an output error), then the probability of a correct answer is just

(4) $$P(\text{Correct Answer to Exercise } i) = (1 - \gamma)^{D_i}(1 - \eta)^{C_i}(1 - \epsilon)^{D_i - C_i - 1} \ .$$

As already indicated, it is important to realize that this equation is an approximation of the 'true' probability. However, to compute the exact probability it is necessary to make a definite assumption about how the probability γ of an output error is distributed among the nine possible wrong responses. A simple and intuitively appealing one-parameter model is the one that arranges the 10 digits on a circle in natural order with 9 next to 0, and then makes the probability of an error j steps to the right or left of the correct response δ^j. For example, if 5 is the correct digit, then the probability of responding 4 is δ, of 3 is δ^2, of 2 is δ^3, of 1 is δ^4, of 0 is δ^5, of 6 is δ, of 7 is δ^2, etc. Thus in terms of the original model

$$\gamma = 2(\delta + \delta^2 + \delta^3 + \delta^4) + \delta^5 .$$

Consider now the exercise

$$\begin{array}{r} 47 \\ \underline{+\ 15} \end{array} .$$

Then, where d_i = the i^{th} digit response,

$$P(d_1 = 2) = (1 - \gamma) ,$$

$$P(d_2 = 6) = (1 - \gamma)(1 - \eta) + \eta\delta .$$

Here the additional term is $\eta\delta$, because if the state entered is 0 rather than 1 when the pair (7,5) is input, the only way of obtaining a correct answer is for 6 to be given as the sum of $0 + 4 + 1$, which has a probability δ. Thus the probability of a correct response to this exercise is $(1 - \gamma)[(1 - \gamma)(1 - \eta) + \eta\delta]$. Hereafter we shall ignore the $\eta\delta$ (or $\epsilon\delta$) terms.

Returning to (4) we may get a direct comparison with the linear regression model defined by (3), if we take the logarithm of both sides to obtain:

(5) $\log p_i = D_i \log (1 - \gamma) + C_i \log (1 - \eta) + (D_i - C_i - 1) \log (1 - \epsilon)$,

and estimate $\log 1 - \gamma$, $\log 1 - \eta$, and $\log 1 - \epsilon$ by regression with the additive constant set equal to zero. We also may use some other approach to estimation such as minimum χ^2 or maximum likelihood. An analytic solution of the standard maximum-likelihood equations is difficult, but the maximum of the likelihood function can be found numerically.

The automaton model naturally suggests a more detailed analysis of the data. Unlike the regression model, the automaton provides an immediate analysis of the digit-by-digit responses. Ignoring the $\epsilon\delta$-type terms, we can in fact find the general maximum-likelihood estimates of γ, ϵ, and η when the response data are given in this more explicit form.

Let there be n digit responses in a block of exercises. For $1 \leq i \leq n$ let x_i be the random variable that assumes the value 1 if the i^{th} response is correct and 0 otherwise. It is then easy to see that

$$P(x_i = 1) = \begin{cases} (1 - \gamma) & \text{if } i \text{ is a ones'-column digit,} \\ (1 - \gamma)(1 - \epsilon) & \text{if it is not a ones' column and there is no carry to the } i^{th} \text{ digit,} \\ (1 - \gamma)(1 - \eta) & \text{if there is a carry to the } i^{th} \text{ digit,} \end{cases}$$

granted that $\epsilon\delta$-type terms are ignored. Similarly for the same three alternatives

$$P(x_i = 0) = \begin{cases} \gamma \\ 1 - (1 - \gamma)(1 - \epsilon) \\ 1 - (1 - y)(1 - \eta) \ . \end{cases}$$

So for a string of actual digit responses $x_1,\ldots,x_n$ we can write the likelihood function as:

$$(6)\quad \not{L}(x_1,\ldots,x_n) = (1-\gamma)^a\gamma^b(1-\epsilon)^c(1-\eta)^d[1-(1-\gamma)(1-\epsilon)]^e[1-(1-\gamma)(1-\eta)]^f ,$$

where a = number of correct responses, b = number of incorrect responses in the ones' column, c = number of correct responses not in the ones' column when the internal state is 0, d = number of correct responses when the internal state is 1, e = number of incorrect responses not in the ones' column when the internal state is 0, and f = number of incorrect responses when the internal state is 1. (In the model statistical independence of responses is assured by the correction procedure.) It is more convenient to estimate $\gamma' = 1 - \gamma$, $\epsilon' = 1 - \epsilon$, and $\eta' = 1 - \eta$. Making this change, taking the logarithm of both sides of (6) and differentiating with respect to each of the variables, we obtain three equations that determine the maximum-likelihood estimates of γ', ϵ', and η':

$$\frac{\partial L}{\partial \gamma'} = \frac{a}{\gamma'} - \frac{b}{1-\gamma'} - \frac{e\epsilon'}{1-\gamma'\epsilon'} - \frac{f\eta'}{1-\gamma'\eta'} = 0 ,$$

$$\frac{\partial L}{\partial \epsilon'} = \frac{c}{\epsilon'} - \frac{e\gamma'}{1-\gamma'\epsilon'} = 0 ,$$

$$\frac{\partial L}{\partial \eta'} = \frac{d}{\eta'} - \frac{f\gamma'}{1-\gamma'\eta'} = 0 .$$

Solving these equations, we obtain as estimates:

$$\gamma' = \frac{a-c-d}{a+b-c-d} ,$$

$$\hat{\epsilon}' = \frac{c(a+b-c-d)}{(c+e)(a-c-d)} ,$$

$$\hat{\eta}' = \frac{d(a + b - c - d)}{(d - c)(a - b - d)} .$$

Estimates of the parameters for the same third-grade data already described, as well as a graph of the observed and predicted response probabilities for the exercises shown in Table 1, are given in Chapter 4 of Suppes and Morningstar (1972). (This chapter was written in collaboration with Alex Cannara and he is responsible for the data analysis.) The estimates are: $\hat{\gamma} = .0430$, $\hat{\delta} = .0085$ and $\hat{\eta} = .0576$. The graph of response probabilities is reproduced as Figure 1. A detailed discussion

----
Insert Figure 1 about here
----

of the fit of the model and further analysis of some of the discrepancies are to be found in the chapter mentioned. Here I have tried only to give a sense of how this kind of model can be brought into direct confrontation with data.

<u>Register machines with perceptual instructions</u>. To introduce greater generality and to deepen the analysis to include specific ideas about the perceptual processing of a column-addition exercise, I move on to register machines for the reasons described in Section 4 of the main text. This research is being conducted in collaboration with Lindsay L. Flannery.

For column addition three registers suffice in our scheme of analysis. First there is the stimulus-supported register [SS] that holds an encoded representation of a printed symbol to which the student is perceptually attending. In the present case the alphabet of such symbols consists of the 10 digits and the underline symbol '_'. As a new symbol is attended to, previously stored symbols are lost unless transferred to a

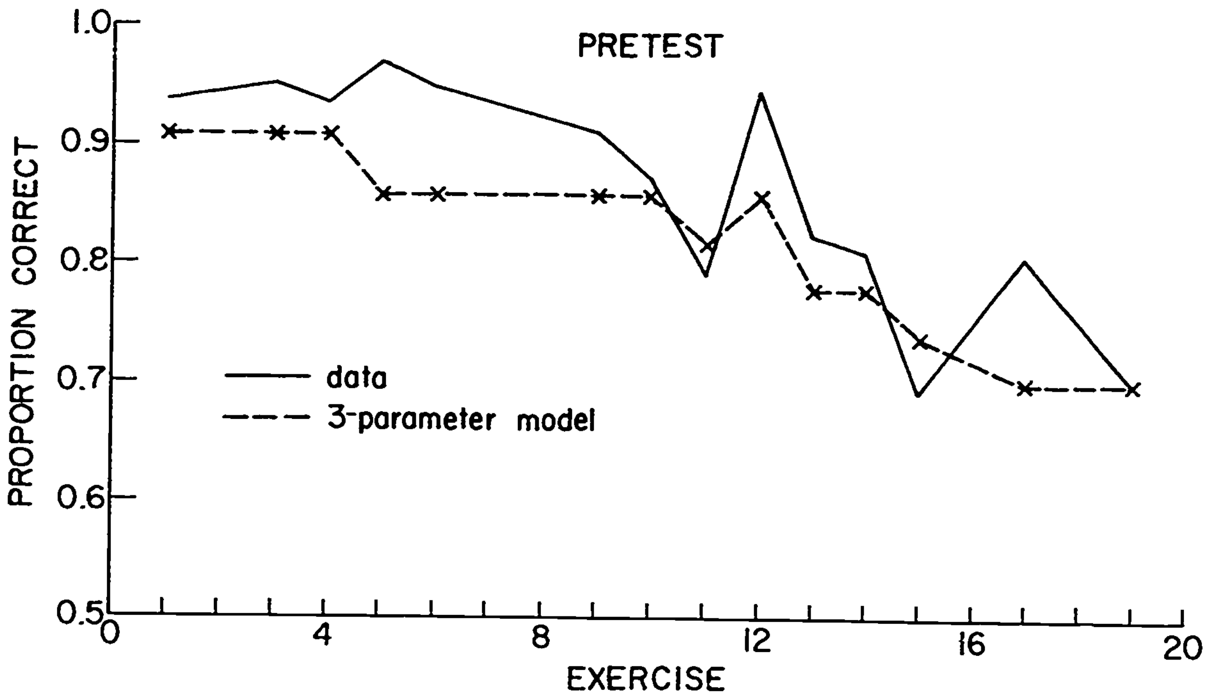

Fig. 1. Predicted and observed probability correct for 3-parameter automaton model.

non-stimulus-supported register. The second register is the non-stimulus-supported register [NSS]. It provides long-term storage for computational results. The third register is the operations register [OP] that acts as a short-term store, both for encodings of external stimuli and for results of calculations carried out on the contents of other registers. It is also primarily non-stimulus-supported.

As already stated in the main text, we drastically simplify the perceptual situation by conceiving each exercise as being presented on a grid with at most one symbol in each square of the grid. For column addition we number the coordinates of the grid from the upper right-hand corner. Thus, in the exercise

$$\begin{array}{r} 15 \\ 24 \\ \underline{+\ 37} \end{array}$$

the coordinates of the digit 5 are (1,1), the coordinates of 4 are (2,1), the coordinates of 7 are (3,1), the coordinates of 1 are (1,2) and so forth, with the first coordinate being the row number and the second being the column number.

The restricted set of instructions we need for column addition are the following 10.

Attend (a,b):	Direct attention to grid position (a,b).
$(\pm a, \pm b)$:	Shift attention on the grid by $(\pm a, \pm b)$.
Readin [SS]:	Read into the stimulus-supported register the physical symbol in the grid position addressed by Attend.
Lookup [R1] + [R2]:	Look up table of basic addition facts for adding contents of register [R1] and [R2] and store the result in [R1].

Copy [R1] in [R2]:	Copy the content of register [R1] in register [R2].
Deleteright [R]:	Delete the rightmost symbol of register [R].
Jump L:	Jump to line labeled L.
Jump (val) R,L:	Jump to line labeled L if content of register [R] is val.
Outright [R]:	Write (output) the rightmost symbol of register [R] at grid position addressed by Attend.
End:	Terminate processing of current exercise.
Exit:	Terminate subroutine processing and return to next line of main program.

Of the 10 instructions only Lookup does not have an elementary character. In our complete analysis it has the status of a subroutine built up from more primitive operations such as those of counting. It is, of course, more than a problem of constructing the table of basic addition facts from counting subroutines; it is also a matter of being able to add a single digit to any number stored in the non-stimulus-supported register [NSS] or [OP], as, for example, in adding many rows of digits in a given column. I omit the details of building up this subroutine.

It should also be obvious that the remaining nine instructions are not a minimal set; for example, the unconditional jump instruction is easily eliminated. We do think the nine are both elementary and psychologically intuitive for the subject matter at hand.

To illustrate in a simple way the use of subroutines, we may consider two that are useful in writing the program for column addition.

The first is the vertical scan subroutine, which is needed for the following purpose. In adding rows of numbers with an uneven number of digits, we cannot simply stop when we reach a blank grid square on the left of the topmost row. We must also scan downward to see if there are digits in that column in any other row. A second aspect of this same problem is that in our model the student is perceptually processing only one grid square at a time, so that he must have a check for finding the bottom row by looking continually for an underline symbol. Otherwise he could, according to an apparently natural subroutine, proceed indefinitely far downward encountering only blanks and leaving entirely the immediate perceptual region of the formatted exercise. Here is the subroutine. In the main program it is preceded by an Attend instruction.

Vertical Scan Subroutine

```
V-scan (0-9,_ )

    Rd      Readin
            Jump (0-9,_ ) SS, Fin
            Attend (+1,-1)
            Readin
            Jump (_ ) SS, Fin
            Attend (+0,+1)
            Jump Rd
    Fin     Exit
```

The labels Rd and Fin of two of the lines are shown on the left.

The second subroutine is one that outputs all the digits in a register working from right to left. For example, in column addition, after the leftmost column has been added, there may still be several digits remaining to print out to the left of this column in the 'answer' row.

```
Output [R]
     Put     Outright [R]
             Deleteright [R]
             Jump (0-9) R, Put
             Exit
```

Using these two subroutines the program for vertical addition is relatively straightforward and requires 26 lines. I number the lines for later reference; they are not a part of the program.

```
     Vertical Addition
1.                  Attend (1,1)
2.                  Readin
3.                  Copy [SS] in [OP]
4.                  Attend (+1,+0)
5.                  Readin
6.        Opr       Lookup [OP] + [SS]
7.        Rd        Attend (+1,0)
8.                  Readin
9.                  Jump (0-9) SS, Opr
10.                 Jump (Blank) SS, Rd
11.                 Attend (+1,0)
```

12.		Outright [OP]
13.		Deleteright [OP]
14.		Copy [OP] in [NSS]
15.		Attend (1,+1)
16.		V-scan (0-9, _)
17.		Jump (_) SS, Fin
18.		Jump (0-9) SS, Car
19.		Copy [SS] in [OP]
20.		Jump Rd
21.	Car	Copy [NSS] in [OP]
22.		Jump Opr
23.	Fin	Jump (Blank) NSS, Out
24.		Attend (+1,0)
25.		Output [NSS]
26.	Out	End

To show how the program works, we may consider a simple one-column addition exercise. I show at the right of each line the content of each register just before the next row is attended to, i.e., after all operations have been performed.

	[SS]	[OP]	[NSS]
4	4	4	
5	5	9	
3	3	12	
8	8	20	
—	—	20	
0	0		2

This kind of analysis can be generalized to prove that the program is correct, i.e., will output the correct answer to any column-addition exercise, but this aspect of matters will not be pursued further here.

By attaching error parameters to various segments of the program, performance models are easily generated. For comparative purposes we may define a performance model essentially identical to the two-state probabilistic automaton already introduced for column addition restricted to two rows. To lines 6-12 we attach the output error parameter γ, and to lines 13-19 we attach the 'carry' error parameter η if there is a carry, and the error parameter ϵ if there is not. Given this characterization of the error parameters the two performance models are behaviorally identical. On the other hand, it is clear that the program for the three-register machine is much more general than the two-state probabilistic automaton, since it is able to solve any vertical addition exercise. It is also obvious that other performance models can easily be defined for vertical addition by introducing error parameters attached to different segments of the program.

Learning. In an earlier article (Suppes, 1969), I proved that given any connected finite automaton there is a stimulus-response model of learning that is asymptotically isomorphic to the automaton, i.e., as the number of trials approaches infinity, and initially all stimuli may be unconditioned to any of the desired responses. In one clear sense, however, the theorem proved is too weak because of the special character of the reinforcement schedule. What is required is reinforcement of the transitions from each response-stimulus pair to the next response, where the responses, internal or external, constitute the states of the automaton. The response on trial n must become conditioned to the pair consisting of the response of trial $n - 1$ and the stimulus on trial n. A complete matching of the reinforcement schedule to such conditioning connections is often not experimentally feasible.

At the other end of the scale, Rottmayer (1970) proved the following theorem. Let $\mathcal{C}$ be a classification scheme for dividing a possibly infinite set of stimuli or stimulus patterns into two classes, such that the classification of any pattern can be accomplished by a finite automaton. Then there is a stimulus-response model that can learn the classification scheme $\mathcal{C}$ given as reinforcement only the information of whether its classification of successively presented patterns is correct or incorrect. The weakness of this theorem is that the learning is very slow, and machinery for building up a hierarchy of concepts is not directly provided.

The shift from automata to register machines seems promising not only for the development of performance models, but also for the construction of learning models. Learning in this framework consists of building internal programs of increasing complexity. The reinforcement procedures realistically fall between the two extremes described above. Verbal directions and corrections correspond closely, but not exactly, to segments of an appropriate program (I emphasize an because the internal program constructed is not necessarily unique).

I restrict myself here to an example of this approach. I take as the class of exercises single-column addition, but with an indefinite number of rows. The program is simpler than the general one given above, and it is easy to see the relation between what is said to the student by the teacher or computer to the desired internal program. In Figure 2 I show the verbal instructions on the right with the physical

Insert Figure 2 about here

pointing to the relevant part of the displayed exercise indicated in parentheses. When errors are made, still more detailed instructions, tailored to the particular error, can be given, but I do not consider such error messages here.

In Figure 2, learning parameters c_1, c_2, c_3 and c_4 are shown for the four segments of the program. Various learning models can be formulated in terms of these four parameters. The simplest is the one that assumes independence of the four parts. If we treat the probability of successive errors combining to yield a correct response as having probability zero, then the mean probability for a correct response on trial n for the independence model is simply:

$$P_n(\text{Correct Response}) = \prod_{c=1}^{4} (1 - (1 - c_i)^{n-1}) .$$

	Internal Program		Verbal Instructions
	Attend (1,1) Readin	c_1	Start here (pointing).
   Opr	Transfer [SS] to [OP] Attend (+1,0) Readin Lookup [OP] + [SS]	c_2	Add first two digits (pointing)
	Attend (+1,0) Readin Jump (0-9) SS, Opr	c_3	Now add again (pointing) (if conditional jump satisfied) or Notice end of column (pointing at __) (if conditional jump not satisfied)
	Attend (+1,0) Output [OP] End	c_4	Write answer here (pointing)

Fig. 2. Single-column addition.

At the other extreme a hierarchical model postulates that the i^{th} segment of the program cannot be learned until the $i-1^{st}$ segment is learned. This hierarchical model leads to the following transition matrix, where state 0 represents all segments as unlearned, state 1 represents the first segment only as learned, etc.

$$
\begin{array}{c|ccccc}
 & 4 & 3 & 2 & 1 & 0 \\
\hline
4 & 1 & 0 & 0 & 0 & 0 \\
3 & c_4 & 1-c_4 & 0 & 0 & 0 \\
2 & 0 & c_3 & 1-c_3 & 0 & 0 \\
1 & 0 & 0 & c_2 & 1-c_2 & 0 \\
0 & 0 & 0 & 0 & c_1 & 1-c_1
\end{array}
$$

Detailed comparison of these two models, especially for testing against data, requires considerable further development, but the relevant mathematical and probabilistic techniques are familiar in the literature of mathematical learning theory.

What is missing from a theoretical standpoint is a deeper conceptualization of the relation between verbal instructions and reinforcements on the one hand and the construction of appropriate segments of internal programs on the other. In the example given above, the crucial concept of iteration or recursion embodied in the conditional jump instruction is presumed to be learned from the instruction "Now add again," with <u>again</u> mainly carrying the force of the recursion. I hope to have something more to say in the near future about this difficult and important problem.

References

Beilin, H. Developmental stages and developmental processes. In D. R. Green, M. P. Ford and G. B. Flamer (Eds.), Measurement and Piaget. New York: McGraw-Hill, 1971. Pp. 172-197.

Bever, T. G., Fodor, J. A., and Garrett, M. Formal limitation of association. In T. R. Dixon and D. L. Horton (Eds.), Verbal behavior and general behavior theory. Englewood Cliffs, N. J.: Prentice-Hall, 1968. Pp. 582-585.

Bloomfield, L., and Barnhart, C. L. Let's read, a linguistic approach. Detroit: Wayne State Univ. Press, 1961.

Buckingham, B. R. Adding up or down: A discussion. Journal of Educational Research, 1925, 12, 251-261.

Campbell, D. T., and Stanley, J. C. Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), Handbook for research on teaching. Chicago: Rand McNally, 1963. Pp. 171-246.

Chomsky, N. Language and mind. New York: Harcourt Brace Jovanovich, 1972.

Cronbach, L. J., and Suppes, P. (Eds.) Research for tomorrow's schools. New York: Macmillan, 1969.

Goldberg, A., and Suppes, P. A computer-assisted instruction program for exercises on finding axioms. Educational Studies in Mathematics, 1972 (in press).

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

Hume, D. Enquiries concerning human understanding and concerning the principles of morals. (2nd ed.) Oxford: Clarendon Press, 1902.

Inhelder, B., and Piaget, J. The early growth of logic in the child. New York: Harper and Row, 1964. English translation by E. A. Lunzer and D. Papert.

Kane, M. T. Variability in the proof behavior of college students in a CAI course in logic as a function of problem characteristics. Technical Report No. 192, 1972, Stanford University, Institute for Mathematical Studies in the Social Sciences.

Laplace, P. S. Theorie analytique des probabilities. Troisième édition revue et augmentée par l'auteur. Paris: Courcier, 18_0.

Moloney, J. M. An investigation of college student performance on a logic curriculum in a computer-assisted instruction setting. Technical Report No. 183, 1972, Stanford University, Institute for Mathematical Studies in the Social Sciences.

Piaget, J. Discussion in J. M. Tanner and B. Inhelder (Eds.), Discussions on child development. Vol. 4. New York: International Universities Press, 1960.

Piaget, J. Discussion in D. R. Green, M. P. Ford and G. B. Flamer (Eds.), Measurement and Piaget. New York: McGraw-Hill, 1971.

Skinner, B. F. The technology of teaching. New York: Appleton-Century Croft, 1968.

Suppes, P. On the behavioral foundations of mathematical concepts. Monographs of the Society for Research in Child Development, 1965, 30, 60-96.

Suppes, P. The desirability of formalization in science. Journal of Philosophy, 1968, 65, 651-664.

Suppes, P. Stimulus-response theory of finite automata. Journal of Mathematical Psychology, 1969, 6, 327-355.

Suppes, P., and Ginsberg, R. A fundamental property of all-or-none models, binomial distribution of responses prior to conditioning, with application to concept formation in children. Psychological Review, 1963, 70, 139-161.

Suppes, P., and Groen, G. J. Some counting models for first grade performance data on simple addition facts. In J. M. Scandura (Ed.), Research in mathematics education. Washington, D. C.: National Council of Teachers of Mathematics, 1967.

Suppes, P., Hyman, L., and Jerman, M. Linear structural models for response and latency performance in arithmetic on computer-controlled terminals. In J. P. Hill (Ed.), Minnesota symposia on child psychology. Minneapolis: Univ. of Minnesota Press, 1967. Pp. 160-200.

Suppes, P., Jerman, M., and Brian, D. Computer-assisted instruction: The 1965-66 Stanford arithmetic program. New York: Academic Press, 1968.

Suppes, P., and Morningstar, M. Computer-assisted instruction at Stanford, 1966-68: Data, models, and evaluation of the arithmetic programs. New York: Academic Press, 1972.

Thorndike, E. L. Educational psychology. Vols. 1, 2, 3. New York: Teachers College, Columbia University, 1913-1914.

Thorndike, E. L. The psychology of arithmetic. New York: Macmillan, 1922.

Footnotes

[1]Roger Brown puts the matter nicely, ". . . one naturally falls into the habit of speaking of stage I and stage II and so on. There is no harm in that so long as we recognize that these are imposed stages, laid upon continuous data by the investigator as an analytic convenience [Psycholinguistics, 1970, p. 100]."

[2]The research reported in the Appendix has been supported by the National Science Foundation Grant NSFGJ-443X and U. S. Office of Education Grant OEG-970-0024(057).

(Continued from inside front cover)

96 R. C. Atkinson, J. W. Breisford, and R. M. Shiffrin. Multi-process models for memory with applications to a continuous presentation task. April 13, 1966. (J. math. Psychol., 1967, 4, 277-300).
97 P. Suppes and E. Crothers. Some remarks on stimulus-response theories of language learning. June 12, 1966.
98 R. Bjork. All-or-none subprocesses in the learning of complex sequences. (J. math. Psychol., 1968, 1, 182-195).
99 E. Gammon. The statistical determination of linguistic units. July 1, 1966.
100 P. Suppes, L. Hyman, and M. Jerman. Linear structural models for response and latency performance in arithmetic. (In J. P. Hill (ed.), Minnesota Symposia on Child Psychology. Minneapolis, Minn.: 1967. Pp. 160-200).
101 J. L. Young. Effects of intervals between reinforcements and test trials in paired-associate learning. August 1, 1966.
102 H. A. Wilson. An investigation of linguistic unit size in memory processes. August 3, 1966.
103 J. T. Townsend. Choice behavior in a cued-recognition task. August 8, 1966.
104 W. H. Batchelder. A mathematical analysis of multi-level verbal learning. August 9, 1966.
105 H. A. Taylor. The observing response in a cued psychophysical task. August 10, 1966.
106 R. A. Bjork. Learning and short-term retention of paired associates in relation to specific sequences of interpresentation intervals. August 11, 1966.
107 R. C. Atkinson and R. M. Shiffrin. Some Two-process models for memory. September 30, 1966.
108 P. Suppes and C. Ihrke. Accelerated program in elementary-school mathematics--the third year. January 30, 1967.
109 P. Suppes and I. Rosenthal-Hill. Concept formation by kindergarten children in a card-sorting task. February 27, 1967.
110 R. C. Atkinson and R. M. Shiffrin. Human memory: a proposed system and its control processes. March 21, 1967.
111 Theodore S. Rodgers. Linguistic considerations in the design of the Stanford computer-based curriculum in initial reading. June 1, 1967.
112 Jack M. Knutson. Spelling drills using a computer-assisted instructional system. June 30, 1967.
113 R. C. Atkinson. Instruction in initial reading under computer control: the Stanford Project. July 14, 1967.
114 J. W. Breisford, Jr. and R. C. Atkinson. Recall of paired-associates as a function of overt and covert rehearsal procedures. July 21, 1967.
115 J. H. Stelzer. Some results concerning subjective probability structures with semiorders. August 1, 1967
116 D. E. Rumelhart. The effects of interpresentation intervals on performance in a continuous paired-associate task. August 11, 1967.
117 E. J. Fishman, L. Keller, and R. E. Atkinson. Massed vs. distributed practice in computerized spelling drills. August 18, 1967.
118 G. J. Groen. An investigation of some counting algorithms for simple addition problems. August 21, 1967.
119 H. A. Wilson and R. C. Atkinson. Computer-based instruction in initial reading: a progress report on the Stanford Project. August 25, 1967.
120 F. S. Roberts and P. Suppes. Some problems in the geometry of visual perception. August 31, 1967. (Synthese, 1967, 17, 173-201)
121 D. Jamison. Bayesian decisions under total and partial ignorance. D. Jamison and J. Kozielecki. Subjective probabilities under total uncertainty. September 4, 1967.
122 R. C. Atkinson. Computerized instruction and the learning process. September 15, 1967.
123 W. K. Estes. Outline of a theory of punishment. October 1, 1967.
124 T. S. Rodgers. Measuring vocabulary difficulty: An analysis of item variables in learning Russian-English and Japanese-English vocabulary parts. December 18, 1967.
125 W. K. Estes. Reinforcement in human learning. December 20, 1967.
126 G. L. Wolford, D. L. Wessel, W. K. Estes. Further evidence concerning scanning and sampling assumptions of visual detection models. January 31, 1968.
127 R. C. Atkinson and R. M. Shiffrin. Some speculations on storage and retrieval processes in long-term memory. February 2, 1968.
128 John Holmgren. Visual detection with imperfect recognition. March 29, 1968.
129 Lucille B. Mlodnosky. The Frostig and the Bender Gestalt as predictors of reading achievement. April 12, 1968.
130 P. Suppes. Some theoretical models for mathematics learning. April 15, 1968. (Journal of Research and Development in Education, 1967, 1, 5-22)
131 G. M. Olson. Learning and retention in a continuous recognition task. May 15, 1968.
132 Ruth Norene Hartley. An investigation of list types and cues to facilitate initial reading vocabulary acquisition. May 29, 1968.
133 P. Suppes. Stimulus-response theory of finite automata. June 19, 1968.
134 N. Moler and P. Suppes. Quantifier-free axioms for constructive plane geometry. June 20, 1968. (In J. C. H. Gerretsen and F. Oort (Eds.), Compositio Mathematica. Vol. 20. Groningen, The Netherlands: Wolters-Noordhoff, 1968. Pp. 143-152.)
135 W. K. Estes and D. P. Horst. Latency as a function of number or response alternatives in paired-associate learning. July 1, 1968.
136 M. Schlag-Rey and P. Suppes. High-order dimensions in concept identification. July 2, 1968. (Psychom. Sci., 1968, 11, 141-142)
137 R. M. Shiffrin. Search and retrieval processes in long-term memory. August 15, 1968.
138 R. D. Freund, G. R. Loftus, and R.C. Atkinson. Applications of multiprocess models for memory to continuous recognition tasks. December 18, 1968.
139 R. C. Atkinson. Information delay in human learning. December 18, 1968.
140 R. C. Atkinson, J. E. Holmgren, and J. F. Juola. Processing time as influenced by the number of elements in the visual display. March 14, 1969.
141 P. Suppes, E. F. Loftus, and M. Jerman. Problem-solving on a computer-based teletype. March 25, 1969.
142 P. Suppes and Mona Morningstar. Evaluation of three computer-assisted instruction programs. May 2, 1969.
143 P. Suppes. On the problems of using mathematics in the development of the social sciences. May 12, 1969.
144 Z. Domotor. Probabilistic relational structures and their applications. May 14, 1969.
145 R. C. Atkinson and T. D. Wickens. Human memory and the concept of reinforcement. May 20, 1969.
146 R. J. Titiev. Some model-theoretic results in measurement theory. May 22, 1969.
147 P. Suppes. Measurement: Problems of theory and application. June 12, 1969.
148 P. Suppes and C. Ihrke. Accelerated program in elementary-school mathematics--the fourth year. August 7, 1969.
149 D. Rundus and R.C. Atkinson. Rehearsal in free recall: A procedure for direct observation. August 12, 1969.
150 P. Suppes and S. Feldman. Young children's comprehension of logical connectives. October 15, 1969.

(Continued on back cover)

FILMED FROM BEST AVAILABLE COPY

(Continued from inside back cover)

151 Joaquim H. Laubsch. An adaptive teaching system for optimal item allocation. November 14, 1969.
152 Roberta L. Klatzky and Richard C. Atkinson. Memory scans based on alternative test stimulus representations. November 25, 1969.
153 John E. Holmgren. Response latency as an indicant of information processing in visual search tasks. March 16, 1970.
154 Patrick Suppes. Probabilistic grammars for natural languages. May 15, 1970.
155 E. Gammon. A syntactical analysis of some first-grade readers. June 22, 1970.
156 Kenneth N. Wexler. An automaton analysis of the learning of a miniature system of Japanese. July 24, 1970.
157 R. C. Atkinson and J.A. Paulson. An approach to the psychology of instruction. August 14, 1970.
158 R.C. Atkinson, J.D. Fletcher, H.C. Chetin, and C.M. Stauffer. Instruction in initial reading under computer control: the Stanford project. August 13, 1970.
159 Dewey J. Rundus. An analysis of rehearsal processes in free recall. August 21, 1970.
160 R.L. Klatzky, J.F. Juola, and R.C. Atkinson. Test stimulus representation and experimental context effects in memory scanning.
161 William A. Rottmayer. A formal theory of perception. November 13, 1970.
162 Elizabeth Jane Fishman Loftus. An analysis of the structural variables that determine problem-solving difficulty on a computer-based teletype. December 18, 1970.
163 Joseph A. Van Campen. Towards the automatic generation of programmed foreign-language instructional materials. January 11, 1971.
164 Jamesine Friend and R.C. Atkinson. Computer-assisted instruction in programming: AID. January 25, 1971.
165 Lawrence James Hubert. A formal model for the perceptual processing of geometric configurations. February 19, 1971.
166 J. F. Juola, I.S. Fischler, C.T. Wood, and R.C. Atkinson. Recognition time for information stored in long-term memory.
167 R.L. Klatzky and R.C. Atkinson. Specialization of the cerebral hemispheres in scanning for information in short-term memory.
168 J.D. Fletcher and R.C. Atkinson. An evaluation of the Stanford CAI program in initial reading (grades K through 3). March 12, 1971.
169 James F. Juola and R.C. Atkinson. Memory scanning for words versus categories.
170 Ira S. Fischler and James F. Juola. Effects of repeated tests on recognition time for information in long-term memory.
171 Patrick Suppes. Semantics of context-free fragments of natural languages. March 30, 1971.
172 Jamesine Friend. Instruct coders' manual. May 1, 1971.
173 R.C. Atkinson and R.M. Shiffrin. The control processes of short-term memory. April 19, 1971.
174 Patrick Suppes. Computer-assisted instruction at Stanford. May 19, 1971.
175 D. Jamison, J.D. Fletcher, P. Suppes and R.C.Atkinson. Cost and performance of computer-assisted instruction for compensatory education.
176 Joseph Offir. Some mathematical models of individual differences in learning and performance. June 28, 1971.
177 Richard C. Atkinson and James F. Juola. Factors influencing speed and accuracy of word recognition. August 12, 1971.
178 P. Suppes, A. Goldberg, G. Kanz, B. Searle and C. Stauffer. Teacher's handbook for CAI courses. September 1, 1971.
179 Adele Goldberg. A generalized instructional system for elementary mathematical logic. October 11, 1971.
180 Max Jerman. Instruction in problem solving and an analysis of structural variables that contribute to problem-solving difficulty. November 12, 1971.
181 Patrick Suppes. On the grammar and model-theoretic semantics of children's noun phrases. November 29, 1971.
182 Georg Kreisel. Five notes on the application of proof theory to computer science. December 10, 1971.
183 James Michael Moloney. An investigation of college student performance on a logic curriculum in a computer-assisted instruction setting. January 28, 1972.
184 J.E. Friend, J.D. Fletcher and R.C. Atkinson. Student performance in computer-assisted instruction in programming. May 10, 1972.
185 Robert Lawrence Smith, Jr. The syntax and semantics of ERICA. June 14, 1972.
186 Adele Goldberg and Patrick Suppes. A computer-assisted instruction program for exercises on finding axioms. June 23, 1972.
187 Richard C. Atkinson. Ingredients for a theory of instruction. June 26, 1972.
188 John D. Bonvillian and Veda R. Charrow. Psycholinguistic implications of deafness: A review. July 14, 1972.
189 Phipps Arabie and Scott A. Boorman. Multidimensional scaling of measures of distance between partitions. July 26, 1972.
190 John Ball and Dean Jamison. Computer-assisted instruction for dispersed populations: System cost models. September 15, 1972.
191 William R. Sanders and John R. Ball. Logic documentation standard for the Institute for Mathematical Studies in the Social Sciences. October 4, 1972.
192 M.T. Kane. Variability in the proof behavior of college students in a CAI course in logic as a function of problem characteristics. October 6, 1972.
193 P. Suppes. Facts and fantasies of education. October 18, 1972.