The Formation of Mathematical Concepts in Primary-Grade Children

Patrick Suppes

THE aims of this paper are primarily constructive, that is, to contribute to the development of a scientific theory of concept formation. However, before turning to this subject, there are two general aspects of the teaching of mathematical concepts upon which I want to comment.[1]

Teaching of Mathematical Concepts

The first concerns the much heard remark that the newer revisions of the mathematics curriculum are particularly significant because of the emphasis they are placing on *understanding* concepts as opposed to the perfection of *rote* skills. My point is not to disagree with this remark, but to urge that it is essentially banal. It is a good thing to understand; it is a bad thing to possess mere rote skill. The banality arises from not knowing what we mean by *understanding*. This failure is not due to disagreement over whether the test of understanding should be a behavioral one. I am inclined to think that most people concerned with this matter would admit the central relevance of overt behavior as a measure of understanding. The difficulty is rather that no one seems to be very clear about the exact specification of the behavior required to exhibit understanding. Moreover, apart even from any behavioral questions the very notion of understanding seems fundamentally vague and ill-defined.

[1] This research has been supported by the United States Office of Education, Department of Health, Education and Welfare, the National Science Foundation and the Carnegie Corporation of New York.

To illustrate what I mean, let us suppose that we can talk about understanding in some general way. Consider now the concept of triangularity. Does understanding this concept entail understanding that the sum of the interior angles is 180°, or that triangles are rigid whereas quadrilaterals are not, or being able to prove that if the bisectors of two angles of a triangle are equal then the triangle is isosceles? This example suggests one classical philosophical response to our query, namely, to understand a concept means, it is said, to know or to believe as true a certain set of propositions that use the concept.

Unfortunately, this set is badly defined. It is trivial to remark that along these lines we might work out a comparative notion of understanding that is a partial ordering defined in terms of the inclusion relation among sets of propositions that use the concept. Thus one person understands the concept of triangularity better than a second if the set of propositions that use the concept and that are known to the first person includes the corresponding set for the second person. (Notice that it will not do to say simply that the first person knows more propositions using the concept, for the second person might know fewer propositions but among these would be some of the more profound propositions not known by the first person; this situation corresponds to the widely held and probably correct belief that the deepest mathematicians are not necessarily the best mathematical scholars.)

Yet this partial ordering does not take us very far. A more behavioral line of thought that at first glance may seem more promising is the response of the advocates of programed learning to the charge that the learning of programed material facilitates rote skills but not genuine understanding of concepts. The response is to assert that if the critics will simply specify the behavior they regard as providing evidence of understanding, they, the programers, will guarantee to develop and perfect the appropriate repertoire of responses. This approach has the practical virtue of sidestepping any complex discussion of understanding, and supposes, with considerable correctness, no doubt, that without giving an intellectually exact analysis of what it means to understand a concept, we still can obtain a rough consensus at any given time as to what body of propositions we expect students to master about a given concept. This is the appropriate practical engineering approach, but it scarcely touches the scientific problem.

In this paper I do not pretend to offer any serious characterization of what it means to understand a concept. I do think that the most promising direction is to develop a psychological theory of concept transfer and generalization. The still relatively primitive state of the theory of the much simpler phenomena of stimulus transfer and generalization does not make me optimistic about the immediate future. For immediate

purposes, however, let me sketch in a very rough way how the application of ideas of transfer and generalization can be used to attack the banality mentioned earlier in the standard dichotomy of understanding versus rote skill. We would all agree, I think, that such matters as learning to give the multiplication tables quickly and with accuracy are indeed rote skills. Yet there is also what I consider to be a mistaken tendency to extend the label "rote skill" to many parts of the traditional mathematics curriculum at all levels. The body of mathematical material tested, for example, by the British Sixth Form examinations is sometimes so labeled by advocates of the newer mathematics curriculum. In terms of the accepted notion of rote skill developed and studied by psychologists, this is a mistake, for the production of a correct response on these examinations cannot be explained by any simple principles of stimulus-response association. Moreover, the problems of transfer involved in solving typical British Sixth Form examination problems in comparison with the kind of examination set by advocates of the newer mathematics curriculum may in fact require more transfer of concepts—at least more transfer in one obvious way of measuring transfer, that is, in terms of the number of hours of training spent in relation to the ability to solve the problems on the part of students matched as to general background and ability. I recognize that these are complicated matters and I do not want to pursue them here. Also I am fully in sympathy with the general objectives of the newer mathematics curriculum. I am simply protesting here against some of the remarks about understanding and rote skills that occur in the pedagogical conversations and writing of mathematicians.

A second general point relates to the many current discussions of the efficacy of the discovery method of teaching. The discussions of this method seem to provide yet one more remarkable example in the history of education of a viewpoint coming to prominence without any serious body of empirical evidence to support or refute its advocates. From the standpoint of learning theory I do not even know of a relatively systematic definition of the discovery method. I do not doubt that some of the advocates of this method are themselves remarkably capable teachers and are able to do unusual and startling things with classes of elementary school children.

The intellectual problem, however, is to separate the pedagogical virtuosities of personality of these advocates from the systematic problem of analyzing the method itself. Workable hypotheses need to be formulated and tested. I know that it is a standard objection of some advocates of the discovery method that any quick laboratory examination of this teaching method versus a more standard immediate reinforcement method, particularly as applied to young children, is bound not to yield an unbiased test. The results and the implications of the method, it is said,

can only be properly evaluated after a long period. I personally rather doubt that this is the case but, if it is so, or if it is propounded as a working hypothesis by advocates of the method, then it seems to me that it is their intellectual responsibility to formulate proper tests of a sufficiently sustained sort. I realize that my remarks on this subject in this paper have the character of *obiter dicta.* On the other hand, in a more complete treatment of mathematical concept formation in young children, I would consider it necessary to probe more deeply into the issues of motivation, reinforcement and concept formation surrounding the controversy between the discovery method and other more classical methods of reinforcement.

I turn now to specific topics I would like to develop more systematically. In the next section an intuitive version of stimulus-sampling learning theory is sketched; it is this theory we have attempted to apply to the experiments dealing with mathematical concept formation in young children described in the following section. A description of these experiments is concluded by attempting to summarize some of the more important findings from the standpoint of pedagogical practice.

In the final section I give a brief description of the related pedagogical program in elementary-school mathematics. It should be emphasized that the empirical variables considered in the experiments have not been arbitrarily selected. The experimental program has been carried out concurrently with the development of a program of primary-grade mathematics, "Sets and Numbers," which makes use of simple set-theoretical notions to introduce the child at the outset to concepts and operations of arithmetic. A concept to be learned in each experiment and the experimental variables introduced have some bearing on problems which have arisen in the development of the Sets and Numbers program.

Sketch of Stimulus-Sampling Learning Theory

In paradigmatic form the theory was developed for an experimental situation in which a subject is presented on each trial with some stimulus display to which he is required to make a specific response. The subject is always told whether he is correct or incorrect on that trial.

The basic assumption of the theory is that in such a situation the correct response will become associated with, or conditioned to, the stimulus display. If it is conditioned, the subject will continue to respond correctly whenever the same stimulus is presented. On the other hand, when the correct response is not conditioned to a stimulus display, the probability of a correct response is at some chance level which remains constant until the association between the stimulus and appropriate response has been established.

More generally, the postulated sequence of events occurring on a given trial in an experiment may be described as follows: The organism begins the trial in a certain state of conditioning. Among the stimuli available a certain set is sampled. On the basis of this sampled stimuli and their conditioning connections to the possible responses, the response is made. After the response is made, a reinforcement occurs which may change the conditioning of the sampled stimuli. The organism then enters a new state of conditioning ready for the next trial.

In the experiments to be described here we shall apply the theory in a simple fashion by postulating that there is exactly one stimulus element available for sampling on each trial and that at the beginning of the experiment this single element is unconditioned. This special one-element model has been applied with considerable success to paired-associate experiments, that is, to experiments in which the subject must learn an arbitrary association established by the experimenter between, say, a nonsense-syllable of a single stimulus and the response such as one of the numerals, 1 to 8, or by the pressing of one of three keys. The most important psychological implication of this one-element model of the theory is that in the paired-associate situation the conditioning takes place on an all-or-none basis. This means that prior to conditioning the organism is simply guessing the correct response with a fixed probability. Once the stimulus is conditioned the correct response is made with probability one. The kind of learning described with this model is often referred to as insightful learning or the "ah ha" experience. I believe it is important to emphasize that learning of this qualitative character can be given an exact quantitative formulation within stimulus-response theory.

In an earlier paper Rose Ginsberg and I (1963) analyzed a number of experiments, including some of those reported here, to exhibit a simple but fundamentally important fact about this all-or-none conditioning model. The assumptions of the model imply that the sequence of correct and incorrect responses prior to the last error form a binomial distribution of Bernoulli trials with parameter p. This null hypothesis of a fixed binomial distribution of responses prior to the last error admits at once the possibility of applying many powerful classical statistics that are not usually applicable to learning data. What is particularly important from a psychological standpoint is the implication of this hypothesis that the mean learning curve, when estimated over responses prior to the last error, will be a horizontal line. In other words, no effects of learning should be shown prior to conditioning. Ginsberg and I analyzed experiments concerned with children's concept formation, animal learning, probability learning and paired-associate learning in adults from this standpoint. I shall not propose to give as extensive analysis of data in

the present paper as we attempted there. I will, however, attempt to cite some of the results on the question of stationarity because of its fundamental importance for any psychological evaluation of the kind of processes by which young children acquire concepts.

Other features of the experiments summarized in the following pages will be mentioned seriatim, particularly if they have some bearing on pedagogical questions. There is, however, one general methodological point that should be mentioned before individual experiments are described. In many of the experiments described the stimulus displays were different on every trial, so that there was no possibility of establishing a simple stimulus-response association. How is the one-element model to be applied to such data? The answer represents, I think, one of our more important general findings. It is that a very good account of much of the data may be obtained by treating the concept itself as the single element. The schema is then that a simple concept-response association is established. Except for Experiment I, it is with this interpretation that we have applied the one-element model to our experiments.

Experiment I. Binary Numbers

This experiment is reported in detail in Suppes and Ginsberg (1962b). Five- and six-year-old subjects were required to learn the concepts of the numbers 4 and 5 in the binary number system, each concept being represented by three different stimuli; for example, if the stimuli had been chosen from the Roman alphabet, as in fact they were not, 4 could have been represented by abb, cdd and eff, and 5 by aba, cdc and efe. The child was required to respond by placing directly upon the stimulus one of two cards. On one card was inscribed a large Arabic numeral 4 and on the other a large Arabic numeral 5. All the children were told on each trial whether they made the correct or incorrect response, but half of them were also required to *correct* their responses following an error.

Thus in this experiment in addition to testing the one-element model, we were concerned to examine the effect upon learning of requiring the subject to correct overtly a wrong response. There were 24 subjects in each of the two experimental groups. From test responses, after each experimental session, it seemed evident that whereas some subjects in both groups learned the concept as such, others learned only some of the specific stimuli representing the concepts, so that in effect within each group there were two subgroups of subjects. It is interesting to note that this finding is in agreement with some similar results in lower organisms (Hull and Spence, 1938), but contrary to results obtained with adult subjects, for whom an overt correction response seems to have negligible behavioral effects (Burke, Estes and Hellyer, 1954).

The data for both correction and noncorrection groups show that there was a significant difference between the two groups in the rate of learning. The t of 4.00 computed between overall responses of the two groups is significant at the .001 level.

For the analysis of paired-associates and concept formation we restricted ourselves to the 24 subjects of the correction group. To begin with, we analyzed the data as if each of the six stimuli, three for each number, represented an independent paired-associate item. Because a total of only 16 trials were run on each stimulus we adopted a criterion of six successive correct responses and thus the proportion of correct responses prior to the last error was examined only for the first ten trials. An X^2 test of stationarity over blocks of single trials supports the null hypothesis ($X^2 = 8.00$, df $=9$, $P > .50$, $N=844$).

Let us now turn to the question of concept formation. The identification we make has already been indicated. We treat the concept as itself the single stimulus, and in this case we regard the experiment as consisting of two concepts, one for the number 4 and one for the number 5. (It should be apparent that the identification in terms of the numbers 4 and 5 is not necessary; each concept can be viewed simply as an abstract pattern.)

Using as the criterion that the concept had been learned if the responses to the last three presentations of each stimuli represented were correct, we divided the data into two parts. The data from the group meeting the criterion were arranged for concept learning analysis, in this case a two-item learning situation. The remaining data were assumed to represent paired-associate learning involving six independent stimulus items. For the paired-associate group over the first 10 trials we had 81 cases. For the concept-formation group we had 21 cases with 48 trials in each. The X^2 test of stationarity was not significant for either group (for the concept subgroup $X^2 = 8.36$, df $= 9$, $P > .30$, $N = 357$; for the paired-associate subgroup $X^2 = 11.26$, df $= 8$, $P > .10$, $N = 570$).

To provide a more delicate analysis of this important question of stationarity we can construct Vincent curves (*cf.* Suppes and Ginsberg, 1963). The proportion of correct responses prior to the last error may be tabulated for percentiles of trials instead of in terms of the usual blocks of trials. There was some evidence of non-stationarity in the final quartile.

It should be noted, of course, that the subjects who take longer to reach criterion are weighted more heavily in the Vincent curves. For example, suppose one subject has sixteen responses prior to his last error whereas another subject has four responses prior to his last error. The first subject contributes four responses to each quartile whereas the second subject contributes only one. Let us turn now to the second experiment.

Experiment II. Equivalence and Identity of Sets

This experiment was performed with Rose Ginsberg, and we have published some of the results in Suppes and Ginsberg (1963) but the detailed report of the experiment has not yet appeared. The learning tasks involved in this experiment were equivalence of sets and the two related concepts of identity of sets and identity of ordered sets.

Ninety-six first grade subjects were run, in four groups of twenty-four each. In Group 1 the subjects were required to learn identity of sets for 56 trials and then equivalence for a further 56 trials. In Group 1 this order of presentation was reversed. In Group 3 the subjects learned first identity of ordered sets and then identity of sets. In Group 4 identity of sets preceded identity of ordered sets. Following our findings in Experiment I, that is, that learning was more rapid when the child was required to make an overt correction response after an error, we included this requirement in Experiment II and most of the subsequent experiments reported here. Also in this experiment and those reported later in this paper as well, no stimulus display on any trial was repeated for an individual subject. This was done in order to guarantee that the learning of the concept could not be explained by any simple principles of stimulus-response association, as was the case for Experiment I. For convenience of reference we have termed concept experiments in which no stimulus display is repeated *pure* property or *pure* concept experiments.

The sets depicted by the stimulus displays consisted of one, two or three elements. On each trial two of these sets were displayed. Minimal instructions were given the subjects to press one of two buttons when the stimulus pairs presented were "the same" and the alternative button when they were "not the same."

Our empirical aims in this experiment were several. First, we wanted to examine in detail whether the learning of simple set concepts by children of this age took place on an all-or-none conditioning basis. Secondly, as the two sequences of learning trials on two different concepts for each group would indicate, we were interested in questions of transfer. Would the learning of one kind of concept facilitate the learning of another, and were there significant differences in the degree of this facilitation? Thirdly, we were concerned to consider the question of finding the behavioral level at which the concepts could be most adequately defined. For example, in learning the identity of sets could the learning trials be satisfactorily analyzed from the standpoint of all trials falling under a single concept? Would it be better to separate the trials on which identical sets were presented from those on which nonidentical sets were presented in order to analyze the data in terms of two concepts? Or would a still finer division of concepts be called for in terms of sets identical in terms of order (O),

sets identical as non-ordered sets ($I\bar{O}$), equivalent sets ($E\bar{I}$) and non-equivalent sets?

In somewhat summary fashion the experimental results were as follows: First, the relevant mean learning curves showed that the number of errors on the concept of identity of ordered sets was extremely small. From the high proportion of correct responses even in the first block of trials of this concept it was evident that this is a very natural and simple concept for children.

Learning curves for trials before the last error were also constructed for all groups. To identify the last error prior to conditioning, we adopted a criterion of 16 successive correct responses. The combined curve for Groups 1a and 4a was clearly stationary. This was also the case for 3a, 4b, 2b and 3b. The results of the X^2 test of stationarity for blocks of 4 trials are shown in Table 1 and confirm these graphic observations. Only the curve for 1b approached significance. (No computation was made for 3a because of the small number of errors; the number of subjects in the final block of 4 trials is shown in the right-hand column of the table.)

Table 1. Stationarity Results for Equivalence and Identity of Sets Experiment (Experiment II)

Group	X^2	df	P >	Ss in last block
1a & 4a	4.95	9	.80	9
1b	16.69	9	.05	12
2a	4.79	9	.80	11
3a	—— Too few errors ——			1
4b	4.89	9	.80	5
2b	5.96	9	.70	5
3b	3.49	9	.90	10

I shall restrict myself to one Vincent curve for this experiment. The 48 subjects of Groups 1 and 4 began with the concept of identity of sets. Of the 48 subjects, 38 met the criterion of 16 successive correct responses mentioned above. The Vincent curve for the criterion subjects showed evidence of non-stationarity in the fourth quartile, as in the case of Experiment I.

From examination of the mean learning curves over all trials, there seemed to be little evidence of transfer. And somewhat surprisingly, the only definite evidence confirmed the existence of negative transfer. The effects of transfer are actually more evident when we examine the data from the standpoint of 2 or 4 concepts. Here I shall try to summarize only what appear to be the most important points. For learning of identity of sets, prior training on equivalence has positive transfer for class $I\bar{O}$ (pairs

of sets identical only in the sense of unordered sets) and negative transfer for $E\bar{I}$ (pairs of sets that are equivalent but not identical). The qualitative explanation seems obvious. The initial natural dichotomy seems to be O, $\bar{O}$, and for this dichotomy $I\bar{O}$ is a class of "different" pairs, but the task of equivalence reinforces the treatment of $I\bar{O}$ pairs as the "same." The situation is reversed for the class $E\bar{I}$, and thus the negative transfer for under equivalence $E\bar{I}$ pairs are the "same," but under identity of sets they are "different."

When (3b) the task is again identity of sets but the prior training is on identity of ordered sets rather than equivalence, there is, as would be expected by the kind of argument just given, negative transfer for the class $I\bar{O}$. There is also some slight evidence of positive transfer for $E\bar{I}$.

From the data of Groups 2a and 1b, we inferred positive transfer for the class $I\bar{O}$ when the task is equivalence and the prior training is on identity of sets. What is surprising is the relatively slight amount of negative transfer for the class $E\bar{I}$.

Finally comparing Groups 3a and 4b, we observed negative transfer for the class $I\bar{O}$, as would be expected. The response curves for the other three classes are too close to probability one to make additional inferences, although there is a slight negative transfer for $E\bar{I}$ which cannot be explained on the principles stated above.

It seems apparent from these results that the analysis of transfer in the learning of mathematical concepts may often be facilitated if a fine-scale breakdown of the concepts in question into a number of subconcepts is possible. What is most needed is a quantitative theory to guide a more detailed analysis of the transfer phenomena.

Experiment III. Polygons and Angles

This experiment is reported in detail in Stoll (1962), and I present some of the data here with her permission. The subjects were 32 kindergarten children who were divided into two equal groups. For both groups the experiment was a successive discrimination, three-response situation, with one group discriminating between triangles, quadrilaterals and pentagons, and the other group discriminating between acute, right and obtuse angles. For all subjects a typical case of each form (*i.e.*, one of the three types of polygons or three types of angles) was shown immediately above the appropriate response key. As in the case of Experiment II, no single stimulus display was repeated for any one subject. Stimulus displays representing each form were randomized over experimental trials in blocks of nine, with three of each type appearing in each block of nine. The subjects were run to a criterion of nine successively correct responses, but with not more than 54 trials in any one session.

Table 2. Stationarity Order and Binomial Distribution Results (Stoll Experiment on Geometric Forms)

	X^2	df	P >
Quadrilateral, p = .609			
Stationarity (N = 273)	1.68	4	.70
Order (N = 262)	0.65	1	.40
Binomial distribution (N = 65)	1.77	2	.40
Pentagon, p = .600			
Stationarity (N = 275)	2.40	4	.60
Order (N = 269)	1.76	1	.15
Binomial distribution (N = 65)	2.07	2	.35
Acute angle, p = .674			
Stationarity (N = 338)	7.96	4	.05
Order (N = 348)	3.17	1	.05
Binomial distribution (N = 85)	2.66	2	.25
Right angle, p. = .506			
Stationarity (N = 313)	6.34	4	.10
Order (N = 326)	2.41	1	.10
Binomial distribution (N = 80)	10.52	2	.001
Obtuse angle, p = .721			
Stationarity (N = 268)	1.10	4	.85
Order (N = 256)	7.32	1	.001
Binomial distribution (N = 63)	2.90	2	.20
Quadrilateral and pentagon, p = .604			
Stationarity (N = 548)	0.71	4	.90
Binomial distribution (N = 130)	1.77	2	.40
All angles, p = .624			
Stationarity (N = 919)	0.97	4	.90

I omit data for the triangles because only a small number of errors were made.

The pentagons, quadrilaterals and right angles had quite stationary Vincent curves, whereas there was a definite increase in the fourth quartile of the Vincent curves for the acute and obtuse angles, and in fact in the case of the obtuse angles a significant increase in the third quartile. Statistical tests of stationarity of these Vincent curves support these results of visual inspection. Each test has three degrees of freedom because the analysis is based on the data for the four quartiles. In the case of the quadrilaterals $X^2 = 1.75$, for the pentagons $X^2 = 1.33$, for the right angles $X^2 = .95$, for the obtuse angles $X^2 = 12.63$, and for the acute angles $X^2 = 16.43$. Only the last two values are significant.

Using responses before the last error, for all concepts except that of a

triangle, goodness-of-fit tests were performed for (a) stationarity in blocks of 6 trials, (b) binomial distribution of responses as correct or incorrect in blocks of 4 trials, and (c) independence of responses — with the test being for zero-order *vs.* first-order dependence. The results of these facts are presented in Table 2 (page 109). The results shown strongly support the adequacy of the one-element model for this experiment.

Experiment IV. Variation in Method of Stimulus Display

In this study conducted with Rose Ginsberg we compared the rate of learning in two experimental situations, one in which stimulus displays were presented individually in the usual way and the other in which the same stimulus displays were presented by means of colored slides to groups of four children. The concept to be learned was identity of sets, and in both situations the children were required to respond by pressing one of two buttons, depending upon whether the stimulus display on that trial was identical or nonidentical. Of the 64 subjects 32 were from first grade and 32 from kindergarten classes. For the children receiving individual displays the experimental situation was essentially identical with that of Experiment II.

However, each group was divided into two subgroups. One subgroup received the stimulus material in random order, and the other in an order based on anticipated difficulty, in particular, with presentations of one-element sets coming first, then two-element sets and finally three-element sets.

Analysis of the mean learning curves for the two subgroups with random presentation indicated that presentations by slides are a less effective learning device for younger children, and the younger the child the more this finding seems to apply. At all levels of difficulty the kindergarten children learned more efficiently when the stimuli were presented to them in individual sessions. With one- or two-element sets displayed, Grade 1 subjects learned only slightly better in the individual session situation than in the slide situation, but when the task was more difficult (stimulus displays of three-element sets) the individual learning situation was clearly the most adequate. In interpreting these results it should be emphasized that the individual session was strictly experimental so that the amount of interaction between subject and experimenter was paralleled in both individual and slide situations.

Why these two experimental situations should produce different results in terms of efficiency of learning is not yet clear to us. One possibility is the following. It has been shown both with lower organisms (Murphy and Miller, 1955) and with young children (Murphy and Miller, 1959) that the ideal situation for learning is when the stimulus, response and re-

inforcement are contiguous. In the individual sessions these requirements were met, for the response buttons were 1.5 inches below the stimulus displays and the reinforcement lights were 1.0 inch from the stimuli. On the other hand, in the slide presentations, although the stimulus displays and reinforcements were immediately adjacent to each other, the response buttons were about 3 ft. from the screen on which the stimulus display was projected. Experimentally it has been shown (Murphy and Miller, 1959) that with children of this age group a separation of 6.0 inches is sufficient to interfere with efficient learning.

Experiment V. Incidental Learning

This experiment represents a joint study with Rose Ginsberg. Thirty-six kindergarten children in three groups of 12 each were run for 60 trials a day on two successive days of individual experimental sessions during which they were required to learn equivalence of sets. On the first day the stimulus displays presented to the subjects on each trial differed in color between the three groups but were otherwise the same. In Group 1 all displays were in one color, black. In Group 2 equivalence sets were presented in red and nonequivalence sets in yellow. For the first 12 trials in Group 3, equivalent sets were in red and nonequivalent sets in yellow; for the remaining 48 trials on that day the two colors were gradually fused until discrimination between them was not possible. On the second day all sets were presented to all three groups in one color, black.

As is apparent from the brief description of the experimental design, Group 1 simply had two days' practice under the same conditions with the concept of equivalence. In Group 2 the child did not actually need to learn the concept of equivalence but he could simply respond to the color difference on the first day, and it is well known that such a color discrimination for young children is a simple task. If the child in this group learned anything about equivalence of sets the first day, therefore, we may assume this to be a function of incidental learning. If incidental learning is effective, his performance on the second day when the color cue is dropped should at least be better than the performance of children in Group 1 on the first day. In Group 3 where we give the child the discriminative cue of color difference in the first trial and then very slowly withdraw that cue, the child should continue to search the stimulus displays very closely for a color stimulus, and is thus obliged to pay very close attention to the stimuli.

Analysis of the mean learning curves for the three groups indicated that of the three groups, only Group 2 approached perfect learning on the first day. In this group, of course, only color discrimination was necessary. Both the other groups did not improve over the first 60 trials, although

Group 3 showed some initial improvement over those trials when the color cues remained discriminable. On the second day Group 1 showed no improvement, and the learning curves for this group and for Group 2 are practically identical. For Group 3 on the other hand, the results were conspicuously better on the second day than for those of any other group. It was apparent from the learning curves that the task chosen was relatively difficult for the age of the children, because essentially no improvement was shown by Group 1 over the entire 120 trials. The conditions in Group 3 in which the children were forced to pay very close attention to the stimuli, do seem to have significantly enhanced the learning.

Experiment VI. Variation of Response Methods

This study was made jointly with Rose Ginsberg, and its object was to study the behavioral effects of different methods of response. Specifically, three groups, each composed of 20 kindergarten children, were taken individually through a sequence of 60 trials on each of two successive days for a total of 120 trials. The task for all three groups was equivalence of sets.

In Group 1 the child was presented with pictures of two sets of objects and indicated by pressing one of two buttons whether the sets "went together" or did not "go together" (were equivalent or nonequivalent).

In Group 2 the child was presented with one display set and two "answer" sets and was required to choose the answer which "went together" with the display set.

In Group 3 the child was presented with one display set and three "answer" sets and made his choice from the three possible answers.

This situation has fairly direct reference to teaching methodology in the sense that Group 2 and Group 3 represent multiple-choice possibilities. In Group 1, in which the child is required to identify either the presence of the concept or its absence on each trial, the situation is comparable to one in which the child is to indicate whether an equation or statement is correct or incorrect.

On the first day each group of children learned the task described above. On the second day they were run on an alternative method. Specifically, Group 1 was run under Group 3 conditions and Groups 2 and 3 were run under Group 1 conditions.

Analysis of the mean learning curves for all groups on both days indicated that in Group 2, where the subjects were required to choose from one or two available responses, the subjects learned slightly more quickly and to a slightly better level of achievement on Day 1 than the other groups, but, on the second day, when the experimental conditions

were shifted, Group 2 subjects did less well than the subjects in the other two groups. The clear superiority of Group 1 on the second day, when they were transferred to Group 3 conditions, indicated some positive transfer from learning to judge whether or not a concept is present to the multiple-choice situation, whereas the results for Groups 2 and 3 on the second day indicated some negative transfer from the multiple-response methods to the presence-or-absence method.

These results are further supported when we examine separately the data from subjects achieving a criterion of 12 successive correct responses on the first day. The more successful method was clearly that used in Group 1. The subjects in this group were conspicuously more successful than the other groups on the second day, making, in fact, no errors on that day from Trial 30 to Trial 60. Group 3 achieved perfect scores on the second day only on the last six trials, and Group 2 never reached that level on the second day, although they, like the other criterion subjects, had achieved perfect learning on the first day.

It seems reasonable tentatively to conclude that the method used with Group 1, in which subjects were required to recognize the presence or absence of some property on each trial, is the more successful method in establishing understanding of a concept well enough to permit transfer to a different response method.

Support for the all-or-none model of conditioning is also to be found in this experiment. In Table 3, X^2 goodness-of-fit tests of stationarity over trials before the final error for each group on each day are shown. The six values are all non-significant and thus support the basic assumption of the all-or-none models.

Table 3. Test for Stationarity over Trials before the Final Error (Experiment VI)

	Group 1		Group 2		Group 3	
	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
X^2	4.97	2.41	10.76	4.255	16.07	2.87
df	8	1	9	8	9	7
$P >$	.70	.10	.20	.80	.05	.80

On the basis of the six experiments just discussed I would like to draw some tentative conclusions, some of which are important for pedagogical procedures (*cf.* Suppes and Ginsberg, 1962a). I want to emphasize, however, that I would not wish to claim that the evidence from these experiments is conclusive enough to establish any one of the six conclusions in any final way, but what I do hope is that the attempt to summarize some of the implications of these experiments will stimulate other research workers to investigate these and related propositions in more adequate detail.

1. Formation of simple mathematical concepts in young children is approximately an all-or-none process. Evidence indicates, however, that significant deviations from the all-or-none model are present.

2. Learning is more efficient if the child who makes an error is required to make an overt correction response in the presence of the stimulus to be learned (Experiment I).

3. Incidental learning does not appear to be an effective method of acquisition for young children and in Experiment IV the group of children who responded to a color discrimination did not subsequently give any indication of having learned the underlying concepts.

4. Contiguity of response, stimulus and reinforcement enhances learning (Experiment V).

5. In the learning of related mathematical concepts, the amount of overall transfer from the learning of one concept to another is surprisingly small. However, considerable positive or negative transfer between specific sub-concepts is often present (Experiment II).

6. Transfer of a concept is more effective if the learning situation has required the subject to recognize the presence or absence of a concept in a number of stimulus displays than if learning has involved matching from a number of possible responses (Experiment VI).

Several of these conclusions are at variance with generally accepted results for adult learning behavior, for example, the efficacy of an immediate overt correction response (see Burke, Estes and Hellyer, 1954, for negative results on this method in adults), the variation of response method, or the relative specificity of learning of concepts with relatively little transfer. What is much needed is a wider range of systematic studies to isolate the factors of learning in young children that are particularly distinct from common features of adult learning behavior.

Sets and Numbers Pedagogical Program

All mathematics can be developed from the notions of sets and operations upon sets. The study of sets would therefore seem to provide the ideal initial introduction to mathematics but is ordinarily considered too "abstract" to introduce early in the student's development. The term "set" is in fact a simple concrete term—easy both to explain and understand. A set is simply any collection or family of objects. Thus, we may speak of the set or collection of all students now in the first grade, the set or collection of dolls now owned by Mary Jones, the set or collection of all Irishmen, the set or collection of all cowboys. The method of displaying problems by grouping objects as used in the standard workbooks approaches the set notion, but the lack of an explicit notation makes precise definitions difficult, if not impossible. Our empirical findings indicate that simple

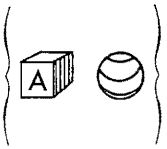
mathematical concepts can be understood by the beginner provided that these are presented precisely with the help of a consistent notation.

There are at least two major reasons to begin first grade arithmetic with the explicit introduction of the notion of set and appropriate notation for sets and operations upon them. In the first place, sets are more concrete objects than numbers. At the same time operations upon sets are more meaningful to the student than manipulation of numbers. The putting together of sets of physical objects, for example, is a more concrete operation than the addition of numbers. The many exercises in the grouping of objects displayed in current books is in fact a recognition of the greater concreteness of such objects.

In the second place, the prior introduction of sets and additional explicit notation permits mathematically exact and precise definitions and concepts rather than the often vague and ambiguous notions encountered in explanations of relations between concrete groups of objects and the Arabic numerals. For example, students can learn a clear, simple and meaningful characterization of numbers as properties of sets. Children who have learned the notion of a property can learn that a precise answer to the question, "What is a number?" is, "A number is a property of a set." Continual emphasis of this definition of numbers precludes learning number symbols as meaningless symbols or addition as a meaningless operation performed on meaningless number symbols. Numbers are properties of sets and the operation of addition of numbers is simply a general way of combining families of sets of things without paying any real attention to the things themselves.

With this definition *zero* presents no special difficulty and takes on no special mystique but is introduced as the numercial property of the empty set, the set which has no members. Empirical evidence strongly suggests that less difficulty is encountered with zero in arithmetical operations than with any other number when it is introduced as a property of the empty set.

The introduction of explicit notation makes the definition of number clear. Ordinarily, we can only assume understanding of the relationship when we make the great leap in abstraction from groups of objects to numerals which name their particular number properties. The use of set notation allows the steps in abstraction to be made explicit. The first step in abstraction is describing a set in the following way:

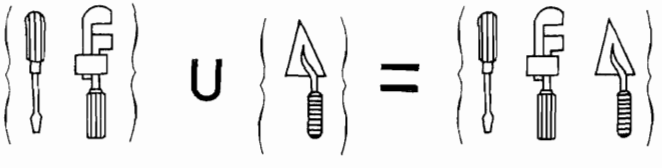


The next step in abstraction is the N notation:

This notation names a number but at the same time it maintains the pictorial character of the set description. In this sense it may be considered a "transition" to the Arabic numerals. Here we have abstracted from the particularity of the objects in the set to the single property of number. We must assume that children understand this step if they are to have any understanding of the way a number is related to a set of objects. The notation makes the step clear and precise in a way not permitted by verbalization at the primary grade level. The final step is to the Arabic numeral which in our example is:

2

Thus the explicit notation for sets introduces the student at the very beginning of his mathematical experience to the easily comprehended operations on sets—rather than to the more difficult and more abstract operations on numbers. Moreover the introduction of notation for sets permits consideration of addition and subtraction of numbers without commitment to the particular notation of Arabic numerals. The possibility of breaking free from the reiterated use of this single notation for numbers does much to eliminate the tendency to focus on the numerals themselves without consideration of their meaning. To illustrate these points about concreteness and meaningfulness the introduction of addition at three levels is illustrated as follows:



2 + 1 = 3

In the first line the concrete operation of putting sets together, that is, of forming their *union*, is indicated. In the second line the notation of N in front of a set is used to indicate the number of objects in the set. This notation permits a very simple transition from the concrete putting together of objects to the abstract use of Arabic numerals, introduced in the third line.

Throughout the beginning materials, set operations are presented as concrete analogues to numerical operations. Union of sets and addition of numbers are presented in sequence, difference of sets and subtraction of numbers are presented in sequence, etc.

Compared to most books, the present program introduces an extensive mathematical vocabulary and notation. It is believed that the use of explicit notation permits a clarity of definition of concepts that is otherwise impossible so that the student can develop a deeper understanding of number concepts. The essential nature of notation for precision in mathematics cannot be overlooked, for it is a truism of the history of mathematics that many strides of major importance have depended almost solely on the introduction of an appropriate mathematical notation. It is indeed difficult to penetrate far in mathematics without the development of an explicit notation.

Experimental work with the material during the academic years, 1959-60 and 1960-61, has shown that primary-grade youngsters do not have any difficulty in grasping new notation (Suppes and McKnight, 1961). As long as the notation introduced is explicit and precise and corresponds to simple notions, no difficulties of comprehension arise. The notation and vocabulary introduced in the program have already been tested and have been found to be well within the capacity of the primary-grade child.

The results of a general achievement test given to 25 experimental and 25 control first grade classes in May 1961 are shown in Table 4 (page 118). This achievement test was constructed by the project staff in order to provide a detailed profile of achievement levels on the standard arithmetical topics. It is important to emphasize, of course, that this test did not include any of the special notation introduced in the Sets and Numbers program, because the same test was administered to both experimental and control groups. This achievement test was used rather than one of the standard achievement tests like the Metropolitan because we wanted to get a more detailed picture of achievement levels.

Similar results were obtained in the spring of 1962 with the arithmetic portions of the Metropolitan Achievement Tests, but I shall not cite the numerical data here.

Finally let me state that we have also conducted an experimental pedagogical program in the teaching of mathematical logic to selected

fifth- and sixth-grade students. The project was conducted as a demonstration project under the sponsorship of the Office of Education. The purpose of the project was to introduce the academically able elementary school child to modern mathematics and mathematical methods at a level which is completely rigorous and simple enough in presentation and context to permit relatively easy comprehension. We are particularly interested in the capacity of children in the fifth- and sixth-grade age levels to

Table 4. Sets and Numbers Project

Results of General Achievement Tests, First Grade Classes, May 1961*

Possible Total Score 118

	N of Students	N of Classes	Mean	Median
Experimental Group	595	25	97.39	104
Control Group	539	25	86.68	90

Difference of means Significant $P < .01$

Comparison of Groups on Item Type

	Items	Proportion of Correct Answers to Total Possible Per Type of Item: Experimental Group	Control Group
1-4	Recognition of groups	.99	.99
5-6	Recognition of numerals	.99	.99
7-10	Writing numerals	.98	.97
11-22	Sequence of numerals	.95	.93
23-25	Telling time	.97	.97
26	Fractional part—half	.93	.92
27-29	Ordinals	.94	.94
30-35	Decimal system of numeration	.99	.97
36-43	Column addition	.94	.90
44-51	Column subtraction	.86	.79
52-54	Addition—2 digit numbers	.69	.50
55-59	Addition of 3 numbers	.65	.30
60-67	Column addition and subtraction	.67	.55
68-79	Addition and subtraction—equations	.85	.55
80-95	Equations—fill in blanks	.80	.75
96-105	Complex equations	.61	.41
106-112	Multiplication—additive	.72	.65
113-118	Operation symbols	.84	.77

* Tests were administered by classroom teachers in all classes, experimental and control.

do the kinds of deductive proofs which are characteristic of modern mathematics and to examine the possible transfer of skills of analysis and correct reasoning to other subject matter areas. In connection with this pedagogical program we have also conducted a number of experiments on the grasp of principles of logical inference and the learning of the beginning elements of mathematical logic from a learning-theoretic standpoint, but I do not have the space nor the time here to report in detail on these experiments. We hope to be able to make such reports generally available in the future (for a preliminary report, see Suppes, 1962).

References

C. J. Burke, W. K. Estes and S. Hellyer. "Rate of Verbal Conditioning in Relation to Stimulus Variability." *Journal of Experimental Psychology* 48: 153-61; 1954.

C. L. Hull and K. W. Spence. " 'Correction' vs. 'Non-correction' Method of Trial-and-Error Learning in Rats." *Journal of Comparative Psychology* 25: 127-45; 1938.

J. V. Murphy and R. E. Miller. "The Effect of Spatial Contiguity of Cue and Reward in the Object-Quality Learning of Rhesus Monkeys." *Journal of Comparative Physiological Psychology* 48: 221-29; 1955.

J. V. Murphy and R. E. Miller. "Spatial Contiguity of Cue, Reward and Response in Discrimination Learning by Children." *Journal of Experimental Psychology* 58: 485-89; 1959.

E. Stoll. "Geometric Concept Formation in Kindergarten Children." Ph.D. dissertation, unpublished, Stanford University, 1962.

P. Suppes. "Mathematical Logic for the Schools." *The Arithmetic Teacher* 9: 396-99; 1962.

P. Suppes and R. Ginsberg. "Experimental Studies of Mathematical Concept Formation in Young Children." *Science Education* 46: 230-40; 1962a.

P. Suppes and R. Ginsberg. "Application of a Stimulus Sampling Model to Children's Concept Formation With and Without an Overt Correction Response." *Journal of Experimental Psychology* 63: 330-36; 1962b.

P. Suppes and R. Ginsberg. "A Fundamental Property of All-or-None Models, Binomial Distribution of Responses Prior to Conditioning, with Application to Concept Formation in Children." *Psychological Review* 70: 139-61; 1963.

P. Suppes and Blair A. McKnight. "Sets and Numbers in Grade One, 1959-60." *The Arithmetic Teacher* 8: 287-90; 1961.