DOCUMENT RESUME

ED 040 878 | 24 | SE 008 994

AUTHOR Frayer, Dorothy Ann
TITLE Effects of Number of Instances and Emphasis of Relevant Attribute Values on Mastery of Geometric Concepts by Fourth- and Sixth-Grade Children (Parts 1 and 2).
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
REPORT NO TR-116
BUREAU NO BR-5-0216
PUB DATE Mar 70
CONTRACT OEC-5-10-154
NOTE 125p:

EDRS PRICE EDRS Price MF-$0.50 HC-$6.35
DESCRIPTORS *Elementary School Mathematics, *Geometric Concepts, Grade 4, Grade 6, *Instruction, *Learning, Mathematical Concepts, *Mathematics Education, Research

ABSTRACT

Reported are the results of study (approximately 20 minutes a day for four days) by fourth- and sixth-grade children of programed lessons dealing with geometric concepts. Independent variables of number of instances (4 or 8) and emphasis of relevant attitude values (presence or absence of attention directing and review questions) were varied systematically. Questions were given to test recognition and production of attribute examples, attribute names, concept examples, and non-examples, concept definitions, and relationships of the concepts to one another. The results showed that: (1) increasing the number of instances significantly improved recognition of concept-non-examples for fourth grade children, and (2) emphasis of relevant attribute values significantly increased recognition of attribute names for fourth grade and sixth grade children, production of attribute names for fourth grade children, and overall concept mastery for fourth grade children. (Author/RS)

3

2

OE/BR
50216
PA24

·ort No. 116

EFFECTS OF NUMBER OF INSTANCES AND EMPHASIS OF
LEVANT ATTRIBUTE VALUES ON MASTERY OF GEOMETRIC CONCEF
OURTH- AND SIXTH-GRADE CHILDREN (Part I) (Chapters I

Report from the Project on
Situational Variables and
Efficiency of Concept Learning

.Office of Education Center No. C-03 Contract OE 5-10-154

Technical Report No. 116

EFFECTS OF NUMBER OF INSTANCES AND EMPHASIS OF RELEVANT ATTRIBUTE VALUES ON MASTERY OF GEOMETRIC CONCEPTS BY FOURTH- AND SIXTH-GRADE CHILDREN (Part I) (Chapters I - IV)

Report from the Project on
Situational Variables and
Efficiency of Concept Learning

By Dorothy Ann Frayer

Herbert J. Klausmeier, Professor of Educational Psychology
Chairman of the Examining Committee and Principal Investigator

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

March 1970

This Technical Report is a doctoral dissertation reporting research supported by the Wisconsin Research and Development Center for Cognitive Learning. Since it has been approved by a University Examining Committee, it has not been reviewed by the Center. It is published by the Center as a record of some of the Center's activities and as a service to the student. The bound original is in The University of Wisconsin Memorial Library.

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

STATEMENT OF FOCUS

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. It includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and that they are applied to the improvement of educational practice.

This Technical Report is from the Situational Variables and Efficiency of Concept Learning Project in Program 1. General objectives of the Program are to generate new knowledge about concept learning and cognitive skills, to synthesize existing knowledge, and to develop educational materials suggested by the prior activities. Contributing to these Program objectives, the Concept Learning Project has the following five objectives: to identify the conditions that facilitate concept learning in the school setting and to describe their management, to develop and validate a schema for evaluating the student's level of concept understanding, to develop and validate a model of cognitive processes in concept learning, to generate knowledge concerning the semantic components of concept learning, and to identify conditions associated with motivation for school learning and to describe their management.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

Table		Page

LIST OF FIGURES

ACKNOWLEDGMENTS

The author wishes to express her sincere appreciation to Dr. Herbert J. Klausmeier who served as mentor throughout her doctoral program and supervised the preparation of this thesis. Appreciation is also extended to Dr. Robert E. Davidson and Dr. Gary A. Davis who provided thoughtful criticism as members of the reading committee. Dr. T. Anne Cleary offered generous assistance in the measurement aspects of the thesis and Dr. Thomas A. Romberg in the mathematical aspects. Dr. Cleary and Dr. Romberg also served on the examining committee.

The author is indebted to the Wisconsin Research and Development Center for Cognitive Learning for the opportunity to utilize the Center's resources and facilities during the preparation of this thesis. Special gratitude is extended to the following Center staff members: Barbara E. Sterrett who assisted in test construction, administration of the pilot study, and lesson scoring; Thomas J. Fischbach who provided competent guidance in the design and analysis of the experiment; Dr. John G. Harvey, James S. Braswell, and Larry K. Sowder who critically reviewed the lessons; Peter A. Lamal, Kenneth J. Simpson, Shelby J. Johnson, and Carolyn J. Gornowicz who helped administer the lessons and tests.

The author is also indebted to the personnel of the Menomonee Falls Public Schools in which the studies were carried out. Special thanks are due Mr. Winston Brown, Assistant Superintendent; Mr. August Schreiner, Principal of Shady Lane Elementary School; Mr. Clifton Scheffler, Principal of Benjamin Franklin Primary School; and Mr. William Evans, Principal of Thomas Jefferson Middle School.

ABSTRACT

The object of this study was to determine the effect of the number of instances and the emphasis of relevant attribute values on the level of concept mastery.

Eight versions of programed lessons dealing with geometric concepts were prepared in which the independent variables of number of instances (4 or 8) and emphasis of relevant attribute values (presence or absence of attention-directing and review questions) were varied systematically. The subjects, who were fourth- and sixth-grade children, studied these lessons approximately 20 minutes a day for four days.

After study of the lessons, the children were given a multiple-choice test and a completion test, each consisting of eleven types of questions related to concept learning. These questions were formulated to test recognition and production of attribute examples, attribute names, concept examples and non-examples, concept names, relevant and irrelevant attributes, concept definitions, and relationships of the concepts to one another.

The essential findings of the study were:

(1) Increasing the number of instances from 4 to 8 did not significantly affect overall concept mastery for either fourth- or sixth-grade children.

(2) Increasing the number of instances significantly improved recognition of concept non-examples for fourth-grade children.

(3) Emphasis of relevant attribute values significantly increased overall concept mastery for fourth-grade children. The increase in overall concept mastery for sixth-grade children was not significant.

(4) Emphasis of relevant attribute values significantly increased recognition and production of attribute names for fourth-grade children and recognition of attribute names for sixth-grade children.

Chapter I

INTRODUCTION

Concept learning is a topic of great theoretical interest and practical importance. Glaser (1968) has suggested that concepts are based on simple behavioral acts and form the elements of higher-order behaviors. This position of intermediate complexity provides an ideal situation to study the interplay of various elemental processes in the accomplishment of learning outcomes. Laboratory studies by many persons affiliated with the Wisconsin Research and Development Center for Cognitive Learning have contributed to a better understanding of the stimulus variables and cognitive operations related to concept learning (Fredrick, 1966, 1968; Jones, 1968; Klausmeier, Harris, Davis, Schwenn, & Frayer, 1968; Lemke, Klausmeier, & Harris, 1967; Lynch, 1966; Ramsay, 1965; Smuckler, 1966). Further, concepts comprise a large and important segment of the subject matter taught in the schools. Guidelines for effective concept teaching can therefore make a significant contribution to improved learning in a wide range of situations. For this reason, personnel at the Center have also extended concept learning principles through research in school settings (Blount, Klausmeier, Johnson, Fredrick, & Ramsay, 1967; Kalish, 1966; Kennedy, 1968; Steffe, 1966; Steffe & Parr, 1968).

A sizeable body of knowledge has accrued concerning laboratory concept learning. This research has made notable contributions to psychological theory and has indicated that certain instructional variables have powerful effects on concept attainment. Laboratory research, however, has been restricted in certain respects. This has limited its potential contribution to both theory and practice.

Laboratory experiments have generally utilized specially-chosen concepts embodying only a few of the aspects which may influence learning of concepts encountered in daily life. Most concepts studied in the laboratory are comprised of characteristics already known to the learner. Thus, the effect of variables influencing the learning of those characteristics is minimized. Laboratory concepts usually do not have meaningful concept labels, even though associations between such labels may have a powerful effect on transfer and interference in learning. Also, a large amount of response learning may be required to acquire the labels. In order to examine aspects of concept learning such as these, a wider range of concepts should be employed. To permit meaningful comparisons, the concepts should be analyzed to determine their relevant characteristics and how these characteristics are combined, their irrelevant characteristics, the associative structure of the concept label, and the relationship of the concept to other concepts. In this manner, many kinds of concepts could be studied but their similarities and differences specified.

A second way in which concept research has been restricted is in techniques of instruction. Typically, only concept examples and non-examples have been presented, the sequence of presentation being determined by either the experimenter or the subject. Few studies have dealt with concept learning by definition, synonyms, or sentence context. Additional consideration should also be given to the effect of advance organizers, questions, and review.

Measurement of concept learning has also focused on only a small segment of behaviors by which learning might be inferred. Trials, errors, or time required to reach a criterion of successive correct identifications have been commonly used measures. These techniques, however, do not assess the verbal aspects of concept learning such as labeling or definition. Nor do they test knowledge of relationships between concepts. Yet much of the power and utility of concept learning derives from the possibility of communicating by naming or describing and from relating the concept to other concepts. A set of behaviors which would permit inference of these various aspects of concept attainment would, therefore, be more informative than a single measure. It is probable that a particular instructional technique may facilitate some aspects of concept learning more than others. Also, past research has for the most part measured rate of acquisition rather than long-term retention. Both should be assessed.

A final limitation of past concept learning research has been the preponderance of studies employing young adults as subjects. In order to discover possible developmental trends in concept learning, experiments should be replicated at various age levels.

In summary, concept learning research should be extended in several ways: (1) a wider range of concepts should be examined, with careful specification of the essential characteristics of the concepts; (2) various instructional procedures should be utilized, including verbal as well as nonverbal strategies; (3) a set of differentiated response measures should be employed to assess both short-term and long-term retention; and (4) performance of subjects of different ages on the same task should be compared. Simultaneous extension of these four aspects of concept learning research may reveal the cognitive processes entailed in concept learning and permit description of the interactive effect of concept type, instructional procedure, and the age of the subject in determining degree of learning.

The present study attempted to implement the extension of concept learning research in each of the suggested directions.

Age of subjects. Subjects were fourth- and sixth-grade children.

Nature of the concepts. Concepts taught were geometric forms which had names, some of which were unfamiliar to the children who served as subjects. Further, some degree of attribute learning was required for the younger children. The concepts studied bore complex interrelationships to one another. A strategy for characterizing the concepts was developed, which consisted of determining the attributes relevant and irrelevant to the concept, and of determining the relationship of each concept to the others.

Measurement of learning. On the basis of logical analysis of the nature of a concept, a review of the literature on cognitive processes in concept learning, and a review of previously employed concept learning measures, eleven tasks were identified which would permit inference of various aspects of concept learning. These tasks were incorporated into test items to measure attainment of each concept.

Instructional procedures. Concepts were taught by a combination of definitions and examples. Lessons varied in the number of examples presented and the relative emphasis of relevant attribute values.

The importance of providing a variety of examples has been stressed by educational psychologists: "the defining attributes of a concept are learned most readily when the concept is encountered in a large number of different contexts . . . . By de-emphasizing the particularity of single or homogeneous instances, multicontextual learning facilitates the abstraction of commonality, strengthens the generality and transferability of the resulting concept, and endows it with greater stability [Ausubel, 1968, p. 531]." "A grasp of abstract concepts, of course, must come from experience with many objects. To understand squareness, for instance, the student must see . . . a variety of concrete objects, all having the common property of squareness [Stephens, 1956, p. 369]."

Despite the seemingly obvious truth of the notion that a wide variety of instances facilitates concept learning, the research evidence concerning the effect has been contradictory. Callantine and Warren (1955) and Morrisett and Hovland (1959) found better transfer to new concept instances when a wide variety of examples were used during training. Podell and Carter (1963) found that a large variety of instances promoted concept acquisition and resulted in greater generalization, especially when the concept was relatively difficult.

The results of Gagné and Bassler (1963) indicated significantly lower retention for students given only a minimal variety of examples.

On the other hand, under some circumstances small variety has been associated with improved acquisition (Amster, 1966) and generalization (Amster & Marascuilo, 1965). Remstad (1969) noted little effect due to number of instances. A significant interaction between number of instances and specific concept was observed by Shore and Sechrest (1961).

Thus, the effect of number of instances has not been clearly established. Determination of the particular aspects of concept learning affected by this variable may provide clarification.

Emphasis of relevant attribute values has also been suggested as an effective technique of increasing concept attainment (Klausmeier & Goodwin, 1966; Ausubel, 1968). The research evidence clearly supports this principle. Inducing a set to respond to the relevant dimension of a concept by pretraining with words describing that dimension facilitated learning performance (Gelfand, 1958). Rasmussen and Archer (1961) also noted improved learning when the relevant dimension had been labeled during pretraining. Wittrock, Keislar, and Stern (1964) discovered that children who were given the label for the relevant characteristic of a concept had significantly higher scores than children given more general hints or specific names of instances. Addition of a single-word verbal cue drawing attention to the common attribute of concept instances greatly increased transfer (Remstad, 1969).

Thus, the emphasis of relevant attribute values has consistently been shown to improve concept learning. It should be noted, however, that in all cases, the concept tasks were inductive in nature. An important question is whether this same facilitation would occur if the subject were given a definition of the concept, making the task a deductive one. Further, it would be of interest to determine what aspects of concept learning performance were most affected by this variable.

Purposes and Hypotheses of the Study

The first objective of this study was to devise a set of prototypic tasks which would test various aspects of concept learning. The second objective was to determine the effect of two instructional variables on performance of these tasks.

Two hypotheses were tested. The first was that level of concept mastery would increase as a function of the number of instances presented. Thus, it was predicted that children who had seen eight examples of a concept would correctly answer more questions concerning that concept than children who had seen only four instances. The second hypothesis was that emphasis of relevant attribute values would facilitate concept learning.

Method

Subjects were 154 fourth-grade and 126 sixth-grade children. Eight versions of programed lessons dealing with geometric concepts were prepared to vary systematically the independent variables of number of instances and emphasis of relevant attribute values. The children studied these lessons approximately 20 minutes a day for four days.

After completion of the lessons, children were given a multiple-choice test and a completion test, each consisting of eleven types of questions related to concept learning.

Multivariate analyses of covariance were carried out on total test scores and on scores for each type of question in order to determine the effects of number of instances and emphasis of relevant attribute values.

Significance of the Study

The method of analyzing concepts developed for this study may be applied to various subject-matter areas. Such a detailed description of the concepts being learned would permit some control of the nature of the concept as a source of variability. The technique of measuring concept learning devised for the study provides exact description of the responses from which concept learning may be inferred.

Specification of the nature of the concept and of the response has great utility for both research on instructional variables and theory of concept learning. The effect of instructional variations may be related to the type of concept and response being considered. The differentiation of tasks may clarify the cognitive processes entailed in concept learning.

Chapter II

REVIEW OF RELATED LITERATURE

To extend our knowledge about concept learning, a wider range of concepts should be studied. A system for characterizing the essential aspects of subject-matter concepts will be described in the first section of this chapter. Utilizing this system, diverse concepts may be classified according to their similarities and differences. In the second section of the chapter, a strategy for testing concept mastery will be derived, based on a review of the literature dealing with cognitive processes and testing procedures. Finally, research related to the effects of increasing the number of concept examples and of emphasizing relevant attribute values will be discussed.

Analysis of Subject-Matter Concepts

From its earliest beginnings, experimental study of concept learning focused on concepts having clearly specifiable characteristics. This trend was crystallized by the publication of an influential article by Hovland (1952) which related the amount of information conveyed by a concept instance to the number of relevant dimensions and values on those dimensions. In order to utilize the Hovland informational analysis, the exact number of relevant dimensions and values must be determined. This requirement has tended to restrict experimental material to geometric figures or artificially devised stimulus populations. These materials are intended to be analogues of naturally occurring concepts. Indeed, Bruner, Goodnow, and Austin (1956) used numerous "real-world" examples to demonstrate that concepts may be described in terms of criterial attributes and conceptual rules. The learning of these "real-world" concepts, however, has rarely been studied experimentally. The paucity of such experimentation may be due to the fact that the exact number of relevant attributes and attribute values is sometimes indeterminate in naturally occurring concepts. In general, however, some set of characteristics common to all examples of a concept may be specified, although the set may not be exhaustive.

It would be desirable to study naturally occurring concepts in order to assess the effects of meaningful labels, mode of representation, and relationship to other concepts. A report by Berzonsky (1968) indicated that analysis of biological concepts in terms of defining attributes could be carried out. Due to the nature of the concepts, experimental results based on this analysis had direct implications for both psychology and pedagogy.

A concept may be described in terms of six characteristics: (1) relevant attribute values (properties which are common to all instances of the concept); (2) concept definition (comprised of the relevant attribute values of the concept and the rule combining those values); (3) irrelevant attributes (properties which vary from instance to instance); (4) concept label; (5) supraordinate concepts (more generic concepts having some but not all of the relevant attribute values of the given concept); and (6) subordinate concepts (more specific concepts having all of the relevant attribute values of the given concept and other in addition).

Although other characteristics might be specified, this set provides a basis for comparing concepts which differ in complexity, rule difficulty, type of label, and degree of interrelatedness.

Measurement of Concept Mastery

Johnson and O'Reilly (1964) have posited that a concept is a single hypothetical construct with specifiable relations to different observable measures. Although a concept is seen as a single entity, Johnson (1964) has suggested that two or more tests be made in order to provide convergent evidence for its existence and characteristics. Differences which occur between two tests are ascribed by Johnson and O'Reilly (1964) to variability in response difficulty, response practice, or scoring. The agreement of two tests, however, is not a necessary consequence of viewing a concept as a single hypothetical construct. The tests may provide evidence concerning different aspects of the concept. Thus, for example, if the subject had cognized the relevant attribute values of a concept but had not cognized the label for them, he would be able to classify concept examples but unable to verbalize the concept definition. Indeed, a discrepancy between the ability to classify and define has been frequently noted (Phelan, 1968; Deno, 1968; Remstad, 1969).

Although a concept held by a given individual at a given time is a single entity, the characteristics of the concept may differ (1) among individuals, and (2) in the same individual over time (Klausmeier, Sterrett, Frayer, Lewis, Lee, & Bavry, 1969). Therefore, our goal should be to characterize the concept held by an individual with sufficient precision to detect these differences.

Several authors have suggested that concept learning is the result of complex interplay between elementary cognitive processes. Gagné (1968) views classification as an intellectual skill. It is one member of a hierarchical set of intellectual skills in which subordinate skills provide positive transfer to the learning of superordinate skills. Gagné cites research, for example, which indicates that dimensional discrimination training facilitates classification. Other skills which may transfer to concept learning include signal learning, stimulus-response learning, learning of verbal associates, and multiple discrimination (Gagné, 1965).

Guilford (1967) has proposed a three-way classification of intellectual abilities, the structure-of-intellect (SI) model. The dimensions of the model specify the operation, content, and product of a given intellectual act. Each factor in the model is defined by specifying a level on each of the three dimensions. The processes relevant to concept learning may be clarified by identifying them with tests of the abilities postulated in Guilford's model. "In complex learning situations such as a concept-learning task, S does not sit passively, learning only at the whim of the experimenter's manipulation of conditions . . . S recognizes attributes common to the stimuli, he produces and tests hypotheses concerning which attributes are relevant, and he remembers what occurred on previous trials. The last statement implies all five of the operations of the SI model. The potential for the understanding of concept learning is in the investigation of these processes that S performs between receiving of the stimulus and the production of the overt response [Dunham, Guilford, & Hoepfner, 1966, p. 4]."

Support for the contention that various processes come into play during concept formation derives from results indicating that different abilities are relevant to different types of concept problems, that relevant abilities vary over stages of practice, and differ for solvers and non-solvers (Dunham, Guilford, & Hoepfner, 1966; Jones, 1968; Manley, 1965).

Still another theory of the development of classificatory concepts has been proposed by Inhelder and Piaget (1959). Seven of the developmental steps identified were: (1) grouping of two objects on the basis of resemblance, (2) grouping of more than two objects, (3) grouping of all objects which are in some way alike, (4) grouping regardless of the physical proximity of the objects, (5) categorizing the same object into more than one group, (6) grouping objects in different ways, and (7) forming classes that stand in an inclusive relationship to one another. Kofsky (1966) administered tasks designed to measure these developmental steps in children 4 - 9 years of age, and found that there was a significant correlation between age and number of task successes. Further, the tasks exhibited the predicted order of difficulty.

Klausmeier (Klausmeier, Harris, Davis, Schwenn, & Frayer, 1968) has identified three sets of process configurations involved in concept learning: (1) analyzing the situation, (2) securing information, and (3) processing the information. The information processing steps hypothesized by Klausmeier were as follows:

> Sensing external and internal stimulus situations
> Acquiring or manifesting non-labeling responses
> Associating responses with stimuli along physical dimensions
> Associating sequential S-R associations
> Acquiring labels and associating proper labels with each stimulus situation
> Discriminating among many stimuli and responding appropriately
> Cognizing common properties of at least two instances and responding with the appropriate single label
> Discriminating the relevant from the irrelevant characteristics of the stimulus situation
> Cognizing that two or three instances do or do not belong to the same set
> Cognizing a relevant attribute or rule by comparing the information presented in positive and negative instances
> Hypothesizing relevant attributes
> Remembering attributes, rules, and hypotheses
> Evaluating relations among attributes or values and rules in terms of an hypothesis
> Inferring the concept by inductively arriving at the common defining properties and rules; by cognizing logical relations among properties and rules [Klausmeier, Harris, Davis, Schwenn, & Frayer, 1968, p. 7].

Drawing upon the work of Gagné, Guilford, Piaget, and Klausmeier, the author has postulated the following processes as related to concept learning: (1) cognition of the attribute values of concept instances, (2) association of attribute values with their labels, (3) cognition of an instance as an example or non-example of the concept, (4) association of a concept instance with its label, (5) cognition of the characteristics common to all concept examples. (6) cognition of the conceptual rule relating the common characteristics, and (7) cognition of the relationship between the concept and concepts subordinate and supraordinate to it. Tests by which the occurrence of these processes may be inferred have been developed for use in the present study and will be detailed in Chapter III.

Development of tests related to each of these aspects of concept learning may permit detection of specific instructional effects. It is probable, of course, that there will be a high correlation between measures. Therefore, a strong differential effect would be necessary for differences to be apparent. Johnson and Stratton (1966), for example, failed to detect specific relationships between training method (classifying examples, defining the concept, using it in a sentence, and giving synonyms), and tests of comparable behaviors. On the other hand, Nuthall (1968) found that comparison of concepts during instruction had a detrimental effect on the recognition of positive instances and identification of the relationship between the concept taught and other concepts. The recognition of negative instances, however, was facilitated. Thus, it appears that differential effects of instruction may in fact occur.

Number of Instances

One of the variables examined in the present study was number of instances. Since no instances were repeated, an increase in number of instances implied an increase in variety of instances.

Several of the experiments dealing with number of instances have related to the formation of a "learning set," rapid solution of a new problem of a specified type following practice on a large number of problems having the same general solution. In these studies the problems which are presented vary in attributes but have a common rule (often relational) joining the attributes. The goal, then, is rule learning.

Adams (1954) presented examples of four concepts consisting of horizontal or vertical arrangements of two different attributes. One group of subjects received 24 different examples of each concept while another group received repeated presentations of a single example of each concept. Both groups then received three presentations of a transfer problem. The group trained with a single example of each concept was clearly superior on the transfer problem.

In contrast to Adams' finding that single-example training led to more positive transfer, Callantine and Warren (1955) found that multiple-example training produced greater transfer. Six groups were compared, having 20, 10, 4, 2, 1, or 0 (control group) different examples per concept. During training, the group having only one example per concept committed the fewest errors. On the transfer task, however, the group having 20 examples per concept was significantly superior. Callantine and Warren suggest that due to rapid pacing, a non-correction procedure, and difficult stimulus materials, Adams' subjects who were trained on multiple examples never mastered the concepts prior to the transfer task.

Replicating Adams' procedure, Morrisett and Hovland (1959) compared performance of subjects having 1, 3, or 24 examples per concept. Subjects having three examples performed at the highest level on the transfer test. Results support Callantine and Warren's interpretation that the poor performance of Adams' multiple-example group was due to low level of learning of each concept. Greater transfer, then, is associated with learning the conceptual rule and generalizing it to a variety of contexts. In the present study, where the conceptual rule is a simple conjunction, one would expect that generalization would be facilitated by increasing the variety of instances.

Although the task employed was not a typical concept learning task, results of an experiment by Gagné and Bassler (1963) indicate greater retention due to variety of task examples. Gagné and Bassler taught concepts in nonmetric geometry to sixth-grade children by means of a self-instructional program. The concepts taught were related to each other in such a fashion that knowledge of subordinate concepts facilitated acquisition of higher-level concepts. One of the variables examined in the study was variety of examples presented for each subordinate concept. No differences were observed in retention of the subordinate concepts over a nine-week period due to variety of examples. Retention of the highest-level concept was significantly inferior for those students who had a minimal variety of examples of the subordinate concepts.

In the present study, variety of attribute examples and variety of concept examples were inversely related. Students who had a larger variety of attribute examples had a smaller variety of concept examples. Since attribute examples are analogous to the subordinate task examples in Gagné and Bassler's study, a wide variety of attribute examples would be expected to facilitate retention of concepts. The net effect of increasing the variety of concept examples, therefore, would depend on the relative importance of variety of attribute examples and variety of concept examples.

Stern (1965) predicted that training with a large number of instances of a few concepts would result in greater transfer to new instances of those concepts. On the other hand, training with a large number of concepts would result in greater transfer to new concepts. To test these predictions she presented kindergarten and first-grade children with 12 instances of two concepts, six instances of four concepts, or three instances of eight concepts. Contrary to prediction, children trained with six instances of four concepts (intermediate variety of both instances and concepts) showed greatest transfer to both new instances and new concepts.

A series of experiments have been carried out by Amster [Podell] investigating the interaction of variety of instances with age and cognitive set. The hypothesis underlying these studies is that under an intentional set to learn, deductive strategies will predominate. In these circumstances, a large variety of examples would facilitate learning by permitting false hypotheses to be rejected in fewer trials than a small variety. Under an incidental set to learn, no difference due to variety is predicted since a large variety might permit more efficient summation of correct associations, but a small variety would permit better recall of instances. Under intentional set, older children are expected to benefit from a large variety of instances. For younger children, no corresponding facilitation is expected since the deductive reasoning process is not well-developed (Amster, 1966).

Although Amster has studied the interactions of variety, age, and set, only the effect of variety under intentional set for older children and adults will be summarized. Podell (1958) presented college students with examples of a figural concept having six relevant attribute values. Half of the subjects received 12 examples (large variety), the other half two examples (small variety). Under intentional set, the subjects who had seen 12 examples were able to identify significantly more relevant attribute values than subjects who had seen only two examples. On the other hand, subjects receiving the small variety recalled significantly more irrelevant features.

Podell and Carter (1963) taught verbal concepts to fourth- and fifth-grade children by presenting sentences in which the word whose meaning was to be learned was replaced by a nonsense word. Four concepts were taught: two nouns intended to be easy, and two verbs intended to be difficult. In the large variety condition, each concept was presented in the context of six different sentences. In the small variety condition, each concept was presented in three different sentences. For the easy concepts, no significant effects were observed as a function of variety. For difficult concepts, however, subjects who had received the large variety of instances were better able to define the concept.

The two studies just mentioned found differences in favor of a large variety of instances. Later experimentation (Amster & Marascuilo, 1965; Amster, 1966) has failed to confirm this advantage. Amster and Marascuilo (1965) taught the concept of set union and set intersection to fourth-grade children using either a large or small variety of figural instances as examples. No significant differences were noted in concept learning due to variety of instances. On a generalization test which employed words or letters as instances of the concepts, subjects who learned the concepts with a small variety of instances performed significantly better. Amster and Marascuilo interpret the greater generalization to be indicative of a higher degree of learning for subjects receiving a small variety of instances. No effect of variety was observed by Amster (1966) when fourth- and fifth-grade children were taught concepts based on complex aggregates of features contained in flags. The dependent variable in this case was accuracy of constructing a new flag to exemplify the concept.

The effect of number of instances may also differ depending on the specific concept being learned. Shore and Sechrest (1961) presented college students with 3, 9, or 18 different instances for each of four concepts based on common characteristics of nonsense syllables. The concepts were chosen to represent four levels of difficulty based on results of a preliminary study. With recognition of new instances as the dependent variable, a significant interaction was obtained. When concepts were based on obvious characteristics, a wide variety of instances resulted in better performance. When concepts were based on less obvious characteristics or required mental transformation of stimuli, a small variety of instances was preferable.

In a study which applied response surface methodology to the optimization of concept learning performance, Remstad (1969) carried out a sequence of experiments varying a combination of independent variables. One of these independent variables was number of instances, 6 or 9. Although response surface methodology does not test hypotheses concerning the effect of individual variables, little change in response was noted when the number of instances was increased from 6 to 9.

In summary, increasing the number of concept instances has not always resulted in improved concept learning performance. There is evidence, however, that greater rule learning occurs with a greater number of instances. Definition and verbalization of relevant attribute values may be higher following training with a wide variety of instances. In addition, presenting more instances may increase performance when the common characteristic is an obvious one.

Emphasis of Relevant Attribute Values

All instances of a concept share certain common properties. These common properties may be called the relevant attribute values of the concept. Instances of a particular concept also differ from one another in certain ways. These properties which differ from instance to instance are not relevant to determining the concept and are therefore called irrelevant attributes.

It is evident that any factor which directs attention to the common characteristics of the instances may be expected to speed concept learning and increase reliability of instance identification. Factors which may increase the noticeability of relevant attribute values have been summarized by Trabasso and Bower (1968). These factors have been classified as: (1) innate (differential sensitivity of the subject to cues); (2) stimulus-bound (perceptual arrangements emphasizing cues); and (3) past-training (discriminative history and verbal instructions). In the present study, verbal cues were employed which focused attention on the relevant attributes. This would be classified by Trabasso and Bower as a past-training factor.

Verbal cues have been shown consistently to facilitate concept learning. Gelfand (1958) taught different word lists to three groups of college students prior to a concept identification task. The lists contained words describing relevant dimensions of the concept, words describing irrelevant dimensions, or words unrelated to the concept. The concept learning task required the subject to sort instances of a geometric concept having two relevant dimensions and zero, two, or four irrelevant dimensions. Mean errors to solution were significantly less for subjects who had previously learned the list of words describing the relevant dimensions. The greatest facilitation was noted for problems having four irrelevant dimensions.

A more direct procedure was used by Rasmussen and Archer (1961) to direct attention to the relevant dimension of a concept. Subjects were assigned to one of two pretraining groups: language pretraining in which two paralogs were associated with two nonsense shapes, or aesthetic pretraining in which pleasantness judgments were elicited for the two nonsense shapes. Contrary to expectation, the group given aesthetic pretraining performed better than the group given language pretraining. Rasmussen and Archer suggested that the aesthetic pretraining may have led subjects to attend to and discriminate among the dimensions of the stimuli. Nevertheless, for subjects given language pretraining, performance was significantly better when the dimension labeled during pretraining was relevant than when it was irrelevant. The latter finding lends further support to the facilitatory effect of directing attention to relevant attribute values by verbal cues.

During a three-month period, Wittrock, Keislar, and Stern (1964) taught kindergarten children a hierarchy of associations of words which were later used as cues in a concept identification experiment. The word article (general cue) was associated with the words la and le (class cues) which were in turn associated with the French names (specific cues) for twelve objects and animals. After the preliminary training period, children were given a task requiring them to select one of two pictures which went with a model picture. The basis for matching was the gender of the French name of the model picture. Children were assigned to four treatment groups which differed in the type of cue given during a training task: no cue, general cue, class cue, or specific cue. Following the training task, children were given a retention test and another transfer test. Children who received the class cue during training performed significantly better on immediate transfer, delayed transfer, and retention tests than children receiving more general or more specific cues.

The effect of preexperimentally-learned general, class, and specific cues was examined by Wittrock and Keislar (1965). Second- and third-grade children were taught geometric concepts. During instruction they were presented with a general cue ("color or shape or some other way"), a class cue (color, size, shape, or number), or specific cue (e.g. red, black, orange). Note that, unlike the cues used in the Wittrock, Keislar, and Stern (1964) experiment, these cues and their relationship to one another were known to the children prior to the experiment. Learning, retention, and transfer to new instances were better for children who received the class or specific cue than for children who received the general cue.

One of the variables included in Remstad's (1969) study of concept learning optimization was presence or absence of a verbal cue intended to draw attention to the relevant attribute. Examples of concepts such as quadrilateral and trapezoid were accompanied either by the concept name or by the concept name plus a one word cue (e.g. for quadrilateral, "count;" for trapezoid, "opposite"). The addition of the single word cue produced one of the largest increases in mean correct instance identifications of any individual variable.

In summary, emphasis of relevant attribute values by verbal cues has been shown to improve immediate concept learning performance, transfer, and retention.

Chapter III

DEVELOPMENT OF INSTRUCTIONAL MATERIALS AND TESTS

Three factors led to the selection of geometric concepts as the subject matter of this experiment. First, the logical structure of geometry permitted explicit specification of relevant and irrelevant attributes of the concepts. Second, geometric concepts could be suitably presented by textual materials, allowing control of the teacher variable. Finally, similarity of these concepts to those frequently used in laboratory concept identification experiments admitted the possibility of comparison with laboratory research results.

Since the effect of instructional variations on concept learning by both fourth- and sixth-grade children was to be studied, it was necessary to select concepts of difficulty appropriate to both grade levels. Inspection of elementary mathematics texts suggested that _quadrilateral_, _trapezoid_, _parallelogram_, _rectangle_, _rhombus_, and _square_ are concepts which can be learned by children in fourth grade but are seldom completely mastered by sixth graders. Also, examples of these six concepts are comprised of the same attributes, making it feasible to present the necessary prerequisite knowledge in a brief period of time. The concept of _kite_[1] which is not usually taught in

1. Use of the concept _kite_ was suggested by Dr. Thomas A. Romberg. The definition of _kite_ also was provided by Dr. Romberg.

elementary geometry was also included, since its examples are made up of the same attributes as the other selected concepts.

Test Construction

Each of the selected concepts, quadrilateral, trapezoid, kite, parallelogram, rectangle, rhombus, and square, was analyzed by determining relevant and irrelevant attributes, definition, exemplars and nonexemplars, subordinate and supraordinate concepts. Table 1 lists the attributes and attribute values of the concepts.

TABLE 1

Attributes and Attribute Values Comprising Examples of the Concepts of Quadrilateral, Trapezoid, Kite, Parallelogram, Rectangle, Rhombus, and Square

Attribute	Attribute Values
Closed vs. Open Figure	Closed, Open
Dimensionality of Figure	Plane, Solid
Number of Sides	Three, Four, Five, etc.
Simple vs. Non-Simple Figure	Simple, Non-Simple
Parallelness of Sides	Only one pair of parallel sides, Two pair of parallel sides (opposite sides equal)
Relative Length of Sides	Two pairs of adjacent sides of equal length, All four sides of equal length
Size of Angles	Right angles, Acute angles, Obtuse angles
Size of Figure	Large, Small, etc.
Orientation of Figure	Horizontal, Tipped, etc.
Color of Figure	Black, Red, etc.

Figure 1 indicates the relationship between the concepts. Relevant attribute values for each concept are indicated in parentheses. Lines indicate supraordinate-subordinate relationships between the concepts, with the subordinate concept being placed lower in the figure than the supraordinate. Note that in each case, subordinate concepts have all of the relevant attribute values of the supraordinate, and others in addition.

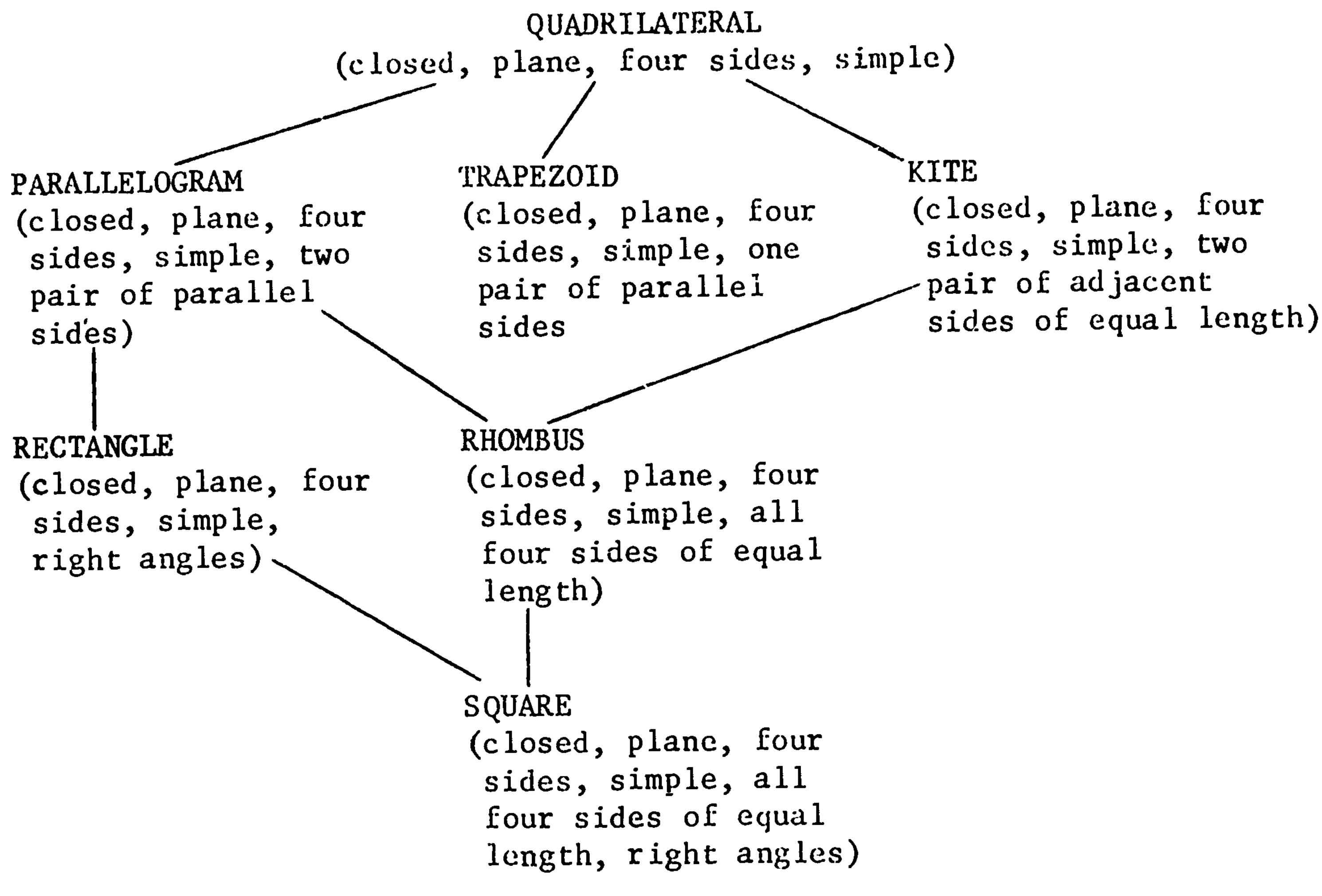


Figure 1. Relationships between the concepts of quadrilateral, parallelogram, trapezoid, kite, rectangle, rhombus, and square. Relevant attribute values of each concept are indicated in parentheses.

A set of behavioral objectives related to concept learning was developed for use in this study. These objectives were based on the analysis of cognitive processes in concept learning which was detailed in Chapter II. The objectives which were developed are listed in the left columns of Tables 2 and 3. Test items exemplifying the behavioral objectives related to the concept quadrilateral are given in the right columns of Tables 2 and 3. Objectives 1a-11a in Table 2 require the selection of a response from several alternatives. Objectives 1b-11b in Table 3 are parallel to the objectives in Table 2 but require the production of a response.

Although specifically developed for this study, the behavioral objectives were intended to reflect cognitive processes in concept learning regardless of subject-matter content. Examples of the application of these same behavioral objectives to transformational grammar concepts may be found in a recent report by Frayer, Fredrick, and Klausmeier (1969).

Selection test item types 1a and 2a require cognition of the attribute value being tested and association of the attribute value with its label. Item types 3a, 4a, and 5a require cognition of an instance and association of the concept with its label. Item types 6a and 7a require discrimination of the relevant from the irrelevant attributes of a concept and association of those attributes with their labels. Item types 8a and 9a require cognition of the conceptual rule and all relevant attribute values and association of the rule and attribute values with their labels. Item types 10a and 11a require cognition of the relationship between the concept and concepts subordinate and supraordinate to it. Production test items are similar

TABLE 2

Behavioral Objectives Related to Concept Learning and Sample Selection Test Items for the Concept of Quadrilateral

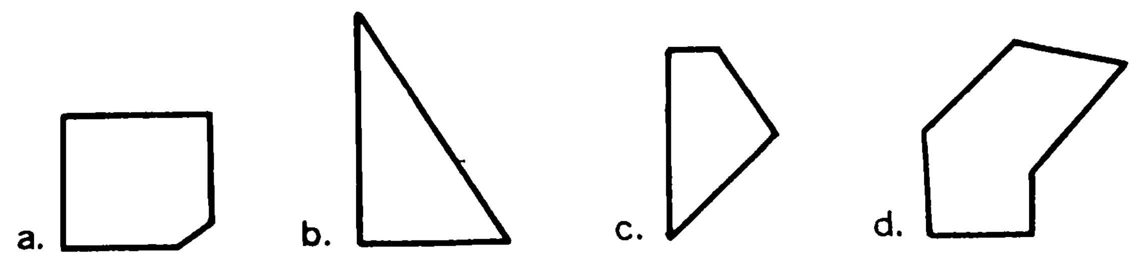
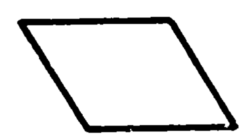
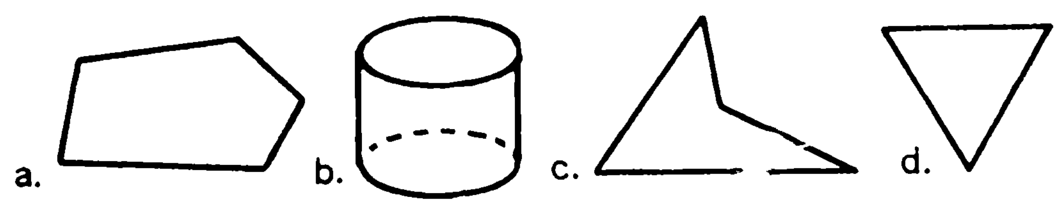
Behavioral Objective	Sample Selection Item
1a. Given the name of an attribute value, the student can select the example of the attribute value.	1a. Which drawing has 4 sides? 
2a. Given an example of an attribute value, the student can select the name of the attribute value.	2a. This figure is:  A. non-simple B. solid C. simple D. open
3a. Given the name of a concept, the student can select the example of the concept.	3a. Which figure is a quadrilateral? 

TABLE 2 (continued)

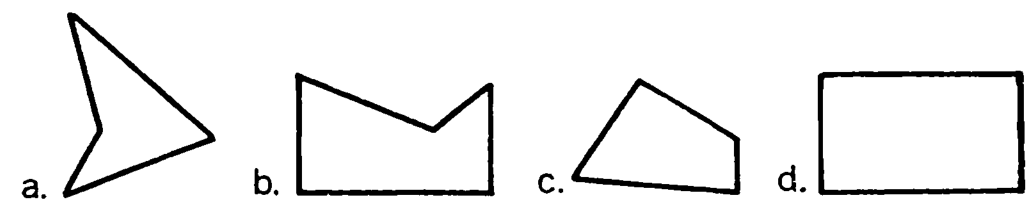
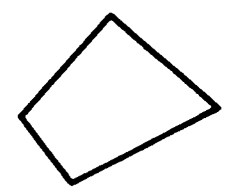
Behavioral Objective	Sample Selection Item
4a. Given the name of a concept, the student can select the non-example of the concept.	4a. Which figure is <u>not</u> a quadrilateral? 
5a. Given an example of a concept, the student can select the name of the concept.	5a. This figure is a:  A. parallelogram B. rectangle C. rhombus D. quadrilateral
6a. Given the name of a concept, the student can select the names of the relevant attribute values of the concept.	6a. <u>All</u> quadrilaterals have: A. no sides of equal length B. no sides parallel C. closed sides D. 2 sides of equal length

TABLE 2 (continued)

Behavioral Objective	Sample Selection Item
7a. Given the name of a concept, the student can select the names of the irrelevant attributes of the concept.	7a. <u>Not</u> all quadrilaterals have: A. straight sides B. 4 angles C. 1 pair of parallel sides D. closed sides
8a. Given the definition of a concept, the student can select the name of the concept.	8a. <u>All</u> plane closed figures with 4 sides may be called: A. rhombuses B. kites C. quadrilaterals D. squares
9a. Given the name of a concept, the student can select the correct definition of the concept.	9a. <u>All</u> quadrilaterals are: A. plane closed figures with opposite angles equal B. plane closed figures with 5 sides C. 4-sided closed plane figures with 4 right angles D. 4-sided closed plane figures

TABLE 2 (continued)

Behavioral Objective	Sample Selection Item
10a. Given the name of a concept, the student can select the name of a concept supraordinate to it.	10a. All quadrilaterals may also be called: A. rhombuses B. kites C. trapezoids D. polygons
11a. Given the name of a concept, the student can select the name of a concept subordinate to it.	11a. Which is true? A. All pentagons may also be called quadrilaterals. B. All kites may also be called quadrilaterals. C. All triangles may also be called quadrilaterals. D. All circles may also be called quadrilaterals.

TABLE 3

Behavioral Objectives Related to Concept Learning and Sample Production Test Items for the Concept of Quadrilateral

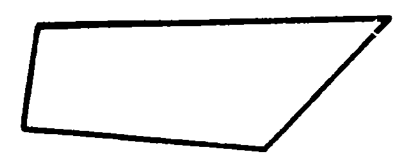
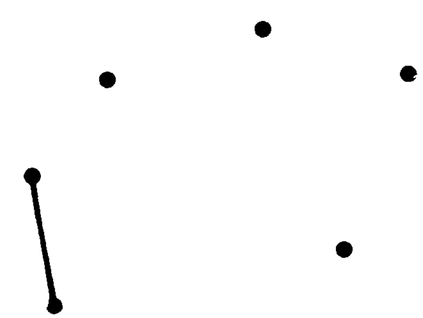
Behavioral Objective	Sample Production Item
1b. Given the name of an attribute value, the student can supply an example of the attribute value.	1b. Draw a closed figure.
2b. Given an example of an attribute value, the student can supply the name of the attribute value.	2b. How many sides does this figure have? 
3b. Given the name of a concept, the student can supply an example of the concept.	3b. Using a ruler, connect as many points as you need to finish the figure so it is a quadrilateral. 

TABLE 3 (continued)

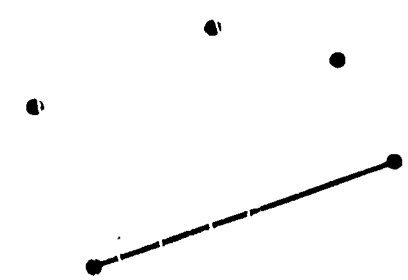
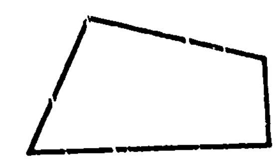
Behavioral Objective	Sample Production Item
4b. Given the name of a concept, the student can supply a non-example of the concept.	4b. Using a ruler, connect as many points as you need to close the figure so it is <u>not</u> a quadrilateral. 
5b. Given an example of a concept, the student can supply the name of the concept.	5b. This polygon  could also be called a ____________________.
6b. Given the name of a concept, the student can supply the names of the relevant attribute values of the concept.	6b. How many sides do quadrilaterals have?

TABLE 3 (continued)

Behavioral Objective	Sample Production Item
7b. Given the name of a concept, the student can supply the names of the irrelevant attributes of the concept.	7b. What is true about the number of parallel sides in a quadrilateral?
8b. Given the definition of a concept, the student can supply the name of the concept.	8b. <u>All</u> plane closed figures with 4 sides may be called ________________.
9b. Given the name of a concept, the student can supply the correct definition of the concept.	9b. List all that is needed to completely describe <u>quadrilateral</u>.
10b. Given the name of a concept, the student can supply the name of a concept supraordinate to it.	10b. All quadrilaterals may also be called ________________.
11b. Given the name of a concept, the student can supply the name of a concept subordinate to it.	11b. Give the name of one geometric figure which is a special kind of quadrilateral.

to the selection items, but in each case the process which is being tested is what Guilford (1967) would term "convergent production."

The behavioral objectives may be to some extent hierarchical. For example, it seems unlikely that a subject could supply a definition of a concept without being able to identify some of its relevant attribute values. On the other hand, he could correctly identify some of the relevant attribute values, yet misclassify a concept instance. On a logical basis, however, we might predict that item difficulty would generally increase f. om item type 1 to item type 11. Five specific difficulty levels were anticipated: Level 1, item types 1 and 2; Level 2, item types 3, 4, and 5; Level 3, item types 6 and 7; Level 4, item types 8 and 9; and Level 5, item types 10 and 11.

In order to test the effect of instructional variations on attainment of geometric concepts, the behavioral objectives were utilized to construct two parallel forms of a multiple-choice test, with each form comprised of items as indicated in Table 4. In addition, a completion test was constructed, consisting of questions parallel to those on the multiple-choice tests but requiring production of the answers. The sequence of items in the multiple-choice and completion tests was the same. In general, the pattern was to ask one question of each item type 1-11, then repeat the cycle until the questions of each item type for all concepts were completed. The sequence of item types differed for each concept.

TABLE 4

Content of Tests by Item Type and Concept

Item Type	Concept							Total Items
	Quadri-lateral	Kite	Trapezoid	Parallel-ogram	Rec-tangle	Rhombus	Square	
1	4	1	1	1	1	1	1	10
2	4	1	1	1	1	1	1	10
3	1	1	1	1	1	1	1	7
4	1	1	1	1	1	1	1	7
5	1	1	1	1	1	1	1	7
6	1	1	1	1	1	1	1	7
7	1	1	1	1	1	1	1	7
8	1	1	1	1	1	1	1	7
9	1	1	1	1	1	1	1	7
10	1	1	1	1	1	1	1	7
11	1	1	0	1	1	1	0	5
								81

Lesson Construction

Lessons were devised to teach the selected geometric concepts. A modified linear program format was used to present the lessons. This format required the child to respond to questions concerning the ideas presented and provided feedback concerning the correctness of his answers. Since the child could work independently with this type of instructional material, the effect of teacher variability was minimized. An attempt was made to include all background information required to learn the concepts in order to attenuate the effect of differences in previous geometry training.

Four lessons were developed, each requiring an estimated thirty minutes for completion. Lesson content was as follows:

<u>Lesson I</u> point, line segment, line, ray, angle

<u>Lesson II</u> right angle, closed curve, simple curve, plane, polygon, parallel, adjacent, cpposite, equal length

<u>Lesson III</u> quadrilateral, kite, trapezoid, parallelogram, rectangle, rhombus, square

<u>Lesson IV</u> relationships between the concepts included in Lesson III

After the lessons were written, a professor and two graduate students in mathematics education reviewed them to assure accuracy of content. One fourth-grade and one fifth-grade student worked through the lessons on an individual basis with the experimenter. This procedure disclosed several ambiguities and some vocabulary and format problems. Revisions were carried out on the basis of the subject-matter consultants' comments and initial tryout results.

Pilot Study of Lessons and Tests

Purpose

The pilot study has as its objectives:

(1) to determine the appropriateness of lesson format and reading level for fourth- and sixth-grade students,

(2) to determine error rate for each lesson frame, as a basis for lesson revision,

(3) to obtain item statistics and reliability information for the multiple-choice test items as a basis for test revision,

(4) to obtain item statistics and reliability information for the completion test items as a basis for test revision, and

(5) to determine time requirements and revise directions.

Procedure

Subjects. Lessons and tests were administered to 140 fourth-grade and 140 sixth-grade students. The fourth-grade children comprised the entire fourth-grade population of an elementary school in a Midwestern suburban community. The sixth-grade children comprised the population of 5 sixth-grade classrooms in a middle school in the same community.

Treatment. The design of the pilot study is shown in Table 5. Four lessons were administered, one per day for four consecutive school days. On the school day immediately following the administration of the fourth lesson, the multiple-choice test was given. Children at each grade level were randomly assigned to one of the two parallel forms, with half of the children taking one form, the other half the other form.

Eleven days after administration of the multiple-choice test (a school vacation intervened), the completion test was given. All children received the same form of this test.

Results. The distributions of total errors and lesson time requirements were determined for each lesson at each grade level. Descriptive statistics for total errors and working times of lessons are presented in Table 6. Mean total error rate varied from 4% to 11%. All students were able to finish each lesson within a 45-minute class period.

In general, frame error rate was low. Any frame with an error rate higher than 15% at either grade level was revised prior to the main experiment.

TABLE 5

Design of the Pilot Study

	Grade 4		Grade 6	
Day 1	Lesson I		Lesson I	
Day 2	Lesson II		Lesson II	
Day 3	Lesson III		Lesson III	
Day 4	Lesson IV		Lesson IV	
Day 5	Multiple-Choice Test Form AA	Multiple-Choice Test Form BB	Multiple-Choice Test Form AA	Multiple-Choice Test Form BB
Day 6	Completion Test		Completion Test	

TABLE 6

Means and Standard Deviations of Total Errors and Mean Working Times for Pilot Study Lessons for Grades 4 and 6

Lesson	Number of Frames	Grade 4			Grade 6		
		Total Errors		Times (min.)	Total Errors		Times (min.)
		M	SD	M	M	SD	M
I	43	4.9	3.4	23.1	3.5	2.6	20.2
II	73	5.7	5.7	21.3	4.8	3.6	19.5
III	112	7.8	6.5	21.1	6.5	5.7	19.0
IV	64	3.6	3.6	17.0	2.6	2.3	14.9

Summary statistics for the multiple-choice and completion tests are given in Tables 7, 8, and 9. All tests were item analyzed by the FORTAP computer program (Baker & Martin, 1968). Separate analyses were carried out for each test form at each grade level. Scores were obtained for the total tests and each of the item type subtests.

Reliabilities for the total tests at both grade levels were high, .85 or greater. Subtest reliabilities varied substantially among different subtests and between grades. In general, reliabilities were lowest for grade 4 on the higher subtests. Low reliabilities were anticipated for the subtests since the number of items per subtest was small and the items unrefined.

Item difficulties also differed among subtests and between grades, with difficulty being greater for the higher subtests and for fourth-grade children. In all cases, mean scores indicated better than chance performance. Overall test difficulty was more suitable for grade 6 than for grade 4.

Sixth-grade mean total scores on both forms of the multiple-choice test were approximately midway between the expected chance proportion (25%) and 100%. The mean total score for sixth graders on the completion test was also mid-range.

On the basis of item analyses, individual items were revised to eliminate unused distractors, adjust difficulty level, and maximize discrimination. Revision of items was carried out with the goal of maximizing reliabilities for the item type subtests.

TABLE 7

Means, Standard Deviations, and Reliabilities for Multiple-Choice Test: Form AA for Grades 4 and 6

Test	Number of Items	Grade 4 (N=67)			Grade 6 (N=68)		
		M	SD	Hoyt Reliability	M	SD	Hoyt Reliability
Total	81	38.39	10.50	.86	50.53	9.82	.86
Type 1	10	6.93	1.85	.58	8.56	1.24	.38
Type 2	10	6.00	2.22	.63	7.88	1.44	.38
Type 3	7	3.16	1.21	.34	4.38	1.48	.50
Type 4	7	3.69	1.45	.38	4.60	1.27	.35
Type 5	7	3.75	1.48	.42	4.41	1.38	.40
Type 6	7	2.91	1.41	.30	3.93	1.32	.30
Type 7	7	2.39	1.41	.27	3.53	1.61	.47
Type 8	7	2.28	1.28	.22	3.25	1.29	.30
Type 9	7	3.06	1.58	.40	4.47	1.33	.30
Type 10	7	2.51	1.56	.40	3.63	1.53	.39
Type 11	5	1.72	1.06	.02	1.88	1.24	.34

TABLE 8

Means, Standard Deviations, and Reliabilities for Multiple-Choice Test: Form BB for Grades 4 and 6

Test	Number of Items	Grade 4 (N=65)			Grade 6 (N=67)		
		M	SD	Hoyt Reliability	M	SD	Hoyt Reliability
Total	81	37.58	10.09	.85	48.10	10.71	.88
Type 1	10	6.57	1.72	.48	8.34	1.30	.35
Type 2	10	6.31	2.43	.73	7.91	1.56	.46
Type 3	7	3.22	1.45	.42	4.10	1.50	.48
Type 4	7	3.23	1.58	.52	4.04	1.45	.35
Type 5	7	3.37	1.49	.43	4.60	1.54	.47
Type 6	7	2.88	1.32	.26	3.93	1.39	.38
Type 7	7	2.69	1.39	.17	3.06	1.53	.36
Type 8	7	2.43	1.25	.16	3.07	1.34	.34
Type 9	7	2.55	1.26	.23	3.49	1.58	.48
Type 10	7	2.98	1.43	.22	3.45	1.57	.39
Type 11	5	1.35	1.05	.16	2.10	1.21	.23

TABLE 9

Means, Standard Deviations, and Reliabilities for Completion Test for Grades 4 and 6

Test	Number of Items	Grade 4 (N=130)			Grade 6 (N=137)		
		M	SD	Hoyt Reliability	M	SD	Hoyt Reliability
Total	31	22.30	10.05	.90	37.85	10.64	.89
Type 1	10	5.44	1.85	.58	7.34	1.79	.58
Type 2	10	4.17	2.32	.67	7.07	1.40	.34
Type 3	7	2.82	1.75	.65	4.28	1.38	.49
Type 4	7	2.18	1.57	.54	3.91	1.81	.64
Type 5	7	2.06	1.34	.41	3.28	1.43	.35
Type 6	7	1.05	.89	.33	2.23	1.25	.46
Type 7	7	.50	.53	-.03	1.47	1.11	.59
Type 8	7	1.06	1.02	.38	2.28	1.46	.47
Type 9	7	.70	.95	.43	1.83	1.46	.60
Type 10	7	.83	1.07	.45	1.91	1.40	.44
Type 11	5	1.48	1.21	.44	2.26	1.22	.32

Students were to be permitted as much time as they wished to complete the tests. During the pilot study, it became evident that the completion test would require two entire class periods instead of one. Scheduling conflicts prevented allowing this much time. As a consequence, many fourth graders and a few sixth graders did not complete this test. Scores are therefore lower than they would have been had adequate time been available. For the main experiment, the test was divided into two parts to be administered on separate days.

Chapter IV

METHOD

The primary purpose of this experiment was to determine the effect on concept learning of two instructional variables, number of instances and emphasis of relevant attribute values. On the basis of related research and logical analysis, two predictions were made regarding the effect of these variables: (1) level of concept mastery would increase as a function of the number of instances presented, and (2) emphasis of relevant attribute values would facilitate concept learning.

Subjects

The subjects in this study were 154 fourth-grade and 126 sixth-grade children. The fourth-grade children constituted the entire fourth-grade population of an elementary school located in the same community in which the pilot study was conducted. The sixth-grade children comprised the population of five classrooms in the same middle school in which the pilot study was carried out. The five classrooms of children were selected on the basis of convenience of scheduling from ten heterogeneously grouped classrooms. Two low ability classes and one high ability class were deliberately excluded from the study. The study began with 169 fourth-grade and 141 sixth-grade children. Fifteen students at each grade level were lost due to absences during the experiment.

The experimenters were one male and one female graduate student, both of whom became familiar with procedures and materials prior to the experiment.

Instructional Materials

To study the effect of number of instances and emphasis of relevant attribute values on concept learning, lessons were desired which were similar to those normally used in the school, but which rigorously controlled the variables of interest. Based on results of the pilot study, lessons were constructed to attain the twin objectives of realism and control. Lessons were designed to be administered on four successive school days. Lesson I was the same for all children. Variations in Lessons II, III, and IV constituted the experimental treatments. Content of lessons for each treatment group is outlined in Table 10.

TABLE 10

Content of Instructional Treatments

Treatment		Lesson			
		I	II	III	IV
1	A	Background	Attributes 1	Attributes 2	Concepts 1
	B	Background	Attributes 1	Attributes 2	Concepts 2
2	A	Background	Attributes 1	Attributes 2	Concepts 1 (Emphasis)
	B	Background	Attributes 1	Attributes 2	Concepts 2 (Emphasis)
3	A	Background	Attributes 1	Concepts 1	Concepts 2
	B	Background	Attributes 2	Concepts 1	Concepts 2
4	A	Background	Attributes 1	Concepts 1 (Emphasis)	Concepts 2 (Emphasis)
	B	Background	Attributes 2	Concepts 1 (Emphasis)	Concepts 2 (Emphasis)

Descriptions of the lessons are as follows:

Background	Point, line segment, line, ray, and angle were introduced.
Attributes 1	Right angle, closed curve, simple curve, plane, polygon, parallel, adjacent, opposite, and equal length were described and examples given.
Attributes 2	Content was similar to Attributes 1, with rephrasing of descriptions and different examples of the attributes.
Concepts 1	Concepts of quadrilateral, kite, trapezoid, parallelogram, rectangle, rhombus, and square were introduced. For each concept, the definition, two positive and two negative instances were presented.
Concepts 2	Content was similar to Concepts 1. Definitions of concepts were repeated, and two positive and two negative instances different from those in Concepts 1 were given.
Concepts 1 (Emphasis)	Content was precisely the same as that of Concepts 1, but in addition had questions designed to direct attention to the relevant attribute values of the concept instance (e.g. "Does this figure have 4 sides?") and reviewed the relevant attribute values

t the conclusion of each concept's presentation (e.g. "How many sides does a quadrilateral have?").

Concepts 2 (Emphasis) — Content was precisely the same as Concepts 2, but in addition had attention-directing questions and reviewed relevant attribute values.

Tests

Multiple-choice test. A single multiple-choice test was constructed by selecting the best items from the two parallel multiple-choice tests used in the pilot study and revising when necessary to improve item characteristics. The composition of the test was the same as that indicated in Table 4. Selection and revision of items were carried out with the following goals:

(1) to make all items of appropriate difficulty for both fourth- and sixth-grade children,

(2) to produce items which would effectively discriminate at both fourth- and sixth-grade levels,

(3) to eliminate non-functioning distractors, and

(4) to maximize item-type subtest reliabilities.

Completion test. A completion test was devised by revision of the test utilized in the pilot study. The primary changes made were the addition of prompts to verbal questions which had proved too difficult. Also, some items which required completion of a drawing were changed to make the correct answer more perceptually obvious. The test was divided into two parts, so that each part could be administered during a regular class period.

Procedure

The schedule for the study was as follows: Days 1-4, administration of Lessons I-IV; Day 5, multiple-choice test; Day 6, completion test, part 1 (preceded by an interval of 72 hours); and Day 7, completion test, part 2.

On the first day of the experiment, the children were given general instructions concerning the purpose of the study and procedures to be followed in completing the lessons. A copy of these instructions comprises Appendix A. Children were reminded of the essential points in these instructions on Days 2, 3, and 4.

Prior to the beginning of each lesson, new vocabulary was reviewed. A vocabulary list was included in each lesson booklet. The experimenter read aloud each word on the list and had the children repeat it after him. The list was then repeated in random order, and the children were directed to raise their hands when they had found the word which had been pronounced. After allowing time for most children to locate the word, the experimenter called on one of the children to give the number of the word which had been pronounced. Thus, feedback was provided to all children so they could be sure they had found the correct word. This vocabulary review was an attempt to produce sight recognition of the words contained in the lesson.

While the children completed the lessons, the experimenter proctored to be sure directions were followed. No assistance was offered to the children other than to fulfill requests for pronunciation of words or clarification of procedure.

Following the experiment, teachers were requested to complete a questionnaire indicating which of the concepts had been studied by their class during the school year. In addition, teachers were asked to estimate the degree of mastery which their students possessed of each concept. A copy of the questionnaire comprises Appendix B.

Experimental Design

The two independent variables were number of concept examples (4 or 8) and emphasis of relevant attribute values (presence or absence of attention-directing and review questions). Factorial combination of the levels on these variables resulted in four basic treatment groups. In addition, however, treatment groups which received only one concept lesson were counterbalanced so that half of the children received one concept lesson, half another. For example, for Lesson IV of treatment 1, half of the children received Concepts 1, the other half Concepts 2. For treatment groups which received only one attribute lesson, similar counterbalancing was carried out so that half of the children received Attributes 1, half Attributes 2. The nature of this counterbalancing may be clarified by reference to Table 10. Counterbalancing of each of the four treatment groups resulted in a total of eight different treatments.

Within each of the six fourth-grade and five sixth-grade classes, subjects were randomly assigned to one of the eight treatment groups. Thus, subjects were nested within classes. The total design may then be characterized as a treatments x blocks design with subjects nested within class and treatments crossed with class. A two-way fixed effects analysis of variance model was assumed with the mean square error term as the denominator of the F-ratio for both main effects and the interaction. The design of the experiment is illustrated in Table 11.

TABLE 11

Design of the Experiment

Grade	Class	4 Instances				8 Instances			
		Non-Emphasis		Emphasis		Non-Emphasis		Emphasis	
		A	B	A	B	A	B	A	B
	1	S_1 . S_n							
	2								
4	3								
	4								
	5								
	6								
	1								
	2								
6	3								
	4								
	5								

Technical Report No. 116

EFFECTS OF NUMBER OF INSTANCES AND EMPHASIS OF RELEVANT ATTRIBUTE VALUES ON MASTERY OF GEOMETRIC CONCEPTS BY FOURTH- AND SIXTH-GRADE CHILDREN (Part II) (Chapter V to Conclusion)

Report from the Project on
Situational Variables and
Efficiency of Concept Learning

By Dorothy Ann Frayer

Herbert J. Klausmeier, Professor of Educational Psychology
Chairman of the Examining Committee and Principal Investigator

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

March 1970

This Technical Report is a doctoral dissertation reporting research supported by the Wisconsin Research and Development Center for Cognitive Learning. Since it has been approved by a University Examining Committee, it has not been reviewed by the Center. It is published by the Center as a record of some of the Center's activities and as a service to the student. The bound original is in The University of Wisconsin Memorial Library.

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education and Welfare, under the provisions of the Cooperative Research Program. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

Chapter V

RESULTS

Twenty-four scores were obtained for each subject: the total test and eleven subtest scores of the multiple-choice test, and the total test and eleven subtest scores of the completion test. In addition, the score for each student on the Paragraph Meaning test of the Stanford Achievement Test battery (Kelley, Madden, Gardner, & Rudman, 1964) was procured from school records for use as a covariate.

In the initial section of this chapter, reliability estimates and item statistics will be presented for both the multiple-choice and completion tests. Subsequently, the multivariate analyses of covariance on total scores and on subtest scores will be reported. Descriptive statistics for the lessons and results of the teacher questionnaire will comprise the final section.

Psychometric Characteristics of Tests

Sample

The reliability estimates and item statistics reported in this section are based on data for all subjects completing each test during the main experiment. The sample size on which the test statistics are based is somewhat larger than that for the multivariate analyses to be reported later, since subjects missing any lesson or test were dropped from the multivariate analyses. At the fourth-grade level, 167 subjects were given the multiple-choice test, 173 the completion test. At the sixth-grade level, 142 subjects were given the multiple-choice test, 136 the completion test.

Test Statistics

Tests were analyzed by the FORTRAN Test Analysis Package (Baker & Martin, 1968). Separate statistics were computed at each grade level for each of the multiple-choice and completion tests' total tests and subtests. The means, standard deviations, ranges, standard errors of measurement, and Hoyt internal consistency reliabilities (Hoyt, 1941) for the multiple-choice total test and subtests are presented in Tables 12 and 13; for the completion total test and subtests in Tables 14 and 15.

A summary of item characteristics (difficulty, X_{50}, point biserial correlation coefficient, and β) for the correct choices for each of the total tests and subtests is presented in Tables 16-23. Utilization of this set of item characteristics has been suggested by Allen, Feezel, Kauffeld, and Harris (1969). Item difficulty refers to the proportion of subjects responding correctly to an item. X_{50} is the point on the criterion scale, given in standard deviation units, corresponding to the median of the item characteristic curve and is the point of maximum discrimination. The criterion used was the total score on the test being considered. The point biserial correlation coefficient and β are indices of the discriminating power of an item. The point biserial coefficient relates scores on a given item to scores on the total test

TABLE 12

Reliability Estimates for Multiple-Choice Total Test and Subtests for Grade 4

Test	Number of Items	Mean Score	Standard Deviation	Range of Scores	Standard Error of Measurement	Hoyt Reliability
Total	81	33.20	9.30	11-59	3.99	.81
Type 1	10	6.04	1.98	2-10	1.27	.54
Type 2	10	5.08	2.19	1-10	1.39	.55
Type 3	7	3.14	1.30	0-6	1.07	.21
Type 4	7	3.38	1.45	0-7	1.06	.37
Type 5	7	2.81	1.41	0-7	1.12	.26
Type 6	7	2.41	1.31	0-6	1.08	.20
Type 7	7	1.82	1.11	0-5	1.08	-.10
Type 8	7	2.32	1.25	0-5	1.11	.08
Type 9	7	2.64	1.35	0-6	1.11	.22
Type 10	7	2.16	1.30	0-6	1.11	.15
Type 11	5	1.40	0.99	0-4	0.89	.00

TABLE 13

Reliability Estimates for Multiple-Choice Total Test and Subtests for Grade 6

Test	Number of Items	Mean Score	Standard Deviation	Range of Scores	Standard Error of Measurement	Hoyt Reliability
Total	81	44.33	10.33	19-74	3.86	.86
Type 1	10	7.87	1.56	3-10	1.00	.55
Type 2	10	7.30	2.09	2-10	1.19	.64
Type 3	7	3.60	1.27	1-7	1.04	.21
Type 4	7	4.43	1.30	0-7	1.03	.26
Type 5	7	3.49	1.72	0-7	1.10	.52
Type 6	7	3.36	1.47	1-7	1.06	.39
Type 7	7	2.97	1.59	0-7	1.15	.39
Type 8	7	2.95	1.45	0-7	1.10	.33
Type 9	7	3.67	1.41	0-7	1.09	.30
Type 10	7	2.91	1.58	0-7	1.15	.38
Type 11	5	1.79	1.09	0-5	0.94	.08

TABLE 14

Reliability Estimates for Completion Total Test and Subtests for Grade 4

Test	Number of Items	Mean Score	Standard Deviation	Range of Scores	Standard Error of Measurement	Hoyt Reliability
Total	81	26.77	9.39	3-57	3.31	.87
Type 1	10	6.24	1.92	2-10	1.14	.61
Type 2	10	4.99	2.25	0-9	1.20	.68
Type 3	7	3.61	1.58	0-7	1.03	.51
Type 4	7	2.59	1.60	0-7	1.09	.45
Type 5	7	2.27	1.40	0-6	0.96	.45
Type 6	7	1.88	1.14	0-6	0.81	.41
Type 7	7	1.01	0.67	0-4	0.58	.12
Type 8	7	1.08	1.04	0-5	0.77	.36
Type 9	7	0.53	0.83	0-5	0.61	.36
Type 10	7	0.99	1.06	0-5	0.81	.33
Type 11	5	1.59	1.15	0-5	0.84	.32

TABLE 15

Reliability Estimates for Completion Total Test and Subtests for Grade 6

Test	Number of Items	Mean Score	Standard Deviation	Range of Scores	Standard Error of Measurement	Hoyt Reliability
Total	81	39.46	9.79	14-75	3.51	.87
Type 1	10	8.04	1.55	3-10	1.02	.52
Type 2	10	7.27	1.62	2-10	1.13	.46
Type 3	7	4.85	1.31	2-7	1.00	.32
Type 4	7	3.66	1.75	0-7	1.11	.53
Type 5	7	3.20	1.41	0-7	1.02	.38
Type 6	7	3.07	1.38	0-7	1.01	.37
Type 7	7	1.29	0.74	0-6	0.53	.41
Type 8	7	2.05	1.47	0-7	0.95	.52
Type 9	7	1.89	1.53	0-7	0.95	.55
Type 10	7	1.86	1.29	0-6	1.00	.29
Type 11	5	2.27	1.34	0-5	0.85	.50

TABLE 16

Summary of Item Difficulties and X_{50} Parameters for Items of Multiple-Choice Total Test and Subtests for Grade 4

Test	Number of Items	Number of Items Within Middle Difficulty Range (.475-.775)	Range of Item Difficulty	Mean Difficulty	Number of Low X_{50}s (-1.01 or below)	Number of Medium X_{50}s (-1.00 to +1.00)	Number of High X_{50}s (+1.01 or above)
Total	81	24	.12-.92	.41	13	30	38
Type 1	10	7	.20-.92	.60	4	6	0
Type 2	10	6	.35-.68	.51	2	8	0
Type 3	7	3	.17-.75	.45	0	4	3
Type 4	7	2	.19-.79	.48	1	4	2
Type 5	7	2	.20-.65	.40	0	5	2
Type 6	7	1	.12-.66	.34	0	5	2
Type 7	7	0	.16-.34	.26	0	1	6
Type 8	7	1	.18-.53	.33	0	4	3
Type 9	7	2	.22-.71	.38	1	4	2
Type 10	7	0	.26-.41	.31	0	3	4
Type 11	5	0	.19-.40	.28	0	2	3

TABLE 17

Summary of Item Difficulties and X_{50} Parameters for Items of Multiple-Choice Total Test and Subtests for Grade 6

Test	Number of Items	Number of Items Within Middle Difficulty Range (.475-.775)	Range of Item Difficulty	Mean Difficulty	Number of Low X_{50}s (-1.01 or below)	Number of Medium X_{50}s (-1.00 to +1.00)	Number of High X_{50}s (+1.01 or above)
Total	81	23	.15-.99	.55	27	36	18
Type 1	10	0	.21-.99	.79	8	1	1
Type 2	10	6	.48-.86	.73	6	4	0
Type 3	7	1	.27-.90	.51	2	5	0
Type 4	7	3	.25-.89	.63	3	4	0
Type 5	7	3	.35-.77	.50	1	6	0
Type 6	7	2	.15-.89	.48	1	5	1
Type 7	7	2	.28-.50	.42	0	6	1
Type 8	7	2	.20-.75	.42	1	4	2
Type 9	7	2	.31-.89	.52	1	5	1
Type 10	7	2	.29-.54	.42	0	7	0
Type 11	5	0	.22-.46	.36	0	4	1

TABLE 18

Summary of Point Biserial Correlation Coefficients and β Parameters for Items of Multiple-Choice Total Test and Subtests for Grade 4

Test	Number of Items	Number of Items With Point Biserial Correlations ≥ .30	Range of Point Biserial Correlations	Mean Point Biserial Correlation	Range of β s	Mean β
Total	81	29	-.12 to .55	.25	-.19 to 1.07	.37
Type 1	10	9	-.03 to .57	.43	-.04 to 1.09	.78
Type 2	10	9	.29 to .60	.44	.40 to 1.17	.73
Type 3	7	7	.31 to .59	.42	.43 to 1.26	.73
Type 4	7	6	.14 to .57	.45	.20 to 1.37	.84
Type 5	7	7	.36 to .54	.43	.47 to .99	.68
Type 6	7	6	.19 to .52	.41	.33 to .92	.66
Type 7	7	6	.29 to .41	.36	.42 to .73	.58
Type 8	7	4	.27 to .53	.38	.43 to .94	.61
Type 9	7	6	.27 to .52	.42	.39 to 1.04	.68
Type 10	7	6	.28 to .47	.40	.40 to .80	.65
Type 11	5	5	.40 to .49	.45	.59 to .88	.77

TABLE 19

Summary of Point Biserial Correlation Coefficients and β Parameters for Items of Multiple-Choice Total Test and Subtests for Grade 6

Test	Number of Items	Number of Items With Point Biserial Correlations ≥ .30	Range of Point Biserial Correlations	Mean Point Biserial Correlation	Range of βs	Mean β
Total	81	33	.02 to .49	.29	.08 to .98	.44
Type 1	10	9	.05 to .56	.43	.20 to 1.46	1.01
Type 2	10	9	.29 to .67	.49	.39 to 1.88	1.13
Type 3	7	5	.29 to .53	.41	.54 to .96	.70
Type 4	7	7	.35 to .50	.43	.58 to .94	.74
Type 5	7	7	.46 to .56	.51	.70 to 1.02	.86
Type 6	7	6	.24 to .59	.45	.43 to 1.12	.79
Type 7	7	7	.34 to .63	.46	.50 to 1.28	.77
Type 8	7	6	.29 to .52	.44	.46 to .88	.72
Type 9	7	6	.27 to .54	.44	.37 to 1.20	.78
Type 10	7	7	.30 to .52	.46	.42 to .94	.74
Type 11	5	4	.29 to .55	.46	.44 to 1.09	.77

TABLE 20

Summary of Item Difficulties and X_{50} Parameters for Items of Completion Total Test and Subtests for Grade 4

Test	Number of Items	Number of Items Within Middle Difficulty Range (.30-.70)	Range of Item Difficulty	Mean Difficulty	Number of Low X_{50}s (-1.01 or below)	Number of Medium X_{50}s (-1.00 to +1.00)	Number of High X_{50}s (+1.01 or above)
Total	81	26	.00-.97	.33	13	23	45
Type 1	10	5	.31-.97	.62	4	6	0
Type 2	10	6	.08-.86	.50	2	6	2
Type 3	7	4	.21-.89	.52	1	5	1
Type 4	7	4	.23-.57	.37	0	7	0
Type 5	7	3	.09-.73	.32	0	3	4
Type 6	7	2	.01-.87	.27	1	3	3
Type 7	7	0	.00-.76	.14	0	4	3
Type 8	7	1	.04-.53	.15	0	1	6
Type 9	7	0	.02-.21	.08	0	3	4
Type 10	7	0	.02-.27	.14	0	2	5
Type 11	5	1	.23-.62	.32	0	3	2

TABLE 21

Summary of Item Difficulties and X_{50} Parameters for Items of Completion Total Test and Subtests for Grade 6

Test	Number of Items	Number of Items Within Middle Difficulty Range (.30-.70)	Range of Item Difficulty	Mean Difficulty	Number of Low X_{50}s (-1.01 or below)	Number of Medium X_{50}s (-1.00 to +1.00)	Number of High X_{50}s (+1.01 or above)
Total	81	36	.01- 1.00	.49	26	29	26
Type 1	10	4	.46- 1.00	.80	4	6	0
Type 2	10	5	.30-.96	.73	5	4	1
Type 3	7	3	.42-.98	.69	3	4	0
Type 4	7	7	.33-.67	.52	1	6	0
Type 5	7	2	.26-.90	.46	2	3	2
Type 6	7	3	.13-.96	.44	1	3	3
Type 7	7	0	.01-.95	.18	1	5	1
Type 8	7	1	.10-.73	29	0	3	4
Type 9	7	3	.09-.57	.27	0	4	3
Type 10	7	4	.02-.45	.27	0	4	3
Type 11	5	4	.30-.80	.45	1	4	0

TABLE 22

Summary of Point Biserial Correlation Coefficients and β Parameters for Items of Completion Total Test and Subtests for Grade 4

Test	Number of Items	Number of Items With Point Biserial Correlations ≥ .30	Range of Point Biserial Correlations	Mean Point Biserial Correlation	Range of βs	Mean β
Total	81	41	.00 to .56	.29	.00 to 1.51	.52
Type 1	10	8	.22 to .62	.46	.59 to 2.39	1.03
Type 2	10	9	.22 to .69	.50	.43 to 1.70	1.00
Type 3	7	7	.39 to .58	.51	.61 to 2.12	1.08
Type 4	7	6	.25 to .57	.49	.33 to 1.10	.87
Type 5	7	6	.25 to .61	.47	.46 to 1.52	.97
Type 6	7	6	.28 to .66	.46	.00 to 1.73	1.05
Type 7	7	5	.00 to .68	.36	.00 to 2.55	.80
Type 8	7	6	.15 to .61	.44	.37 to 3.33	1.64
Type 9	7	6	.29 to .64	.46	.00 to 3.44	1.20
Type 10	7	6	.08 to .59	.42	.27 to 1.45	.98
Type 11	5	5	.39 to .58	.52	.64 to 1.28	1.01

TABLE 23

Summary of Point Biserial Correlation Coefficients and β Parameters for Items of Completion Total Test and Subtests for Grade 6

Test	Number of Items	Number of Items With Point Biserial Correlations ≥ .30	Range of Point Biserial Correlations	Mean Point Biserial Correlation	Range of βs	Mean β
Total	81	42	.00 to .55	.29	.00 to 1.15	.49
Type 1	10	6	.00 to .62	.37	.00 to 1.40	.84
Type 2	10	8	.23 to .55	.41	.51 to 1.45	.84
Type 3	7	6	.17 to .53	.43	.47 to 1.12	.82
Type 4	7	6	.28 to .63	.51	.40 to 1.27	.90
Type 5	7	7	.39 to .54	.46	.63 to 1.03	.83
Type 6	7	5	.23 to .60	.44	.32 to 1.17	.86
Type 7	7	7	.31 to .64	.49	.00 to 1.80	.42
Type 8	7	7	.43 to .58	.50	.81 to 1.19	1.03
Type 9	7	6	.28 to .64	.50	.56 to 1.34	1.04
Type 10	7	6	.29 to .54	.43	.62 to 1.35	.86
Type 11	5	5	.50 to .65	.57	.86 to 1.41	1.18

of which the item is a part. β is the reciprocal of the standard deviation of the item characteristic curve. The skewed distributions of the point biserial correlation coefficients and β's present problems in use of the mean as a measure of central tendency. General trends may, however, be revealed by inclusion of these statistics (Allen et al., 1969).

Mean scores for the multiple-choice test presented in Tables 12 and 13 are generally lower than the comparable scores obtained during the pilot study. Two differences between the studies may account for this decrease in mean scores. First, item revision may have inadvertently increased difficulty although in most cases this was not an intended outcome. Second, one of the lessons given in the pilot study but not in the main study dealt with the relationships between the concepts being taught. This lesson may have increased performance, particularly on item types 10 and 11.

The reliability estimates for the multiple-choice total test and subtests reported in Tables 12 and 13 are also somewhat lower than those obtained in the pilot study. Particularly low reliabilities were found for scales 3 and 5 - 11 at fourth-grade level, and scales 3, 4, and 11 at sixth-grade level. The reliability of a set of test scores is related to other characteristics of the test. Ebel (1965) indicated that the reliability coefficient will usually be greater: (1) for longer tests than shorter tests, (2) for tests composed of homogeneous items than for tests composed of heterogeneous items, and (3) for tests whose items are of intermediate difficulty than for tests comprised primarily of very difficult or very easy items, and (4) for tests whose items are more discriminating than for tests whose items are less discriminating.

Inspection of Tables 16-19 suggests that a combination of these factors rather than a single factor accounts for the low reliabilities. A smaller number of items, lower item difficulties, and lower mean point biserial correlations appear to be associated with low reliability coefficients.

Mean scores for the completion test (Tables 14 and 15) are generally higher than those obtained for the pilot study. Insufficient time was available for administration of the completion test to fourth-grade children during the pilot study. Thus, it was predictable that mean scores for fourth graders would be higher when a longer time allowance was provided. Also, the retention interval between Lesson IV and the completion test was eleven days for the pilot study, but only four days for the main study.

Problems were encountered with completion item type 7. During the pilot study, the questions of this type proved highly difficult and non-discriminating for fourth grade children. An example of the wording of the questions used during the pilot study for this study was "How many pair of parallel sides does a quadrilateral have?" This proved misleading since any single number (e.g. 0) would be incorrect. In the main study these questions were reworded to say, for example, "What is true about the number of parallel sides in a quadrilateral?" This wording led to only small improvement.

Sequence Effects

Mean scores on the completion test were lower than mean scores on the multiple-choice test. Since there were four alternative choices for each item on the multiple-choice test, chance performance was 25%. Chance performance on the completion test was indeterminate. The fact that some completion items had strong prompts and that a supplementary word list was provided probably raised chance performance above 0%. It is unlikely, however, that it reached 25%. Taking into account the differences in chance levels, performance on the two tests may have been equivalent.

Some evidence exists that there is a facilitative sequence effect when a subject takes the multiple-choice form of a test prior to the completion form. Heim and Watts (1967) administered multiple-choice and open-ended forms of a vocabulary test to sailors. Half of the subjects took the multiple-choice form first, immediately followed by the open-ended form. For the other half of the subjects, the order was reversed. The mean scores for the two groups on the multiple-choice test were almost identical, indicating that experience gained by taking the open-ended test did not facilitate performance on the multiple-choice test. On the other hand, mean scores for the two groups on the open-ended test differed significantly, with the higher mean score being attained by the group which had previously taken the multiple-choice form.

Similar effects were noted by Deno (1968), who required college students to classify, define, and generate examples of psychological concepts. Subjects in Group I were given the classification task first, while subjects in Group II were given the definition and generation of novel examples tasks first. No differences were observed between groups on the classification task. Group I was significantly better than Group II on definition and generation of examples.

In the present experiment, the completion test was administered after the multiple-choice test. Thus, a facilitating sequence effect would be anticipated. On the other hand, the two parts of the completion test were administered three and four days, respectively, after the multiple-choice test. This time interval would be predicted to cause a decrement in performance.

Relationships Among the Subtests

In Chapter III, it was suggested that item difficulties would be expected to increase from item type 1 to item type 11, with five levels of difficulty: Level 1, types 1 and 2; Level 2, types 3, 4, and 5; Level 3, types 6 and 7; Level 4, types 8 and 9; and Level 5, types 10 and 11. Inspecting the mean item difficulties (Tables 16 and 17), we note that two departures from this order occurred for the multiple-choice test. Item type 9 was intermediate in difficulty between item types 5 and 6 and was therefore easier than had been predicted. Item type 7 was more difficult than predicted, especially for fourth-grade children. For the completion test (Tables 20 and 21), three departures from prediction occurred. Item type 11 was easier than expected, while item types 7 and 9 were harder than expected. The differences between item type 9 on the multiple-choice test and item type 9 on the completion test are especially interesting. Selecting the correct definition for a concept is an easier task than anticipated, while supplying the definition is a more difficult task.

Multivariate Analysis of Data

Various guidelines have been offered concerning minimal acceptable reliability coefficents. Relatively low reliability coefficients may still permit accurate conclusions concerning group means. Nevertheless, it is clear that no firm conclusions may be drawn from comparison of scores on tests with low reliabilities. For this reason, the primary analyses were carried out on scores for the total tests rather than the subtests.

All multivariate analyses were carried out using Finn's (1968) Multivariance computer program. Separate multivariate analyses of covariance were carried out for each grade level. Dependent variables were total score on the multiple-choice test (MT) and total score on the completion test (CT). Means and standard deviations of total test scores are presented in Table 24. The covariate was the raw score on the Paragraph Meaning test of the Stanford Achievement Test battery. This was selected as a covariate in order to reduce variability due to differences in reading ability. According to the authors (Kelley, Madden, Gardner, & Rudman, 1964), the Paragraph Meaning test measures the student's ability to comprehend connected discourse involving levels of comprehension from recognition to inference.

TABLE 24

Means and Standard Deviations of Total Test Scores by Treatment Groups for Grades 4 and 6

Grade	Test	1A	1B	2A	2B	3A	3B	4A	4B
4	MT	33.13 (11.88) N = 16	29.55 (9.56) N = 22	36.25 (11.07) N = 20	36.58 (9.79) N = 19	33.80 (9.16) N = 20	31.60 (8.48) N = 20	36.33 (10.03) N = 18	33.05 (11.21) N = 19
	CT	25.25 (11.19) N = 16	25.68 (10.29) N = 22	32.05 (12.53) N = 20	28.95 (10.36) N = 19	27.75 (11.65) N = 20	27.50 (8.56) N = 20	30.56 (9.16) N = 18	25.16 (8.21) N = 19
6	MT	42.42 (10.97) N = 19	43.73 (17.00) N = 15	42.64 (9.71) N = 14	48.18 (13.97) N = 17	40.93 (16.73) N = 15	46.31 (15.01) N = 13	47.33 (7.25) N = 18	45.07 (16.42) N = 15
	CT	38.32 (10.62) N = 19	39.33 (15.76) N = 15	40.43 (12.05) N = 14	39.47 (17.12) N = 17	37.33 (13.24) N = 15	39.54 (10.62) N = 13	39.94 (5.77) N = 18	42.20 (14.34) N = 15

Note. - Standard deviations are given in parentheses.

A multivariate regression analysis was carried out to analyze the relationship of the covariate to the dependent variables. Tables 25 and 26 contain the multivariate and univariate statistics summarizing the regression analysis. The multivariate analysis reveals that the covariate has a highly significant correlation with the dependent variables. The univariate *F*'s indicate that a significant amount of each dependent variable's variance can be predicted by the covariate. Since R^2 equals the percent of variance predicted, we can see that the amount of variance predicted ranges from 52% for fourth-grade children on the completion test to 12% for sixth-grade children on the multiple-choice test.

The relationship between the two dependent variables and the amount of that relationship due to covariation with reading ability are suggested by the following comparisons. The correlation between MT and CT is .72 for fourth grade, .81 for sixth grade. The partial correlations for the same variables after the linear effects of the covariate have been removed are .49 for fourth grade, .78 for sixth grade. Thus, total scores for the multiple-choice and completion tests are highly related for both fourth- and sixth-grade children. For fourth-grade children, however, this relationship is largely due to reading skill.

Since the number of subjects in the cells varied slightly, the multivariate analysis of covariance design is non-orthogonal. Because of this, the effects are not independent and are estimated in stepwise fashion. The effects of greatest interest are ordered last to obtain

TABLE 25

Regression Analysis of the Relationship between the Dependent Variables and the Covariate for Grade 4

Variable	Square Multiple R	Multiple R	F	P Less Than
Multiple-Choice Test	.4164	.6453	74.9035	.0001
Completion Test	.5156	.7181	111.7749	.0001

Degrees of Freedom for Hypothesis = 1

Degrees of Freedom for Error = 105

F Value for Test of No Association between Dependent Variables and Covariate = 63.3898

D.F. = 2 and 104 P Less Than .0001

TABLE 26

Regression Analysis of the Relationship between the Dependent Variables and the Covariate for Grade 6

Variable	Square Multiple R	Multiple R	F	P Less Than
Multiple-Choice Test	.1204	.3470	11.6346	.001
Completion Test	.2119	.4604	22.8585	.0001

Degrees of Freedom for Hypothesis = 1

Degrees of Freedom for Error = 85

F Value for Test of No Association between Dependent Variables and Covariate = 11.4213

D.F. = 2 and 84 P Less Than .0001

unbiased estimates of them. Results of the multivariate analysis of covariance of total test scores for Grade 4 are presented in Table 27. Estimates and tests of effects were carried out in the indicated order.

The significance level adopted in this experiment for the multivariate analyses was .05. When univariate analyses were interpreted, the alpha level was scaled down in order to control the error rate for the tests considered jointly. A strategy suggested by Miller (1966) was employed, setting the significance level for an individual *F* at α/k, where k is the number of *F* tests being interpreted. Thus, when two *F* tests were examined, the critical probability level was set at .025 in order to maintain an overall error rate of .05. When eleven *F* tests were considered, the probability level was set at .005.

There was significant variation among the mean vectors over the six class groups. Univariate *F* statistics were computed for each variable. Only the univariate *F* for the completion test score was significant, suggesting that this element of the vector accounted for most of the variation in the mean vectors. To explore the possibility that the class effect was due to differences between experimenters, a *t* test was carried out on the difference between mean scores on the completion test for subjects run by each experimenter. The pooled within-cell variance was used to obtain an estimate of the standard error of the difference. The experimenter effect was not significant ($t = .72$, $df = 106$, $p < .5$), leading to the conclusion that the significant class effect was due to differences among the class groups rather than between experimenters.

To clarify the nature of the significant class x treatment interaction, the degrees of freedom were partitioned, and separate tests were carried out on seven different class x treatment comparisons. Results suggest that the interaction is due to differences between classes in relative achievement for treatments 4A and 4B. Thus, in some classes, achievement is higher for 4A, in others for 4B. This effect occurs for both the multiple-choice and completion tests.

The effect of primary interest was the highly significant variation in mean vectors due to emphasis of relevant attribute values. The univariate F's for both dependent variables were significant, indicating that scores on both the multiple-choice and completion tests contributed to the effect. The observed mean score was 35.55 on the multiple-choice test for subjects whose lessons emphasized relevant attribute values and 31.90 for subjects whose lessons did not emphasize relevant attribute values. The observed mean score on the completion test for emphasis groups was 29.20, for the non-emphasis groups, 25.82.

The effect of number of instances and the interaction between number of instances and emphasis of relevant attribute values were not significant.

As was mentioned previously, analyses were carried out on total scores since reliabilities for the subtests were low. In order to gain a preliminary notion of which variables were affected most by the emphasis of relevant attribute values, further analyses were carried out using subtest scores as dependent variables. Mean subtest

TABLE 27

Multivariate and Univariate Analyses of Covariance

of Total Test Scores for Grade 4

Source	Multivariate Analysis			Univariate Analysis			
	F	df	Probability	Test	F	df	Probability
Class	2.2319	10, 208	<.0173*	MT CT	1.8439 3.8095	5, 105 5, 105	<.1106 <.0033*
Class X Treatment	1.5554	70, 208	<.0090*	MT CT	1.8287 1.6718	35, 105 35, 105	<.0101* <.0242*
Between Counter-balanced Groups	1.9634	8, 208	<.0526	MT CT	.7542 1.5233	4, 105 4, 105	<.5575 <.2008
Number of Instances	0.6026	2, 104	<.5493	MT CT	1.1970 0.1665	1, 105 1, 105	<.2765 <.6842
Emphasis of Attributes	5.3652	2, 104	<.0061*	MT CT	6.7978 9.1335	1, 105 1, 105	<.0105* <.0032*
Number X Emphasis	1.3668	2, 104	<.2595	MT CT	2.3614 1.6847	1, 105 1, 105	<.1274 <.1972

*Significant at the indicated level

scores for subjects receiving each level of the two independent variables are presented in Table 28. Results of the multivariate and univariate analyses for the effects of number of instances and emphasis of relevant attribute values are presented in Tables 29 and 30.

The multivariate *F*'s for the effect of number of instances were nonsignificant. The one indication of an effect is found in the significant univariate *F* for multiple-choice subtest 4. The mean score was 3.19 on this subtest for subjects who were presented with four examples, and 3.66 for subjects who were presented with eight examples.

With the eleven multiple-choice subtest scores as dependent variables, a significant multivariate *F* was obtained for the comparison between emphasis and non-emphasis groups. Subsequent examination of the univariate *F*'s indicated a significant value for multiple-choice subtest 2. Subtest 6 approached significance. A significant multivariate *F* was also obtained for the comparison between emphasis and non-emphasis groups using the eleven completion subtest scores as dependent variables. Among the univariate *F*'s, only that for completion test 2 was significant.

One must remember, however, that results for all analyses carried out on subtest scores must be interpreted with caution due to low reliabilities.

Results of the multivariate analysis of covariance of total test scores for Grade 6 are presented in Table 31. There was no significant variation among the mean vectors for any of the main effects or interactions. Mean subtest scores for Grade 6 are reported in Table 32, and results of multivariate and univariate analyses of subtest

TABLE 28

Observed Mean Subtest Scores for Grade 4 Students for Each Level of Emphasis of Attribute Values and Number of Instances

Subtest	Number of Items	Standard Deviation	Non-Emphasis (N=78)	Emphasis (N=76)	Difference	4 Instances (N=77)	8 Instances (N=77)	Difference
M 1	10	1.96	5.83	6.39	+ .56	6.31	5.91	-.40
M 2	10	1.97	4.53	5.82	+1.29	5.21	5.12	-.09
M 3	7	1.28	3.08	3.30	+ .22	3.17	3.21	+.04
M 4	7	1.40	3.36	3.50	+ .14	3.19	3.66	+.47
M 5	7	1.28	2.91	2.70	- .21	2.90	2.71	-.19
M 6	7	1.19	2.12	2.79	+ .67	2.47	2.43	-.04
M 7	7	1.07	1.74	1.99	+ .25	1.94	1.79	-.15
M 8	7	1.17	2.31	2.43	+ .12	2.40	2.34	-.06
M 9	7	1.36	2.56	2.74	+ .18	2.71	2.58	-.13
M 10	7	1.25	2.12	2.33	+ .21	2.06	2.38	+.32
M 11	5	0.94	1.35	1.57	+ .22	1.40	1.51	+.11
C 1	10	1.93	5.99	6.63	+ .64	6.49	6.12	-.37
C 2	10	2.01	4.63	5.58	+ .95	5.31	4.88	-.43
C 3	7	1.56	3.50	3.89	+ .39	3.77	3.61	-.16
C 4	7	1.61	2.49	2.90	+ .41	2.68	2.73	+.05
C 5	7	1.35	2.28	2.39	+ .11	2.48	2.19	-.29
C 6	7	1.13	1.74	2.13	+ .39	2.00	1.87	-.13
C 7	7	0.69	0.97	1.11	+ .14	1.05	1.03	-.02
C 8	7	1.06	1.10	1.21	+ .11	1.14	1.17	+.03
C 9	7	0.84	0.56	0.62	+ .06	0.66	0.52	-.14
C 10	7	1.02	1.15	0.91	- .24	1.01	1.05	+.04
C 11	5	1.06	1.41	1.82	+ .41	1.45	1.75	+.30

TABLE 29

Multivariate and Univariate Analyses for Effect of Number of Instances on Subtest Scores for Grade 4

Multiple-Choice Subtests Multivariate F = 1.4647 D.F. = 11 and 95 P < 0.1579			Completion Subtests Multivariate F = 0.5939 D.F. = 11 and 95 P < 0.8296		
Subtest	Univariate F	P Less Than	Subtest	Univariate F	P Less Than
M 1	0.2016	0.6544	C 1	0.0510	0.8218
M 2	0.3941	0.5316	C 2	0.3002	0.5850
M 3	0.4019	0.5275	C 3	0.0425	0.3372
M 4	9.3209	0.0029*	C 4	0.5779	0.4489
M 5	0.0044	0.9474	C 5	0.9193	0.3399
M 6	0.0031	0.9559	C 6	0.0021	0.9633
M 7	0.0401	0.8416	C 7	0.0056	0.9405
M 8	0.0036	0.9526	C 8	1.1287	0.2905
M 9	0.0648	0.7995	C 9	0.0307	0.8613
M 10	3.2813	0.0730	C 10	0.5059	0.4785
M 11	0.4394	0.5089	C 11	3.6097	0.0602

Degrees of Freedom for Hypothesis = 1
Degrees of Freedom for Error = 105

Degrees of Freedom for Hypothesis = 1
Degrees of Freedom for Error = 105

*Significant at the indicated level

TABLE 30

Multivariate and Univariate Analyses for Effect of Emphasis of Attributes on Subtest Scores for Grade 4

Multiple-Choice Subtests Multivariate F=2.4944 D.F.=11 and 95 P < 0.0085			Completion Subtests Multivariate F=2.5317 D.F.=11 and 95 P < 0.0076		
Subtest	Univariate F	P Less Than	Subtest	Univariate F	P Less Than
M 1	2.1942	0.1416	C 1	3.9376	0.0499
M 2	13.6790	0.0004*	C 2	10.2012	0.0019*
M 3	0.6621	0.4177	C 3	1.4793	0.2267
M 4	0.0370	0.8478	C 4	3.4269	0.0670
M 5	0.9658	0.3280	C 5	0.4639	0.4973
M 6	7.9893	0.0057	C 6	4.1183	0.0450
M 7	0.8252	0.3658	C 7	1.2206	0.2718
M 8	0.3229	0.5711	C 8	0.3267	0.5689
M 9	0.5440	0.4625	C 9	0.1091	0.7419
M 10	0.9198	0.3398	C 10	2.9469	0.0890
M 11	1.2750	0.2614	C 11	6.4689	0.0125

Degrees of Freedom for Hypothesis = 1
Degrees of Freedom for Error = 105

*Significant at the indicated level

Degrees of Freedom for Hypothesis = 1
Degrees of Freedom for Error = 105

*Significant at the indicated level

TABLE 31

Multivariate and Univariate Analyses of Covariance

of Total Test Scores for Grade 6

Source	Multivariate Analysis			Univariate Analysis			
	F	df	Probability	Test	F	df	Probability
Class	.4252	8, 168	<.9048	MT	.5556	4, 85	<.6955
				CT	.1577	4, 85	<.9590
Class X Treatment	.7185	56, 168	<.9239	MT	.3825	28, 85	<.9974
				CT	.4034	28, 85	<.9960
Between Counter-balanced Groups	1.4801	8, 168	<.1679	MT	2.0465	4, 85	<.0951
				CT	.5539	4, 85	<.6968
Number of Instances	0.0936	2, 84	<.9108	MT	.1804	1, 85	<.6722
				CT	.1528	1, 85	<.6969
Emphasis of Attributes	0.5269	2, 84	<.5924	MT	1.0450	1, 85	<.3096
				CT	.4980	1, 85	<.4824
Number X Emphasis	0.9926	2, 84	<.3750	MT	.1920	1, 85	<.6624
				CT	1.4057	1, 85	<.2391

TABLE 32

Observed Mean Subtest Scores for Grade 6 Students for Each Level of Emphasis of Relevant Attribute Values and Number of Instances

Subtest	Number of Items	Standard Deviation	Non-Emphasis (N=62)	Emphasis (N=64)	Difference	4 Instances (N=65)	8 Instances (N=61)	Difference
M 1	10	1.60	7.84	8.02	+ .18	8.06	7.82	-.24
M 2	10	2.04	6.71	8.00	+1.29	7.53	7.16	-.37
M 3	7	1.36	3.69	3.55	- .14	3.72	3.51	-.21
M 4	7	1.43	4.50	4.37	- .13	4.35	4.52	+.17
M 5	7	1.75	3.45	3.48	+ .03	3.46	3.48	+.02
M 6	7	1.51	3.19	3.56	+ .37	3.25	3.52	+.27
M 7	7	1.61	2.89	3.08	+ .19	2.94	3.03	+.09
M 8	7	1.46	2.79	3.13	+ .34	2.82	3.11	+.29
M 9	7	1.51	3.56	3.83	+ .27	3.66	3.74	+.08
M 10	7	1.57	2.74	3.13	+ .39	2.71	3.18	+.47
M 11	5	1.06	1.82	1.83	+ .01	1.77	1.89	+.12
C 1	10	1.68	7.89	8.14	+ .25	8.02	8.02	.00
C 2	10	1.50	7.02	7.58	+ .56	7.28	7.33	+.05
C 3	7	1.29	4.76	4.97	+ .21	4.86	4.87	+.01
C 4	7	1.75	3.74	3.59	- .15	3.78	3.54	-.24
C 5	7	1.48	3.21	3.20	- .01	3.37	3.03	-.34
C 6	7	1.53	2.84	3.33	+ .49	2.95	3.23	+.28
C 7	7	0.78	1.24	1.31	+ .07	1.28	1.28	.00
C 8	7	1.50	1.77	2.38	+ .61	1.98	2.18	+.20
C 9	7	1.65	1.94	1.91	- .03	1.85	2.00	+.15
C 10	7	1.31	1.90	1.84	- .06	1.74	2.02	+.28
C 11	5	1.35	2.27	2.20	- .07	2.20	2.28	+.08

scores in Tables 33 and 34. With the multiple-choice subtest scores as dependent variables, a significant multivariate F was obtained for the comparison between emphasis and non-emphasis groups. Examination of the univariate F's for the multiple-choice subtests revealed a significant value only for item type 2. With completion subtest scores as dependent variables, no significant effects were noted.

Lesson Statistics

Although theory and practice in programed instruction has in general stressed the need for low error rate, recent thinking has been to discount error rate as a validating criterion for a program. As Lumsdaine (1965) has pointed out, an end-of-program test following a retention interval is probably a more accurate measure of program effectiveness than error rate since it is free of immediate context cues. The primary usefulness of error rate is in suggesting specific aspects of a program which may need revision. On the other hand, it is unlikely that a program with a very high error rate is an effective program. Therefore, error rate may be considered a necessary but not sufficient condition for a good program. The error rates for all lessons used in the present experiment appeared to be acceptably low for subjects at both grade levels. At the fourth-grade level, total error rates varied from 2.2% to 12.5%. At the sixth-grade level, total error rates ranged from 2.2% to 9.3%.

Since lessons varied in length, it is important to note the average length of time spent in study of the lessons by each treatment

TABLE 33

Multivariate and Univariate Analyses for Effect of Number of Instances on Subtest Scores for Grade 6

Multiple-Choice Subtests Multivariate F=1.1373 D.F.=11 and 75 P < 0.3455			Completion Subtests Multivariate F=0.7738 D.F.=11 and 75 P < 0.6648		
Subtest	Univariate F	P Less Than	Subtest	Univariate F	P Less Than
M 1	0.7223	0.3978	C 1	0.0117	0.9140
M 2	0.5006	0.4812	C 2	0.4702	0.4948
M 3	0.8601	0.3564	C 3	0.0230	0.8800
M 4	0.4371	0.5104	C 4	0.7253	0.3968
M 5	0.0252	0.8743	C 5	1.3714	0.2449
M 6	0.5504	0.4603	C 6	1.4120	0.2381
M 7	0.1634	0.6871	C 7	0.0057	0.9399
M 8	1.2311	0.2704	C 8	1.2552	0.2658
M 9	0.1109	0.7400	C 9	0.2849	0.5950
M 10	3.0339	0.0852	C 10	1.0104	0.3177
M 11	0.4099	0.5238	C 11	0.0841	0.7726

Degrees of Freedom for Hypothesis = 1
Degrees of Freedom for Error = 85

Degrees of Freedom for Hypothesis = 1
Degrees of Freedom for Error = 85

TABLE 34

Multivariate and Univariate Analyses for Effect of Emphasis of Attributes on Subtest Scores for Grade 6

Multiple-Choice Subtests Multivariate F=2.0322 D.F.=11 and 75 P < 0.0367			Completion Subtests Multivariate F=1.4343 D.F.=11 and 75 P < 0.1756		
Subtest	Univariate F	P Less Than	Subtest	Univariate F	P Less Than
M 1	0.2364	0.6281	C 1	0.2595	0.6118
M 2	11.4656	0.0011*	C 2	4.5840	0.0352
M 3	0.5497	0.4606	C 3	0.4220	0.5178
M 4	0.2614	0.6105	C 4	0.2273	0.6348
M 5	0.2933	0.5896	C 5	0.2490	0.6191
M 6	1.3071	0.2562	C 6	3.1367	0.0802
M 7	0.2118	0.6466	C 7	0.0891	0.7661
M 8	0.9565	0.3309	C 8	3.8931	0.0518
M 9	0.4319	0.5129	C 9	0.0746	0.7854
M 10	0.8761	0.3520	C 10	0.2735	0.6024
M 11	0.0014	0.9704	C 11	0.4133	0.5221

Degrees of Freedom for Hypothesis = 1
Degrees of Freedom for Error = 85

Degrees of Freedom for Hypothesis = 1
Degrees of Freedom for Error = 85

*Significant at indicated level

group (Table 35). In interpreting the data, two facts should be taken into account. First, there was fairly wide variability between subjects in completion times. Second, there was some unreliability in the self-reports of time elapsed due to errors in telling time. Nevertheless, there is at least one clear trend in the data. Subjects having lessons emphasizing relevant attribute values spent longer times studying the lessons than the other subjects. Fourth-grade students having lessons emphasizing relevant attribute values spent an average of 70.1 minutes completing the lessons, while students having lessons not emphasizing relevant attribute values spent an average of only 50.0 minutes. For sixth-grade students the comparable times were emphasis, 63.4 minutes; non-emphasis, 46.9 minutes.

Teacher Questionnaires

All six fourth-grade teachers and four of the five sixth-grade teachers responded to the questionnaire regarding their students' knowledge of the concepts dealt with in this experiment. At fourth-grade level the number of teachers indicating study of various concepts during the past school year were: equal length, 5; right angle, 4; rectangle and square, 3; closed curve, simple curve, and opposite, 1. All of these lessons occurred five months prior to the experiment. In addition, one teacher dealt with the concepts of plane figure and parallel one week prior to the experiment. The estimates of fourth-grade teachers regarding their students' mastery of concepts were quite variable. In general, they believed their students were

TABLE 35

Mean Completion Times for Each Lesson

and Total Mean Completion Times (in minutes) by Treatment Group

		Grade 4					Grade 6				
Group		Lesson I	Lesson II	Lesson III	Lesson IV	Total	Lesson I	Lesson II	Lesson III	Lesson IV	Total
1	A	18.1	17.0	17.1	5.6	57.8	16.7	15.9	14.5	8.0	55.1
	B	17.1	16.9	11.5	5.2	50.7	15.4	15.5	12.1	6.6	49.6
2	A	21.7	19.8	16.2	19.8	77.5	14.9	15.1	10.9	16.4	57.3
	B	18.0	17.4	15.3	16.5	67.2	15.2	16.6	12.3	16.2	60.3
3	A	16.6	16.0	6.9	5.4	44.9	13.3	14.4	7.1	6.1	40.9
	B	19.2	15.6	7.3	6.0	48.1	15.5	17.2	8.3	6.6	47.6
4	A	17.3	15.7	18.7	13.3	65.0	17.8	17.1	17.5	14.5	66.9
	B	19.0	16.9	19.0	15.3	70.2	16.9	17.1	19.1	14.1	67.2

completely unfamiliar with most of the concepts, had some mastery of rectangle, square, and equal length, and had not completely mastered any of the concepts.

At sixth-grade level no teachers reported teaching any of the concepts during the past school year. They pointed out that geometry was not dealt with in the sixth-grade textbook but had been stressed in the fifth-grade text. The teachers, on the whole, believed their students had some knowledge, but not mastery, of most of the concepts. Thus, the assumption that neither fourth- nor sixth-grade children had mastered the concepts prior to the experiment was supported by teacher judgments.

Chapter VI

DISCUSSION AND SUMMARY

This study had two major objectives: (1) to devise a set of prototypic tasks which would test various aspects of concept learning, and (2) to determine the effect of two instructional variables, number of instances presented and emphasis of relevant attribute values, on performance of those tasks.

Concept Learning Tests

A set of eleven tasks was developed on the basis of analysis of the nature of a concept, review of the literature on cognitive processes in concept learning, and review of testing procedures used in previous concept learning studies. Each of the basic tasks was then differentiated into two parallel tasks, one requiring recognition of the correct answer, the other requiring production of the correct answer.

An important question to be answered is whether these tasks are equivalent measures of a single mediating construct (Johnson, 1964; Johnson & O'Reilly, 1964; Johnson & Stratton, 1966) or whether they are distinct, hierarchically-related tasks reflecting different aspects of concept learning (Dunham, Guilford, & Hoepfner, 1966; Deno, 1968; Gagné, 1968; Klausmeier, Harris, Davis, Schwenn, & Frayer, 1968). Both the multiple-choice and completion tests comprised of the eleven tasks showed high internal consistency reliabilities (.81-.87), indicating a high degree of relationship between the tasks. The degree to which the subtests based on particular tasks may be distinguished from one another may not be determined from this study. A factor analytic study would clarify the interrelationship among subtests. The fact that emphasis of relevant attribute values and increase in number of instances appeared to affect performance on some tasks more than others lends support to the belief that the tasks may be differentiated.

Item difficulties generally increased from task levels 1-11. Whether this may be regarded as evidence for the existence of a hierarchy of task complexity, however, is open to question. Kropp, Stoker, and Bashaw (1966) in testing the hierarchical nature of the Taxonomy of Educational Objectives: Cognitive Domain (Bloom, 1956) hypothesized that means of scores for taxonomy levels would decrease as the complexity level increased. With few exceptions, the pattern of means supported the hypothesis. On the other hand, simplex analysis suggested that the imputed order of complexity of taxonomy levels might be incorrect, in which case item difficulty and complexity would be discrepant.

Guttman (1954) asserted that there is no necessary relationship between the complexity of intellectual processes and the difficulty of items which are intended to measure them. Support for this assertion was provided by Crawford (1968) who found that the order of difficulty was generally different from the order of complexity for tests based on the levels of a taxonomy of intellectual processes. We may therefore conclude that, although the tasks utilized in this experiment may form a hierarchy, mean task difficulties alone are not sufficient evidence to verify this structure. Refinement of the tests should be carried out and simplex analyses utilized to determine order of complexity.

In Chapter V it was noted that several subtests had low internal consistency reliabilities. The small number of items and wide range of item difficulties on these subtests may largely account for the low reliabilities. Expansion of these subtests might increase reliabilities and also permit more analytical differentiation of levels of concept mastery.

For example, several items might be written in the multiple-choice format of tasks 3 and 4, which require recognition of concept examples and non-examples. These items could contrast examples of the concept with non-examples differing in the number of relevant attribute values which they lack. Such a systematic construction of distractors has been suggested by Guttman and Schlesinger (1967). The concept is analyzed to determine the number of facets (relevant attribute values) which it posesses. Non-examples are classified according to the number of facets on which they differ from the concept. As the number increases, similarity decreases and the distinction between examples and non-examples becomes more obvious. Construction of distractors in this manner would also permit detection of the particular facets on which the student makes his errors.

An alternative method of presenting tasks 3 and 4 in the multiple-choice format would be to present the student with an array consisting of concept examples and non-examples and have him check those which are examples. This would reveal incomplete concepts (some examples not checked) and overgeneralization (some non-examples checked).

For the multiple-choice format of task 5, a list of concept labels could be provided and the student instructed to check all those which are correct. This would be particularly appropriate when one desires to test knowledge of more generic labels. Weaver (1966) has used this approach in an inventory which tests knowledge of the names for geometric figures. The figure is presented, and the student indicates for each of six suggested names whether it is a correct name for that figure. More than one name may be assigned to each figure.

For each concept, multiple-choice items of types 6 and 7 could be written for each relevant and irrelevant attribute.

Definitions offered on task 9 of the completion test could be assigned weighted scores on the basis of their quality. A possible system of scoring definitions has been proposed by Podell (1958) in which points are given for each common feature (relevant attribute value) and variable feature (irrelevant attribute) mentioned in the definition. It should be noted that when variable features are mentioned in a definition, the student should indicate that the feature is variable. For example, if "parallel sides" are included in a definition of quadrilateral, it should be specified that a quadrilateral may or may not have parallel sides. If it were indicated that all quadrilaterals had parallel sides, the definition would be too restrictive and, therefore, less correct.

Tabulation of the answers given to completion items may provide insights into the nature of the problems encountered by students in learning concepts. Interference due to different meanings previously associated with concept labels may be indicated by the following responses: a simple curve (a curve which can be drawn without crossing itself at any point) was called an "easy drawing;" the angles of a rectangle were described as "2 right angles and 2 left angles."

Also, interference may occur due to similarity of concept labels. A large number of students stated that a rectangle has three sides. Although it is possible that the students believed that rectangles have three sides, it is equally plausible that they confused the words "rectangle" and "triangle." The latter interpretation was suggested by the fact that one student began his definition of rectangle with the words "A triangle is . . . ."

In summary, a test comprised of items written to measure the achievement of behavioral objectives related to concept learning had high reliability. The degree to which each objective reflects a different ability a[illegible] way in which these abilities are related to one another should be further explored by factor and simplex analyses. The fact that performances on subtests were differentially affected by instructional variations suggests that the subtests may be measuring somewhat different abilities.

Number of Instances

For fourth-grade students, number of concept instances had no significant effect on total test scores. Further no significant effects were noted for the multiple-choice or completion subtests considered multivariately. The only indication of an effect was on multiple-choice subtest 4. Students who received eight instances performed significantly better than students who received only four instances. Thus, the larger number of instances facilitated recognition of non-examples.

At sixth-grade level, no effects were noted for total test scores or any of the subtest scores.

In interpreting the general lack of effect of this variable, one should observe that all groups received both positive and negative instances of the concepts. Greater effects might have resulted from use of positive instances only. In addition, four instances provide a fairly large variety of examples. It is probable that the effect of four _versus_ zero or one instance would be greater than the effect of four _versus_ eight instances.

Emphasis of Relevant Attribute Values

For fourth-grade students, emphasis of relevant attribute values had significant effects both on total test scores and on the multiple-choice and completion subtest scores considered multivariately. Univariate _F_'s for multiple-choice subtest 2 and completion subtest 2 were significant.

At sixth-grade level, multivariate F's for total test scores and for scores on the completion subtests were not significant. The multivariate F for multiple-choice subtest scores indicated the presence of an effect, which appeared to be due in large part to subtest 2.

Taken together, these results suggest that the greatest effect of emphasis of relevant attribute values is on the ability to correctly label attribute values.

Summary

Eight versions of programed lessons dealing with geometric concepts were developed which systematically varied the independent variables of number of instances and emphasis of relevant attribute values. Two hypotheses were tested: (1) level of concept mastery would increase as a function of number of instances presented, and (2) emphasis of relevant attribute values would facilitate concept learning.

Ss were 154 fourth-grade and 126 sixth-grade children. Students at each grade level were randomly assigned to the eight treatment conditions. The basic design of the experiment was a 2 x 2 factorial design with two levels of number of instances (4 or 8) and two levels of emphasis of relevant attribute values (presence or absence of attention-directing and review questions). Counterbalancing of specific lesson content in the four treatment groups of the 2 x 2 design resulted in the eight treatment conditions.

Ss studied the lessons approximately 20 minutes per day for four days. After studying the lessons, children were given a multiple-choice test and a completion test, each consisting of eleven types of questions related to concept learning.

The essential findings of the study were:

(1) Increasing the number of instances from 4 to 8 did not significantly affect overall concept mastery for either fourth- or sixth-grade children.

(2) Increasing the number of instances from 4 to 8 significantly improved recognition of concept non-examples for fourth-grade children.

(3) Emphasis of relevant attribute values significantly increased overall concept mastery for fourth-grade children. The increase in overall concept mastery for sixth-grade children was not significant.

(4) Emphasis of relevant attribute values significantly increased recognition and production of the names for attribute values by fourth-grade children. It also significantly increased recognition of the names for attribute values by sixth-grade children.

APPENDIX A

INSTRUCTIONS TO STUDENTS

Instructions to Students

My name is ____________________. I work at the University of Wisconsin in Madison. At the university, we are trying to find out the best way to help children learn geometry. We have written four lessons about geometry. Today you will study the first lesson, and Tuesday, Wednesday, and Thursday you will study the other lessons. On Friday, Monday, and Tuesday we will give you tests with questions about the things you studied in the lessons to see how well you learned them. By doing the best job you can on these lessons and tests, you will help us find out how to teach geometry so it is easier to understand.

I am going to hand out the first lesson now. When you get your booklet, please do not open it until I tell you what to do. When I call your name, you may come up and get your booklet and also take one of these pieces of green cardboard.

Now turn to the first page of the lesson where it says "word list." (Read the directions on the page to the children. Then pronounce each word and have them repeat it aloud with you. Briefly explain meaning of <u>represent</u>, e.g. a picture of a cat represents a real cat. Then, go through list in random order and ask them to find the word you pronounce. Provide feedback by telling number of word after pause.)

(Hold up a copy of Lesson I turned to the first page of the lesson, so the children can see it while you give the following instructions.)

This lesson may be a little different than other lessons you have studied. The pages in your lesson will look like this. This side of the page (point) has questions for you to answer. The other side of the page (point) gives the correct answer for each question.

When you do the lesson, you should cover the answers with the green cardboard, like this (demonstrate). After you write your answer to the question, move the cardboard down just far enough to see if the answer you wrote was correct. Then write your answer to the next question, check your answer, and so on. When you finish with one page, turn to the next page, cover the answers and then go on in the same way.

Now turn to the next page in your booklet and we'll do the first page together. Remember to cover the answers.

(Read frame 1, allowing time for children to write in their answers. Prompt them when to move answer cover, and read correct answer aloud to them.)

If you make a mistake on a question, cross out the wrong answer with a single line like this (demonstrate on blackboard) and write in the correct answer below it (demonstrate on blackboard). By correcting your answers this way, we will be able to find out what questions were too hard or confusing, and we can write better lessons next time. Suppose on the last question, "What are the names of the points?", I had answered "P, Q, and R." (Write P, Q, and R on board.) How should I correct it? (Have children tell you how to correct it and follow their directions.)

Does anyone have any questions? You will go through the rest of the lesson by yourself now. If you have any questions or come to any words that you do not know, raise your hand and I will help you. Remember to correct your answers by drawing a single line through the wrong answer and writing the correct answer below it.

Write the time it is now (tell them the correct time) on the front cover of your booklet where it says "starting time." Then, finish the lesson. When you are done, write the time that you finish on the front cover where it says "finishing time." Take your time in doing the lesson so that you will understand it and be able to answer questions about it later on.

APPENDIX B

TEACHER QUESTIONNAIRE:

KNOWLEDGE OF GEOMETRY CONCEPTS

Knowledge of Geometry Concepts

1. Prior to the experiment on April 28, 1969, which of the following geometry concepts had your class studied during the current school year?

	YES	NO	If yes, please give approximate date.
closed curve			
simple curve			
plane figure			
parallel			
equal length			
right angle			
opposite			
adjacent			
number of sides			
quadrilateral			
parallelogram			
kite (in technical sense)			
trapezoid			
rectangle			
rhombus			
square			

2. Prior to the experiment on April 28, 1969, what degree of mastery do you think the majority of students in your class possessed regarding the following concepts?

	Completely Unfamilar	Some knowledge, but not mastery	Mastery
closed curve			
simple curve			
plane figure			
parallel			
equal length			
right angle			
opposite			
adjacent			
number of sides			
quadrilateral			
parallelogram			
kite (in technical sense)			
trapezoid			
rectangle			
rhombus			
square			

REFERENCES

Adams, J. A. Multiple versus single problem training in human problem solving. Journal of Experimental Psychology, 1954, 48, 15-18.

Allen, R. R., Feezel, J. D., Kauffeld, F. J., & Harris, M. L. The development of the Wisconsin Tests of Testimony and Reasoning Assessment (WISTTRA). Technical Report Number 80. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1969.

Amster, H. Effect of instructional set and variety of instances on children's learning. Journal of Educational Psychology, 1966, 57, 74-85.

Amster, H., & Marascuilo, L. Effect of type of pretraining and variety of instances on children's concept learning. Journal of Experimental Child Psychology, 1965, 2, 192-204.

Ausubel, D. P. Educational psychology: A cognitive view. New York: Holt, Rinehart and Winston, 1968.

Baker, F. B., & Martin, T. J. FORTAP: A FORTRAN test analysis package. Madison: Wisconsin Research and Development Center for Cognitive Learning, 1968.

Berzonsky, M. D. The effect of the number and level of conceptualization of the defining attributes on the difficulty of a conjunctive concept. Paper presented at the meeting of the American Educational Research Association, Chicago, February 1968.

Bloom, B. S. (Ed.) Taxonomy of educational objectives: Handbook 1, cognitive domain. New York: McKay, 1956.

Blount, N. S., Klausmeier, H. J., Johnson, S. L., Fredrick, W. C., & Ramsay, J. G. The effectiveness of programed materials in English syntax and the relationship of selected variables to the learning of concepts. Technical Report Number 17. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1967.

Bruner, J. S., Goodnow, J. J., and Austin, G. A. A study of thinking. New York: John Wiley, 1956.

Callantine, M. F., & Warren, J. M. Learning sets in human concept formation. Psychological Reports, 1955, 1, 363-367.

Crawford, W. R. Item difficulty as related to the complexity of intellectual processes. Journal of Educational Measurement, 1968, 5, 103-107.

Deno, S. L. Assessing subject matter concepts: Item forms and the effect of sequence. Paper presented at the meeting of the American Educational Research Association, Chicago, February 1968.

Dunham, J. L., Guilford, J. P., & Hoepfner, R. Abilities pertaining to classes and the learning of concepts. Reports from the Psychological Laboratory, Number 39. Los Angeles: University of Southern California, 1966.

Ebel, R. L. Measuring educational achievement. Englewood Cliffs, N.J.: Prentice-Hall, 1965.

Finn, J. D. Multivariance--univariate and multivariate analysis of variance and covariance: A FORTRAN IV program. Version 4. Buffalo: Department of Educational Psychology, State University of New York at Buffalo, June 1968.

Frayer D. A., Fredrick, W. C., & Klausmeier, H. J. A schema for testing the level of concept mastery. Working Paper Number 16. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1969.

Fredrick, W. C. The effects of instructions, concept complexity, method of presentation, and order of concepts upon a concept attainment task. Technical Report Number 3. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1966.

Fredrick, W. C. Information processing and concept learning at grades 6, 8, and 10 as a function of cognitive style. Technical Report Number 44. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1968.

Gagné, R. M. The conditions of learning. New York: Holt, Rinehart and Winston, 1965.

Gagné, R. M. Learning hierarchies. Presidential address to Division 15, American Psychological Association, San Francisco, August 1968.

Gagné, R. M., & Bassler, O. C. Study of retention of some topics of elementary nonmetric geometry. Journal of Educational Psychology, 1963, 54, 123-131.

Gelfand, S. Effects of prior associations and task complexity upon the identification of concepts. Psychological Reports, 1958, 4, 567-574.

Glaser, R. Concept learning and concept teaching. In R. M. Gagné & W. J. Gephart (Eds.), Learning research and school subjects. Itasca, Ill.: Peacock Publishers, 1968. Pp. 1-36.

Guilford J. P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Guttman, L. A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), Mathematical thinking in the social sciences. Glencoe, Ill.: Free Press, 1954. Pp. 258-348.

Guttman, L., & Schlesinger, I. M. Systematic construction of distractors for ability and achievement test items. Educational and Psychological Measurement, 1967, 27 569-580.

Heim, A. W., & Watts, K. P. An experiment on multiple-choice versus open-ended answering in a vocabulary test. British Journal of Educational Psychology, 1967, 37, 339-346.

Hovland, C. I. A "communication analysis" of concept learning. Psychological Review, 1952, 59, 461-472.

Hoyt, Cyril. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, 153-160.

Inhelder, B., & Piaget, J. La genèse des structures logiques élémentaires. Neuchâtel: Delachaux et Niestlé, 1959. (Kofsky, E. A scalogram study of classificatory development. Child Development, 1966, 37, 191-204.)

Johnson, D. M. Cognitive structures and intellectual processes. In A. H. Passow (Ed.), Intellectual development: Another look. Washington, D.C.: Association of Supervision and Curriculum Development, 1964. Pp. 27-39.

Johnson, D. M., & O'Reilly, C. A. Concept attainment in children: Classifying and defining. Journal of Educational Psychology, 1964, 55, 71-74.

Johnson, D. M., & Stratton, R. P. Evaluation of five methods of teaching concepts. Journal of Educational Psychology, 1966, 57, 48-53.

Jones, D. L. Relationships between concept learning and selected ability test variables for an adult population. Technical Report Number 51. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1968.

Kalish, P. W. Concept attainment as a function of monetary incentives, competition, and instructions. Technical Report Number 8. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1966.

Kelley, T. L., Madden, R., Gardner, E. F., & Rudman, H. C. Stanford Achievement Test. Intermediate I, Form X. Intermediate II, Form W. Directions for Administering. New York: Harcourt, Brace & World, 1964.

Kennedy, B. J. Motivational effects of individual conferences and goal setting on performance and attitudes in arithmetic. Technical Report Number 61. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1968.

Klausmeier, H. J., & Goodwin, W. Learning and human abilities: Educational psychology. (2nd ed.) New York: Harper and Row, 1966.

Klausmeier, H. J., Harris, C. W., Davis, J. K., Schwenn, E., & Frayer, D. A. Strategies and Cognitive Processes in Concept Learning. Final Report, Cooperative Research Project No. 2850. Madison: University of Wisconsin, 1968.

Klausmeier, H. J., Sterrett, B. E., Frayer, D. A., Lewis, S. B., Lee, V. W., & Bavry, J. L. Concept learning: A bibliography, 1950-1967. Technical Report Number 82. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1969.

Kofsky, E. A scalogram study of classificatory development. Child Development, 1966, 37, 191-204.

Kropp, R. P., Stoker, H. W., & Bashaw, W. L. The construction and validation of tests of the cognitive processes as described in the Taxonomy of Educational Objectives. Final report, Cooperative Research Project No. 2117. Tallahassee: Institute of Human Learning and Department of Educational Research and Testing, Florida State University, 1966.

Lemke, E. A., Klausmeier, H. J., & Harris, C. W. Relationship of selected cognitive abilities to concept attainment and information processing. Journal of Educational Psychology, 1967, 58, 27-35.

Lumsdaine, A. A. Assessing the effectiveness of instructional programs. In R. Glaser (Ed.), Teaching Machines and programed learning, II: Data and directions. Washington, D.C.: National Education Association, Department of Audiovisual Instruction, 1965.

Lynch, D. O. Concept identification as a function of instructions, labels, sequence, concept type, and test item type. Technical Report Number 6. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1966.

Manley, M. B. A factor analytic study of three types of concept attainment tasks. Research Bulletin (RB-65-31) and ONR Technical Report. Princeton, N. J.: Educational Testing Service, 1965.

Miller, R. G., Jr. Simultaneous statistical inference. New York: McGraw-Hill, 1966.

Morrisett. L., Jr., & Hovland, C. I. A comparison of three varieties of training in human problem solving. Journal of Experimental Psychology, 1959, 58, 52-55.

Nuthall, G. A. An experimental comparison of alternative strategies for teaching concepts. American Educational Research Journal, 1968, 4, 561-584.

Phelan, J. G. Inhibiting effects of attempts to verbalize on attainment and use of complex and conjunctive concepts. Psychological Reports, 1968, 22, 785-793.

Podell, H. A. Two processes of concept formation. Psychological Monographs, 1958, 72 (15, Whole No. 468).

Podell, H. A., & Carter, H. D. Effects of Cognitive Set and the Variety of Relevant Experience on Concept Formation in Children. Final Report, Cooperative Research Project No. 1459. Berkeley: University of California, 1963.

Ramsay, J. G. The attainment of concepts from figural and verbal instances, by individuals and pairs. Technical Report Number 4. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1965.

Rasmussen, E. A., & Archer, E. J. Concept identification as a function of language pretraining and task complexity. Journal of Experimental Psychology, 1961, 61, 437-441.

Remstad, R. C. Optimizing the response to a concept attainment test through sequential classroom experimentation. Technical Report Number 85. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1969.

Shore, E., & Sechrest, L. Concept attainment as a function of number of positive instances presented. Journal of Educational Psychology, 1961, 52, 303-307.

Smuckler, N. S. Concept formation as a function of method of presentation and ratio of positive to negative instances. Technical Report Number 27. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1966.

Steffe, L. P. The performance of first grade children in four levels of conservation of numerousness and three IQ groups when solving arithmetic addition problems. Technical Report Number 14. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1966.

Steffe, L. P., & Parr, R. B. The development of the concepts of ratio and fraction in the fourth, fifth and sixth years of the elementary school. Technical Report Number 49. Wisconsin Research and Development Center for Cognitive Learning, The University of Wisconsin, Madison, 1968.

Stephens, J. M. Educational psychology: The study of educational growth. (Rev. ed.) New York: Henry Holt, 1956.

Stern, C. Labeling and variety in concept identification with young children. Journal of Educational Psychology, 1965, 56, 235-240.

Trabasso, T., & Bower, G. H. Attention in learning: Theory and research. New York: John Wiley, 1968.

Weaver, J. F. Focal Point: Levels of geometric understanding: An exploratory investigation of limited scope. Arithmetic Teacher, 1966, 13, 322-332.

Wittrock, M. C., & Keislar, E. R. Verbal cues in the transfer of concepts. Journal of Educational Psychology, 1965, 56, 16-21.

Wittrock, M. C., Keislar, E. R., & Stern, C. Verbal cues in concept identification. Journal of Educational Psychology, 1964, 55, 195-200.