DOCUMENT RESUME

ED 056 882 SE 012 653

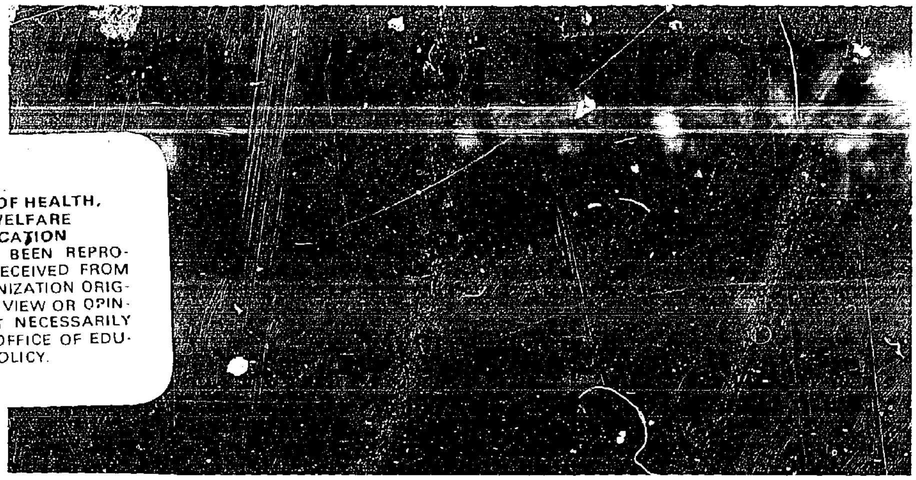
AUTHOR Weinstein, Melissa Starbuck
TITLE Effects of Training on the Concepts of Water Level and Horizontality in the Classroom.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
REPORT NO TR-128
BUREAU NO BR-5-0216
PUB DATE May 70
CONTRACT OEC-5-10-154
NOTE 92p.

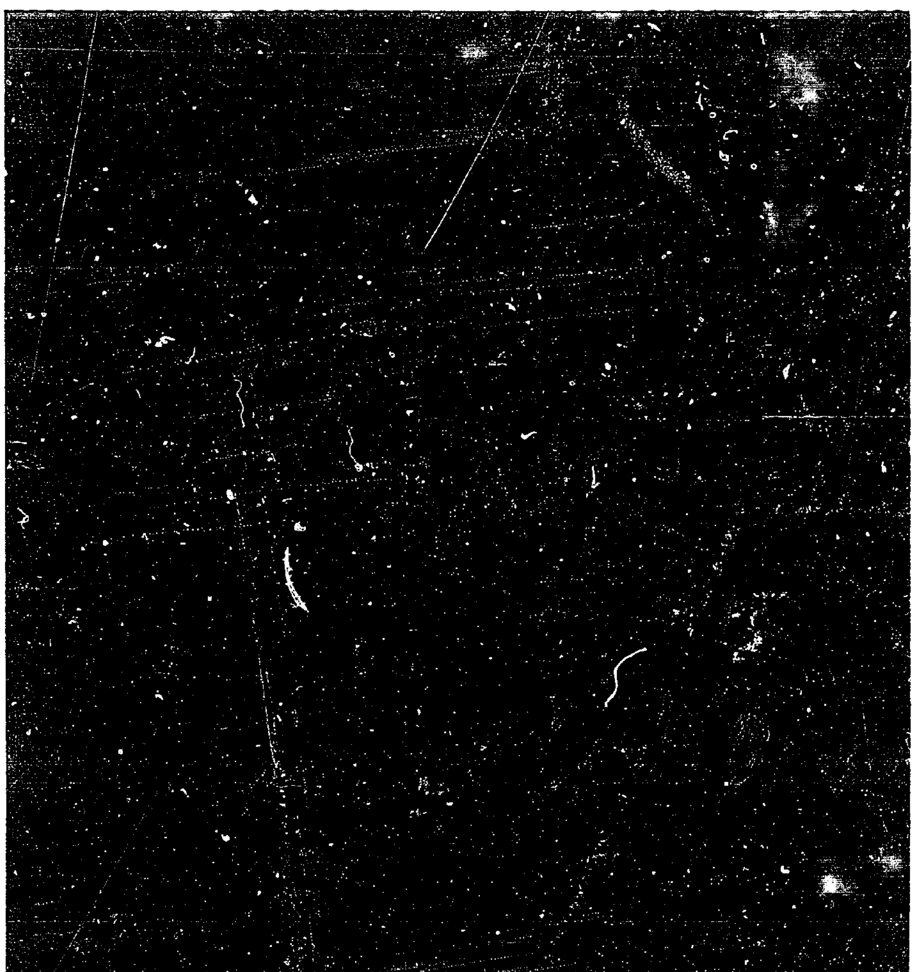
EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Concept Formation; *Elementary School Mathematics; Instruction; *Learning; Perception; *Research; Space Orientation; *Visual Learning

ABSTRACT

This experiment was designed to see if classroom instruction in the concept of water level and horizontality can improve students' knowledge of these concepts. The sample consisted of a kindergarten and a second grade class from one school and a first grade class from another school. Each class was divided into three groups. The first group was given three 20-minute training sessions, the same for each grade. The second group was given a similar treatment in which the words "horizontal" and "vertical" were used and stressed. The third group was a control. All nine groups were given the same test of 16 experimenter-designed, diagram-completion items before and after treatments. The pretest showed substantial differences between the nine groups, so treatment effects were first analyzed by a rank order test on group means. The data supported the prediction that experimental groups would perform better than control groups, but detected no difference between two versions of the experimental treatment. Three multivariate analyses of variance were also performed in an attempt to eliminate the effect of initial group differences. (MM)

BR 5-0216





EC 012 655

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

Technical Report No. 128

EFFECTS OF TRAINING ON THE CONCEPTS OF WATER LEVEL AND HORIZONTALITY IN THE CLASSROOM

Report from the Project on Analysis of Mathematics Instruction

by Melissa Starbuck Weinstein

Robert E. Davidson, Assistant Professor of Educational Psychology
Chairman of the Examing Committee

Thomas A. Romberg and John G. Harvey
Principal Investigators

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

May 1970

NATIONAL EVALUATION COMMITTEE

Samuel Brownell
Professor of Urban Education
Graduate School
Yale University

Henry Chauncey
President
Educational Testing Service

Elizabeth Koontz
Wage and Labor Standards Administration, U.S. Department of Labor, Washington

Patrick Suppes
Professor
Department of Mathematics
Stanford University

Launor F. Carter
Senior Vice President on Technology and Development
System Development Corporation

Martin Deutsch
Director, Institute for Developmental Studies
New York Medical College

Roderick McPhee
President
Punahou School, Honolulu

***Benton J. Underwood**
Professor
Department of Psychology
Northwestern University

Francis S. Chase
Professor
Department of Education
University of Chicago

Jack Edling
Director, Teaching Research Division
Oregon State System of Higher Education

G. Wesley Sowards
Director, Elementary Education
Florida State University

RESEARCH AND DEVELOPMENT CENTER POLICY REVIEW BOARD

Leonard Berkowitz
Chairman
Department of Psychology

Russell J. Hosler
Professor, Curriculum and Instruction

Stephen C. Kleene
Dean, College of Letters and Science

B. Robert Tabachnick
Chairman, Department of Curriculum and Instruction

Archie A. Buchmiller
Deputy State Superintendent
Department of Public Instruction

Clauston Jenkins
Assistant Director
Coordinating Committee for Higher Education

Donald J. McCarty
Dean
School of Education

Henry C. Weinlick
Executive Secretary
Wisconsin Education Association

Robert E. Grinder
Chairman
Department of Educational Psychology

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational Psychology

Ira Sharkansky
Associate Professor of Political Science

M. Crawford Young
Associate Dean
The Graduate School

EXECUTIVE COMMITTEE

Edgar F. Borgatta
Brittingham Professor of Sociology

Robert E. Davidson
Assistant Professor, Educational Psychology

Russell J. Hosler
Professor of Curriculum and Instruction and of Business

Wayne Otto
Professor of Curriculum and Instruction (Reading)

Anne E. Buchanan
Project Specialist
R & D Center

Frank H. Farley
Associate Professor, Educational Psychology

***Herbert J. Klausmeier**
Director, R & D Center
Professor of Educational Psychology

Robert G. Petzold
Associate Dean of the School of Education
Professor of Curriculum and Instruction and of Music

Robin S. Chapman
Research Associate
R & D Center

FACULTY OF PRINCIPAL INVESTIGATORS

Vernon L. Allen
Professor of Psychology

Frank H. Farley
Associate Professor of Educational Psychology

James Moser
Assistant Professor of Mathematics Education; Visiting Scholar

Richard L. Venezky
Assistant Professor of English and of Com

Ted Czajkowski
Assistant Professor of Curriculum and Instruction

Lester S. Golub
Lecturer in Curriculum and Instruction and in English

Wayne Otto
Professor of Curriculum and Instruction (Reading)

Alan Voelker
Assistant Professor of Curriculum and Instruction

Robert E. Davidson
Assistant Professor of Educational Psychology

John G. Harvey
Associate Professor of Mathematics and of Curriculum and Instruction

Milton O. Pella
Professor of Curriculum and Instruction (Science)

Larry Wilder
Assistant Professor of Curriculum and Instruction

Gary A. Davis
Associate Professor of Educational Psychology

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational Psychology

Thomas A. Romberg
Associate Director, R & D Center
Professor of Mathematics and of Curriculum and Instruction

Peter Wolff
Assistant Professor of Educational Psychology

M. Vere DeVault
Professor of Curriculum and Instruction (Mathematics)

Donald Lange
Assistant Professor of Curriculum and Instruction

B. Robert Tabachnick
Chairman, Department of Curriculum and Instruction

MANAGEMENT COUNCIL

Herbert J. Klausmeier
Director, R & D Center
V.A.C. Henmon Professor of Educational Psychology

Thomas A. Romberg
Associate Director

James Walter
Director
Dissemination Program

Dan G. Woolpert
Director
Operations and Business

Mary R. Quilling
Director
Technical Development Program

* COMMITTEE CHAIRMAN

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 education practices. The strategy for research and development is comprehensive. It includes basic research to 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 Phase 2 of the Project on Prototypic Instructional Systems in Elementary Mathematics in Program 2. General objectives of the Program are to establish rationale and strategy for developing instructional systems, to identify sequences of concepts and cognitive skills, to develop assessment procedures for those concepts and skills, to identify or develop instructional materials associated with the concepts and cognitive skills, and to generate new knowledge about instructional procedures. Contributing to the Program objectives, the Mat[illegible], Phase 1, is developing and testing a televised course in arithmetic for Grades 1-6 which provides not only a complete program of instruction for the pupils but also inservice training for teachers. Phase 2 has a long term goal of providing an individually guided instructional program in elementary mathematics. Preliminary activities include identifying instructional objectives, student activities, teacher activities materials, and assessment procedures for integration into a total mathematics curriculum. The third phase focuses on the development of a computer system for managing individually guided instruction in mathematics and on a later extension of the system's applicability.

ACKNOWLEDGMENTS

For his help and patience I would especially like to thank Dr. Thomas Romberg, who was instrumental in the preparation of this thesis, both individually and through the mathematics project. Other members of the project, Mrs. Rochelle Meyer and Mrs. Caroline Gornowicz, were particularly helpful.

Thanks are due to Dr. Robert Davidson, my advisor, whose research orientation is different from that of this thesis, but who was appreciative of a different outlook; and also to Dr. Peter Wolff, who served as third member of my committee.

A special credit must be given to Mr. Tom Fischbach of the staff of the Wisconsin Research and Development Center for his design for the statistical analysis.

The Stoughton Public School District was most gracious to allow the research to be done it its schools.

CONTENTS

CONTENTS (Continued)

LIST OF TABLES AND FIGURES

ABSTRACT

The concepts of water level and horizontality are essential to the coordinatization of the plane. According to Piaget, the development of these concepts is very slow. The concepts are based on structures slowly developed from what is given perceptually, and not coming into the behavioral repertoire until about age seven. The purpose of this experiment was to administer a sequence of training tasks in the classroom to intact classes to see if attending to the phenomena of the concept in a structured way would bring about significant improvement on a test designed to reflect knowledge about these concepts. To see if there was a difference if the treatments used specific language labels or generalized everyday language was a secondary purpose.

The design used was a pretest-posttest, nonequivalent control group involving three classes of three grade levels, K, first and second. The training was given in three 20-minute sessions.

Three multivariate analyses of variance were performed on the data. The analysis was done in a sequence to eliminate initial class effects to be able to determine the significance of the treatments. Results showed treatment effects to be significant, but there was little differentiation between the two treatments. All grades showed significant gains, but a definite grade effect was shown.

Chapter I

INTRODUCTION

In the child's world he has frequent daily experiences of instances of the water level concept: tilted bottles, milk, water, soda, and other pourables such as sand are manipulated by the child and those around him. His very slow ability to represent the phenomenon of the horizontality of the water level is in contrast to its occurence. It is in contrast only if we assume that space is something given in the child's environment and that it is immediately given perceptually. The purpose of this research was to give the child direct experience in attending to the phenomenon in a structured way, i.e. structured in the sense that the phenomenon is pointed out as that to which he should be attending, and to give him a precise language with which to talk about it. The concept of water level, a specific application of the concept of horizontality was to be emphasized in the treatment sessions where the children were to be given training in its representation couched in the language of the larger concept of horizontality.

We are learning that the child's concept of space differs from that of the adult who has had time to build a conceptual construct about space. Piaget and Inhelder[1] study the child's conception of space as an intellec-

[1] J. Piaget and B. Inhelder The Child's Conception of Space (New York, 1967).

tual development, i.e. the development of intelligence as it works on spatial relationships. Piaget keeps separate the notions of space perception itself and the evolution of spatial representation, comparing and contrasting the two, but clearly keeping them as separate entities.

The research reported here is not in the complete spirit of the Piagetian notion of space perception. He apparently believes that if something is given in the child's environment we can automatically assume that he is attending to it, the reason the child can't adequately represent it is a problem of the slow development of intellectual processes. It was felt that we cannot make the assumption that children are attending to those things given in their environment and that this is especially true for the life-style of children in this country at this time of great absorption in the mass media. It is on this point only that there was divergence from Piaget's work. The following exposition of spatial concepts is purely Piagetian.

Spatial representations are built up through the organization of actions[2] performed on objects in space, at first motor actions and later, internalized actions which eventuate in operational systems. What Piaget wants to stress is that the effortless seeing of the adult is really the end product of long and arduous developmental construction; and that

---

2 J. H. Flavell, The Developmental Psychology of Jean Piaget (Princeton, 1963), pps. 82-83. Cognition is at all genetic levels a matter of real actions performed by the subject. According to Piaget, actions performed by the subject constitute the substance or raw material of all intellectual and perceptual adaptation. Generally, actions first occur as motor activity, which with development become progressively internalized, and later evolve into highly organized systems of internal operations. Piaget's theory permits him to see adult logical operations as sensory-motor actions which have undergone a succession of transformations, rather than as a different species of behavior.

the construction itself is more dependent upon actions than upon perceptions *per se*.[3]

The importance of the concept of horizontality becomes apparent when we discover that it is a developmental process necessary for the child so that he is able to coordinatize the plane. Until this is done, Euclidean geometry would be unintelligible. Piaget clearly shows the importance of the concept in his overall system of the development of spatial representation:

System of Reference and Horizontal-Vertical Coordinates[4]

The onto-genetic development of spatial concepts approximates their logical order; first, topological relations (concerning the object itself and its properties) and then later projective concepts (which imply a comprehensive linking together of figures in a single system, based on the co-ordination of a number of different viewpoints) and the idea of euclidean space (involving a co-ordination of objects themselves). The concepts of parallels, angles, and proportion provide the transition between the two latter systems. Assumed is the conservation of distance, together with the evolution of the notion of displacement, or congruent transformation of spatial figures.

At the outset, the co-ordinates of euclidean space are no more than a vast network embracing all objects, and merely consists of relations of order applied simultaneously to all three dimensions. However, a

---

3 *Ibid.*, pps. 327-328.

4 J. Piaget and B. Inhelder, *op cit.*, pps. 375-376; Flavell, *op cit.*, pps. 332-334.

reference frame is not simply a network composed of relations of order between the various objects themselves. It applies equally to the positions within the network as to objects occupying any of these positions and enables the relations between them to be invariantly maintained, independent of potential displacement of the objects. [illegible], the frame of reference constitutes a euclidean space after the fashion of a container, relatively independent of the mobile objects contained [illegible] the container consists of the entire assemblage of the relations of order and the intervals of distances between objects. These relations are not confined to the objects at a particular moment, but include all their successive or potential positions, linking them all together and employing certain favored positions as reference points etc. for all subsequent positions. The essential character of a reference frame resides in the possibility of co-ordinating positions and intervals without limit through constantly enlarging on the orginal system.

The Concept of Horizontality[5]

The concept of horizontality is by nature a physical one. By definition, a horizontal line is perpendicular to that taken by a freely falling body. But as Piaget points out the concept itself is unique in that it paradoxically becomes a mathematical problem the moment the child uses it to develop a co-ordinate system as a "simple tool of geometrical orientation." The rectangular ordinates of geometry are only approximate to the physical reality they represent; that is, in nature the level of a liquid is not truly horizontal but curved, plumb-lines are not absolutely parallel to each other.

---

5 Piaget and Inhelder, op cit., pps. 375-418.

> The dual nature of these physical and geometrical problems raises a question of obvious importance from the point of view of psychology, regardless of whether the answer lies in the direction of the independence or interdependence of these two factors--though in the latter case it would be necessary to establish the precise nature of this interdependence. The problem is in fact none other than that of the physical and experimental nature of mathematics as opposed to its being of an *a priori* and purely intellectual character, together with all the intermediate possibilities available between these two extremes. Now this problem constantly reoccurs in an extremely crude, but all the more impressive, form in each of the experiments we are about to describe. From the very outset, starting with the arrangement and organization of the experimental session itself, one finds oneself at grips with the interdependence of the physical and intellectual functions involved.[6]

In developing an experimental method for studying the notions of horizontality and verticality, the main data available are the schema the child uses to record his perceptions of the world. We must start with the physical so that the mathematical will be apparent.

Piaget experimentally combined the concepts of verticality and horizontality. In discussion, he says that the two are discovered in "one shot." Apparently he feels that they are discovered by the same process of finding external references; not through conceptualizing their interdependence: thus it would seem that separating them for experimental purposes would be a valid procedure. The concepts merge only after the developmental process is nearly completed.

The concept of horizontality was tested by having the children guess the waterline in variously shaped, tilted bottles presented to them at eye level, and to indicate it by gesture or drawing.

---

[6] Piaget and Inhelder, *op cit*., p. 338.

> Care is also taken to make the children draw the edge of the table or the support holding the bottle, in such a way that this horizontal, directly perceived, can assist in judging the position of the liquid.[7]

Drawings were matched with the experimental apparatus and correction or redrawing took place immediately. Variables were the bottle shapes and the form the child used to indicate his response--from gesturing by the smallest child, to a complete line drawing by the older children where bottle, waterline, base, and angle between base and bottle was to be drawn.

As with most Piagetian research, many questions remain unanswered. What was the test's duration, what was the procedure for each child, who were the subjects, how many were there, etc. Therefore, much of the subsequent research has concerned itself with the developmental stages Piaget derived as a result of his work:

Stage I — This stage is characterized by an inability to distinquish surfaces or planes (either fluids or solids). This stage lasts until age 4 or a little later.

Stage II — Generally, spatial orientation is determined by the particular configurations represented. Operationally it is broken into two substages:

Stage IIA — Water level is shown parallel with the base of the jar, movement of water is noted, and the water is indicated as a plane surface.

Stage IIB — Children are able to show the waterline as no longer parallel to the base of the vessel, by connect-

7 Ibid., p. 381.

ing the waterline to both sides of the jar but failing to coordinate it with an external reference system. When the jar is inverted, the waterline is made horizontal.

Transition stage between IIB and IIIA. The child is able to predict the level of the liquid when it is parallel with the sides of the vessel.

Stage III Culminates with the discovery of the horizontal axis as part of the coordinate system. Begins at 7-8 years and may extend into the 9-12 year. This is also broken into two substages:

Stage IIIA Begins at 7-8 years and may last until the age of 9, during which the principle becomes gradually applied to all cases, though at the beginning the level is often made oblique, ignoring reference points external to the jar.

Stage IIIB This stage begins at about age 9, and develops so that the immediate prediction of horizontal and vertical is part of an overall system of coordinates.

There is a great spread in the ages of the subjects Piaget uses to illustrate the various stages, e.g. Stage I is illustrated by examples from ages (4;6) - (7;1); Substage IIA from (5;10) - (7;9); and Substage IIB from (5;1) - (7;0). In children, this age spead would usually bring great changes. Further, we do not know if the children progressed through the stages identically for the concepts of horizontal and vertical. And finally, there is little or nothing to indicate the causes of interor intra-individual differences.

Piaget postulates great change in all realms of intellectual development in children of this age (5 - 7 years). Clearly, spatial concepts are included in this development. What the exact nature of the correlations and inter-relationships would be can only be speculated upon.

Chapter II

RELATED RESEARCH

There have been a few replications of Piaget and Inhelder's work on horizontal representation; however, these have usually been modified or embedded in larger studies on spatial representation. Rivoire (1961), Dodwell (1963), Shantz and Smock (1966), and Smedslund (1963) conducted studies that were more in the nature of replication, while Beilin, Kagan, and Rabinowitz (1966) did a study designed to modify water level representation. There was no research reported that had been done under actual classroom conditions.

Smedslund (1963)

Smedslund carried out a limited replication of the water level experiment with 27 children, 5 - 7 years, from a nursery in Oslo, using a pretest, observation, posttest design. There are two types of test questions: "choice" questions, i.e. picking the correct representation of water level from eight models, and drawing in the waterline in a line drawing of a tilted bottle. The two types did not correlate well, but (given the small sample) the significance of this was not determined. Smedslund's results generally support the findings of Piaget and Inhelder: there is a general absence of a representation of horizontality in that age group; Piaget's intermediate stage was identified; there was an absence of learning in children with no traces of concept initially and only limited improvement from those with initial traces of the concept. This latter supports the point of view that learning cannot be conceptualized in terms of simple empiricism, where the organism acquires habits or expectancies as a function of direct contact with the external world. A more reasonable view would be that every contact is patterned accordingly to the existing schema of the subject. Therefore, when the child makes "mistakes," there exists for him no contradictions.

Schantz and Smock (1966)

Shantz and Smock worked with 20 first graders to test the following hypotheses: 1. That every child using the coordinate system also conserves distance, and every child who is unable to conserve distance is unable to demonstrate the coordinate system concept. The results supported this hypothesis. 2. To determine the comparability of data relevant to two spatial concepts (distance conservation and the coordinate system) derived from two-dimensional (drawings) to three-dimensional stimuli (objects). Both Dodwell and Rivoire suggest that the variability in age of emergence of the coordinate system may be due in part to the way in which the task is presented, specifically a differential performance was associated with the use of either objects or pictures as simuli. However, an ANOVA of correct responses was not highly significant ($p < .12$). There was some indication of an order effect, suggesting that the training of spatial concepts may be most effective with the manipulation of objects preceding the two-dimensional presentation of tasks.

Dodwell (1963)

Dodwell attempted a general assessment of the generality of spa[illegible]al concepts as reported by Piaget and Inhelder: examing age trends, the consistency of activities for each of the stages, and to evaluate these factors as evidence for a theory of cognitive development. Each of the 194 children (5;1) - (11;3), completed 36 items. One of the subgroups of the test dealt with horizontal and vertical axes. All the stages described by Piaget were found. Dodwell's findings suggest that age limits for the stages cannot be precisely identified, e.g. his age range for Stage I was 5.8 - 10.3, the mode (N = 15) being 6.3. The enormously predonderant "mixed" category suggests very little regularity in the developmental pattern. There was a considerable correlation between total score and age, and an even more marked correlation with mental age. Dodwell in conclusion makes two points: that the "mixed" category, which numerically predominated, needs to be further investigated, and that factors which obviously could disrupt development need to be used in a way consonant with developmental theory.

Beilin, Kagan, and Rabinowitz (1966)

Beilin, Kagan, and Rabinowitz attempted to determine whether language and perceptual experience can play a significant role in symbolic imagery. A pretest-training-posttest-transfer design was used. The tests were based on an anticipation method. There were two training procedures: one characterized as perceptual, the other as verbal and which involved the use of a programmed instruction booklet. Second graders (N = 180) ranging from 6.2 - 8.2 years were the subjects. They were divided into nine groups [3 control, 2 perceptual (P - motor, P - no motor response), and 3 verbal (V - water level, V - horizontal - water level, V - water level - horizontal)] based on varying the generality of the concept between the water level principle and the concept of horizontality, varying the order of their presentation, and whether or not the trial was recorded.

Results showed that anticipation imagery indicated in water level representation was improved through training; and that on the whole, perceptual training was more effective than verbal training. This suggests that water level representation is more dependent upon non-verbal than verbal mediating processes.

The horizontal program did lead to stage level change. There was improvement in both specific concept training (water level) and general concept training (horizontality), the latter showing weaker results. There was also evidence to support the Piaget and Inhelder notion that there is a relationship between symbolic imagery and operativity.

Upon analysis, it was found that the oblique positions presented greater difficulty. They suggest further research to find whether or not this problem is reducible to a perceptual process difficulty.

There was also evidenced lack of transfer between tasks utilizing the same principle.

Chapter III

THE PROBLEM

In its work the staff of the Project on Prototypic Instructional Systems in Elemen:ary Mathematics, under whose auspices this research was done, is seeking to integrate geometry into the main instructional mathematics curriculum and to develop it consistently throughout the elementary grades. The coordinization of the plane is an essential step in building a base on which to teach the concepts of Euclidean geometry, and would be essential in other subject areas as well. Piaget postulates that Stage III, the stage that culminates with the discovery of the horizontal axis as part of the coordinate system, does not originate until children are 7 - 8 years old and continues from the age of nine, extending to perhaps the twelfth year. This would effectively delay the geometry program until the sixth grade of school or beyond.

However, based on evidence gathered from research, the age levels for each stage of development might be expected to be somewhat lower than those postulated by Piaget. The research also indicates that stage level achievement was governed more by other factors than the child's chronological age: perhaps one reason why so many children would fall not into any one stage but into a "mixed" group with a very large age overlap.[8]

The purpose of this experiment was to find whether specific training on the water level concept and the concept of horizontality given in a classroom situation would significantly change the performance on a test designed to show understanding of these concepts.[9] It was further postulated that the use of precise language labels would be significantly more effective in teaching these concepts. Roger Brown's discussion of the use of words as tools with which to develop a concept in a more effective way prompted the thinking that the children would be more able to learn the concepts if they were interested in them by way of new terminology.[10] Therefore, the training would be done in two sections: (1) Those children given direct training on the concepts, and (2) those given direct training on the concepts which had been given specific language labels.[11]

---

8 Dodwell, "Children's Understanding of Spatial Concepts," Logical Thinking in Children, I. E. Sigen and F. H. Hooper, editors (New York, 1968) pps. 126-134.

9 See Appendix I for the test, and below, in the discussion of the design for a discussion of its use.

10 Roger Brown, Words and Things (Glencoe, 1958), p. 207. "We take a new word as a lure for cognition because in a long experience of language we have learned that such utterances are attributes of nonlinguistic categories and that these categories are ordinarily worth knowing . . . The semantic utterance is also a selective response elicited by some array of nonlinguistic stimuli . . .;" cf. John B. Caroll, Language and Thought (Englewood Cliffs, 1964). pps. 95-97. See also, below in the discussion of the design.

11 See Appendix II for a complete description and also below, in the discussion of the design.

The questions specifically asked were:

1. Is there a significant difference between the two treatment groups designed to teach the concepts of water level and horizontality and a control group which was given no treatment at each of three grade levels (K, 1, 2)?
2. Is there a significant difference between the two treatment groups which differ only on use of verbal labels at each grade level?
3. Is there a significant grade level effect on the performance of the test?
4. And finally, is there a significant interaction between the treatments, grade level, and the time over which the treatments were given?

Chapter IV

THE DESIGN

The basic design used in this experiment is described by Stanley and Campbell and is designated by them as the pretest-posttest non-equivalent control group design.[12] In this paradigm there is a pretest followed by treatment followed by a posttest. All groups participate in the testing; the control group does not receive treatment. This experiment is a 2 factor (Grade X Treatment) completely crossed investigation. The treatments were administered to three treatment groups (including one control group) within each of three grade levels (kindergarten, first and second grades), with assignment to treatment by class within each grade, i.e. the treatments were given to intact classes within their own classrooms.

The experimental treatments were the same with the exception of the language used to describe the phenomena of the concept. In treatment 1 the language remained general, i.e. "horizontal" was described as "going straight across," being "level," or simply "going this way" and gesturing; "vertical" was described as "up and down," etc. In treatment 2, horizontal and vertical were the terms used and stressed when describing the phenomena. The treatment group, control, did not receive treatment but took the pre- and posttests.

---

12 Experimental and Quasi-experimental Designs for Research (Chicago, 1963), pps. 37-43.

The treatments consisted of three 20-minute sessions, and all grade levels received the same treatment. A teacher from the project[13] taught all the sessions. The first day was an introduction to the concept of horizontality and verticality. Instances were found in the room, and the relationship between horizontal and vertical lines was introduced though not stressed. Water level was shown to be horizontal in all positions that a plastic liter bottle half-filled with colored water was tipped. Each child was given a worksheet with line drawings of bottles (including the baseline) tipped in different directions. The children were grouped in two's, each pair given its own bottle, and were asked to draw in the waterlines on their papers. The teacher did the first two drawings with the children as a class, then they were asked to continue on their own. A bottle was left in each classroom for children to work with if they chose to.

The second day a carpenter's level was shown and explained, and the parallel drawn between it and the waterline or "bubble" formed by a very full bottle of water. The children were given their own "carpenter's level" (see appendix II for description of the apparatus) and were asked to work in pairs to find surfaces in the room which were horizontal.

The third day the original introduction was repeated, this time using transparencies; the teacher checked the responses the children made with a plastic liter bottle filled with water. Then the idea of horizontal

---

13 The teacher was Mrs. Caroline Gornowicz, project specialist, Project on Prototypic Instructional systems in elementary Mathematics and an experienced, certified elementary teacher.

using a non-gravitational vertical axis was introduced. The teacher illustrated this through instances, e.g. having a boy with a cross-ways striped shirt bend about--the stripes remain horizontal on him but he does not retain his horizontality when bending.

The five sessions of the experiment (including the pre- and posttests) were arranged so that there was either a school day or a weekend between each session; for all groups then the whole experiment took about ten days. The control group was tested on the first and tenth days.

The written test used as both pretest and posttest consisted of sixteen items. Each item was scored either "0" if the item was incorrectly marked or left undone, or "1" if the item was correct, for a possible total score of 16. The administration of the tests and all scoring was done by the experimenter. In the analysis, items 1-12 and 16 were analyzed together as part A. They were grouped together because these questions did not include the term "horizontal" in the instructions for the test. Questions 13-15 involved the term horizontal in the instructions, and were designated "B."

The items were of five different types:

1. Items 1 - 5 (page 1 of the test in appendix I) were line drawings complete with the baseline. The children were shown a bottle whose contents were obscured with a covering but which was tipped in the same way as the bottle depicted on the test page, and were asked to draw the waterline as they imagined it would look if they could see it. The bottles were pictured as standing upright, 90 degrees right, 45 degrees right, inverted and 30 degrees left. Scoring was done in accord with descriptions by Piaget in his account of this type of test item.

2. Items 6 - 9 (page 2) were of the recognition type. The children were shown the same bottle tipped in a certain direction. Out of the five depictions of a bottle with water in it they were to mark the one they thought was the correct one.

3. Items 10 - 12 (page 3) were matching items. There was one depiction of a bottle of water in a box on the left side of a row of five bottles, the matching bottle in the row was to be marked.

4. Items 13 - 15 (page 3) were the "horizontal" items. e.g. for item 13, it was asked "What box has horizontal stripes painted in it?" Similarly for items 14 - 15, but these items were complicated by a rotated axis. These items were included as a measure of the specific language approach of treatment 2.

5. Item 16 (page 5) was a general item; given five polygonal shapes, the children were asked to indicate the shape off of which a marble would not roll.

The sample of children was drawn from two schools in the Stoughton Public School District. (Total N = 206. Table 1 shows the number of subjects by grade and treatment.) It was reported that the two schools

Table 1

Number of Subjects, by Grade and Treatment, Participating in the Testing

Grade	Treatment 1	Treatment 2	Control	Total
K	24	20	23	67
1	19	20	19	58
2	27	27	27	81
Total	70	67	69	206

were comparable in every way e.g. the ratio of rural to urban, socio-economic status, etc. Kindergarten and second grade were from one school, first grade from the other. The only anticipated systematic difference was in the kindergarten classes. The morning classes were the rural children, and it was said they didn't have the "cultural enrichment" of the town children who came in the afternoon, and had proven themselves to be slower. This prediction tended to be confirmed in the pretest scores.

Assignments within grade were made in the context of the school environment. Ages of the children were not taken into account and there was no attempt to determine them; it was assumed that the age levels here would be those of a representative school system. The scheduling of library, gym, etc. were predetermined and affected the availability of the children. The teachers knew of the experiment but were not informed of the treatment grouping nor of the status of their class in the experiment. If there was not both a pretest and posttest available for any particular child, he was dropped from the sample. Absence from the treatment sessions was not recorded and not taken into consideration; however, absence from school at the time of the year of the experiment was extremely light and probably would not have proven to have significantly affected the internal validity of the design.

One of the weaknesses of the design for most testing, as mentioned by Stanley and Campbell, is the interaction that often occurs between the pretest and the treatment which for most purposes would cast doubt on the external validity of the design. For the purposes of this experiment, this weakness is considered an asset. If the children were in fact "primed" for the treatment sessions by the pretest, then it was for the good. Usually this would have had to been accomplished through some teaching strat[illegible] to have the pr[illegible]test serve as catalyst was an advantage in view [illegible] brevity of the time available for the training sessions.

The data turn[illegible] to be not amenable to analysis by the groups that the items were [illegible]igned to fall into. Items 10 - 12 were much too easy. All groups w[illegible] [illegible]ble to complete them on both the pre- and post-tests. Item 16 was [illegible] one that was responded to correctly by most of the children. Item[illegible] - 15 were possibly too confounded by the logic necessary to grasp [illegible] concept to be adequate test items for children of this age level. [illegible] this reason, they may have affected the statistical analysis negatively in terms of the hypothesis that they were reflecting the efficacy [illegible]f treatment II. It would take further testing in order to make a pos[illegible]tive statement of this effect, however.

Each class received the treatments as a whole so that the scores on the posttest are not independent. Further, assignment of classes to treatment groups was [illegible] "random" in the statistical sense. Individual students had not previously been assigned "randomly" to these classrooms. The pre-test scores [illegible]hen, were independent except for a linear class effect nested within each grade.

Given the dependencies just noted, the following analysis based on gain scores was designed by Mr. Tom Fischbach, a staff member of the Wisconsin Research and Development Center for Cognitive Learning. This analysis was performed to determine just what inferences can be gained from the data.

Chapter V

RESULTS

The following is a summary of the sequence of analyses performed on the data. The first step was to find how the classes compared on the pretest; for this a least squares estimate for class and grade effects was done. F-tests then showed that the classes were substantially different within each grade level and also across grade levels. This fact of initial intra-grade differences combined with the fact that the data on students were not independent dictated that a rank order test of the hypothesis that there were treatment effects for classes be performed.

A permutation test was performed, the results showed that the treatments were effective. Having established the effectiveness of the treatments for groups, multivariate analysis of variance on pretest and posttest scores was made using the student data as the basic unit affected by treatment. This was followed by an analysis of co-variance on posttest differences corrected for pretest differences. Finally, an analysis of variance on gain scores was also done in an attempt to eliminate the influence of the initial class effects and to establish the magnitude of the effectiveness of the treatments. Each analysis was made by parts: A (items 1 - 12 and 16) and B (items 13 - 15), so that the specific differences in the treatments could be shown. Additionally, to help clarify treatment effects, item data is also reported for each group.

I. Pretest

The raw-score class means are shown in Table 2.

Table 2

Observed Raw Score Means on the Pretest
by Treatment Group and Grade

Pretest

Grade	Treatment	A*	B**	Total***
K	1	7.38	.38	7.75
	2	6.15	.35	6.50
	C	7.87	.61	8.52
1	1	9.68	.32	10.00
	2	8.80	.30	9.10
	C	8.84	.11	8.94
2	1	10.88	.19	11.07
	2	10.59	.15	10.74
	C	11.22	.30	11.52

* Part A included items 1 - 12 and 16.
** Part B included items 13 - 15.
*** Total = A + B

As seen by inspection there are differences in these means between classes and between grade levels. Figure 1 shows the rank order of class means on the pretest. The ordering is nearly the reverse of that which the hypothesis of treatment effects would show.

Grade	Rank Order of Class Means
K	C > 1 > 2
1	1 > 2 > C
2	C > 1 > 2

Figure 1

Rank Order of Class Means on the Pretest

Any "treatment" effects or "treatment x grade" interaction effects on the pretest necessarily reflects only class effects for this test was administered prior to any treatment. The least squares estimates of these effects (Class Effects) and grade effects are shown below (see Table 3). A fixed model is assumed.

Table 3

Least Squares Estimates of Grade and Class Effects for the Pretest

Source	Estimate	Standard Error of Estimate
	GRADE EFFECTS	
Mean	9.35	0.136
Grade		
1st - K	1.76	.346
2nd - K	3.52	.319

Table 3 continued

Source	Estimate	Standard Error of Estimate
	CLASS EFFECTS	
"Treatment 1" - "Control"	-0.05	.331
"Treatment 2" - "Control"	-0.88	.334
Interactions 1	1.82	.842
Interactions 2	.33	.770
Interactions 3	2.17	.885
Interactions 4	1.24	.790

It should be noted that the direction of these effects (i.e., the "treatment" effects before actual treatment) is opposite that which might be expected after the treatments were administered, assuming that in fact the treatments did have a positive effect on test performance. Both treatment control differences are negative, showing that the control groups scored higher than did either of the treatment groups in two instances, and that there was a slight reversal in the treatment groups in the third. By taking the differences between the two effect, e.g., ("Treatment 2" - "Control") - ("Treatment 1" - "Control"), we get the estimate of "Treatment 2" - "Treatment 1" which is also negative (and undoubtedly differs from zero by a statistically significant amount).

F-tests for testing the hypothesis that the two main effects and the interaction effects are zero would also reject the null hypothesis for the main effects at the .05 level (actually the .01 level) and the interactions at the .07 level. From the results (see Table 4), it is clear that the classes differed substantially at each grade level, and of course, also between grade levels, before any treatments were administered.

Because of this substantial difference, the usual procedure of making an analysis of variance directly was postponed until the hypothesis that there had been treatment effects sufficiently great to overcome the initial loss had been tested. It cannot be overly stressed that the initial loss effects could hardly have been more unfavorable for the hypothesis (i.e. the rank order of means for all grades would be 2 > 1 > C) or more favorable for the null hypothesis. For this reason, the first analysis of posttest scores was a permutation test, for the purposes of establishing a rank order of means.

II. Posttest

Class means for the posttest are shown in Table 5. As can be seen there are again differences in the values between classes and between grades. Figure 2 shows rank order of cell means on the posttest. The means have been established more into the order forecast by the hypothesis of treatment effects.

Grade	Rank Order of Class Means
K	1 > 2 > C
1	2 > 1 > C
2	2 > 1 > C

Figure 2

Rank Order of Means on the Posttest

Table 4

Univariate Tests on Hypotheses on the Pretest

Source	Variable	MS	F	P	df
Mean	A	17377.01	4791.38	<.0001	1, 197
	B	18.06	58.47	<.0001	
Grade 1st-K	A	254.28	70.00	<.0235	2, 197
Grade 2nd-K	B	1.16	3.75	<.0001	
Treatment 1	A	5.13	1.42	<.2355	1, 197
Control	B	.02	.07	<.7917	
Treatment 2	A	22.12	6.10	<.0144	1, 197
Control	B	.25	.82	<.3650	
Interactions	A	5.18	1.42	<.2262	4, 197
	B	.37	1.20	<.3116	

Table 5

Observed Cell Means on the Posttest
by Treatment and Grade

Grade	Treatment	A	B	Total
K	1	9.79	.25	10.04
	2	8.35	.80	9.15
	C	8.61	.52	9.13
1	1	11.05	.32	11.36
	2	11.85	1.30	12.35
	C	10.21	.05	10.26
2	1	11.56	.40	11.96
	2	11.85	1.41	13.25
	C	10.41	.41	10.81

If one asked the question, "What is the likelihood of assigning treatments to classes if that treatment is random (see limitations set in the design section) and the rank ordering within classes is fixed such that one would observe the observed or greater agreement between predicted rank order to actual order (as shown in Figure 2)?" There are six possible assignments within each grade and each is equally likely. In grades one and two the actual agreement is the one assignment which is in perfect agreement with the hypothesis, while in kindergarten there is one assignment which could be better. The probability of such agreement in grades one and two is 1/6 in each while in K it is 1/3. This pattern of agreement could have occured in three equally likely ways, e.g., the one reversal could have occured in any one of the grades. The probability of such a "lucky" assignment of rank order of means is:

$$3 \times 1/3 \times 1/6 \times 1/6 + (1/6)^3 = .032.$$

Good fortune could hardly be that good, hence the null hypothesis of no treatment effect will be rejected. It can be concluded that the data are consistent with the predicted treatment effect, i.e., scores are highest under treatment two, followed by treatment one, and are lowest for the control group.

If the posttest scores are analyzed by parts (Part A is the score on items 1-12, and 16; Part B on items 13-15) the results show perfect agreement in all three grades for part B (probability = $(1/6)^3$ = .0046) under the null hypothesis and two reversals among the two treatment groups--but no reversals for either of the treatments vis-a-vis control--for the Total (probability = .088).

In summary then:

1. The data agree well with the prediction made by the hypothesis for a total score;

2. The data are in exact agreement with the prediction for part B; and

3. For part A the agreement is partial, the data do support the prediction that the treated groups could score higher but not the prediction that treatment two produces higher scores thus generalizing the effect of specific language labels than treatment one.

A. Analysis of Variance

Three separate multivariate analyses of variance were done on the data: one an analysis of variance on the pretest and posttest means; the second, an analysis of covariance on the posttest means adjusted for pretest differences; and third, an analysis of variance on gain scores. The models assume that each student within each class was affected independently by the treatments. Clearly this assumption is not correct. F tests were made for both parts (A and B) and all main effects (time, grade x mean gain, treatment and grade). The full table of values for both the MANOVA on pre- and posttest means and the MANCOVA tests are given in Appendix III. Gain score means and standard deviations are shown in Table 6. The differences are apparent, e.g. total score mean ranged from -.70 to 3.25.

For both analyses the multivariate F values for main effects and interactions were large enough to reject the null hypotheses. An examination of the univariate F tests indicated that the interaction effect was due to posttest differences, and not pretest differences because the grade x treatment interactions are significant. These data are not presented in detail here since it merely supports the earlier finding of significant differences without adequately accounting for initial class differences.

B. Analysis of Variance of Gain Scores

The third multivariate analysis of variance using gain score means was done as an attempt to eliminate some of the initial class effects.

The model assumed in analyzing gain score comparisons for the two tests was as follows--for the pre-test score of individual ℓ in class i assigned to treatment k in grade j:

Table 6

Gain Score Means and Standard Deviations

Grade		Treatment Control			Treatment 1			Treatment 2			Total Across Grade		
		B	A	Total	B	A	Total	B	A	Total	B	A	Total
Kindergarten	Mean	-.09	.74	.61	-.13	2.42	2.29	.45	2.2	2.65	.06	1.78	1.8[illegible]
	N			23			24			20			6[illegible]
	SD	.90	1.94	2.21	1.08	2.84	2.91	1.05	1.40	1.73			
1	Mean	-.05	1.37	1.32	0	1.37	1.37	1.0	2.25	3.25	.33	1.67	2.00
	N			19			19			20			5[illegible]
	SD	.40	2.34	2.40	.82	2.67	2.52	1.02	1.59	1.89			
2	Mean	.11	-.81	-.70	.22	.67	.89	.26	1.26	2.52	.53	.37	.90
	N			27			27			27			81
	SD	.85	1.82	1.88	.70	1.80	1.83	1.02	1.23	1.76			
Total	Mean	0	.30	.29	.04	1.46	1.5						
	N			69			70						

$$y_{1ijkl} = \mu + c_{i(j)} + \theta_j + a_{ijkl} + e_{ijkl},$$

where

μ is a constant for location,

$c_{i(j)}$ is the effect of class i nested within grade j (means 0, variance = σ_c^2)

θ_j is the effect of grade j (within school)

a_{ijkl} is the effect of individual l in class i (mean = 0, variance = σ_c^2, independent of $c_{i(j)}$) and

e_{ijkl} is an error of observation of individual l.

For the post-test score, the model assumed is:

$$y_{2ijkl} = \mu + \Delta_t + c_{i(j)} + \theta_j^* + \alpha_k + \beta_{jk} + a_{ijkl}^* + e_{2ijkl};$$

where

Δ_t is the effect of mean gain over the time period of experimentation

θ_j^* is the effect of grade j at the second point in time (it includes a grade X time interaction),

α_k is the effect of the treatment k,

β_{jk} is the interaction between grade j and treatment k,

a_{ijkl}^* is the individual effect * (mean = 0, var σ_{a*}^2

e_{2ijkl}^* is the error at the second observation ($\Sigma(e_{2ijkl}^*) = 0$, var $(e_{2ijkl}^*) = \sigma_{e*}^2$ not necessarily = σ_e^2, it also is independent of all other random variables).

* a_{ijkl} and a^*_{ijkl} are correlated.

Assumptions

Included in the model were the following assumptions:

Any interaction between mean gain and treatment is simply a treatment effect. We do presume no interaction between mean gain and class effect. It is also presumed that there is no class-treatment interaction. In other words, the effect of time is on the individual directly and/or through his grade or age level (aside from the main effect) and that the treatment effects are on the individual (a^*_{ijkl} includes both individual x time and individual x treatment interactions) and on the grade level.

These assumptions can be expressed in the form

$$y_{2ijkl} = \Delta_t + \theta^{**}_j + \alpha_k + \beta_{jk} + y_{1ijkl} + \Delta a_{ijkl} + \Delta e_{ijkl};$$

where

θ^{**}_j is simply $\theta^*_j - \theta_j$,

$\Delta a_{ijkl} = a^*_{ijkl} - a_{ijkl}$,

and

$\Delta e_{ijkl} = e^*_{ijkl} - e_{ijkl}$; or

$$y_{2ijkl} - y_{ijkl} = \Delta_t + \alpha_k + \theta^*_j + \beta_{jk} + Z_{ijkl},$$

where Z_{ijkl} is the sum of the error terms.

The parameters on the right do not include a class effect. Thus, an analysis of the gain scores would permit estimation of the parameters Δ_t, α_k, θ^{**}_j (which is now a "relative" gain over time by class), and β_{jk}.

The present model assumes that without the treatments the relative difference between classes would tend to be maintained while the spread or variation within classes would be increased, unless the individual x mean gain effect is negatively correlated with individual effects. The model also presumes that the treatments affected each student within each class independently of the others. The assumption that there was no class x mean gain interaction was guarded by not informing the teacher of the performance of the classes on the pretest. A violation of this would have invalidated the permutation test as well.

Least squares estimates

The least squares estimates of differences and effects and the standard errors of the estimators are shown in Table 7. The correlations of these estimators across effects are small, ranging from -.09 to +.09, owing to the fact that the cell sizes are almost all the same size. Within effects the correlations are high and the estimates for the same effect or similar effect are correlated across variables the same extent as between those variables. The multivariate and univariate F-tests of the gain scores is shown in Table 8.

1. Mean Gain - the mean gain or the time effect is large relative to standard errors for the test. The increase for all grades and treatments was substantial. This makes suspect the assumption that there was no class x mean gain interaction, because analysis thus far has shown that both of these main effects may be large.

2. Grade x Mean Gain - The effects on the two parts of the test are opposite in direction. On part A the two lower grades appear to have increased more than the second grade. The increase of either the lower grades over the second is large in relation to the standard errors while the

difference between the two lower grades is relatively small.

On part B there is a tendency for the increase in scores to increase with grade level.

3. Grade Effect on the Posttest - In spite of the above relative increases, the posttest scores tended to be higher in part A for each successively higher grade. The differences between the two upper grades and the kindergarten are large relative to standard errors but not so relatively high for the second grade - first grade difference. On part B, none of the differences appear large in relation to standard error.

4. Treatment - On part A the major difference is between the two treatment groups and the control groups. These are relatively large in relation to standard errors while that between the treatment groups themselves is not.

On part B, the marked difference is between treatment 2 and the other two groups. The difference between treatment 1 and the control is quite small in relation to its standard error.

A substantial correlation, about +.05, is shown for the various differences within each effect. The null hypothesis for treatment x grade x mean gain could not be rejected for either part separately or the two parts jointly or for total gain. This may indicate some successful elimination of those effects, or it may also indicate that the treatment x grade x mean gain and class x mean gain effects cancel each other.

Table 7

Least Squares Estimates of Effects

Source	Estimate: Test Part A	Estimate: Test Part B	Estimate: Test Part Total	Standard Error: Test Part A	Standard Error: Test Part B	Standard Error: Test Part Total
Time	1.27	.31	1.58	.14	.06	.15
Grade X Mean Gain						
1st – K	– .12	.24	.13	.36	.16	.39
2nd – K	–1.41*	.45*	– .95	.33	.15	.36
2 – 1	–1.29*	.21	–1.05	.42	.19	.45
Treatment						
1 – C	1.05*	.04	1.11*	.34	.15	.37
2 – C	1.47*	.91*	2.40*	.35	.16	.37
2 – 1	.42	.87	1.29*	.42	.19	.45
Treatment X Grade X Time						
1	–1.68	.09	–1.63	.87	.39	.94
2	– .20	.15	– .09	.80	.36	.86
3	– .58	.52	– .11	.89	.40	.95
4	.61	.61	1.18	.82	.37	.88
Grade on Post-test						
1st – K	1.85*	.03	1.89*	.30	.14	.33
2nd – K	2.35*	.22	2.57*	.28	.12	.30
2 – 1	.50	.19	.68	.35	.16	.39

Table 7 continued

Least Squares Estimates of Effects

Source	Estimate: Test Part A	Estimate: Test Part B	Estimate: Total	Standard Error: Test Part A	Standard Error: Test Part B	Standard Error: Total
Treatment on Post-test						
1 - C	1.06	-.00	1.05	.29	.13	.31
2 - C	.68	.84	1.52	.29	.13	.32
3 - C	- .38	.84	.47	.35	.16	.39
Grade X Treatment on Post-test						
1	- .34	.53	.19	.73	.33	.80
2	- .03	.27	.24	.67	.30	.73
3	1.10	.97	2.07	.74	.33	.81
4	1.70	.72	2.42	.69	.31	.75

* Indicate simultaneous .95 confidence intervals which do not include zero. A separate 5% error is "spent" for each effect group.

Table 8

Multivariate and Univariate Tests of the Gain Score Means

Source	F-ratio for Multivariate Test	df	P	Variable	MS	Univariate F	P	df
Mean	51.88	2, 196	<.0001	B GAIN	21.15	26.22	<.0001	1, 197
				A GAIN	293.77	73.32	<.0001	
Grade 1st-K	7.61	4.392	<.0001	B GAIN	4.07	5.05	<.0073	2, 197
Grade 2nd-K				A GAIN	45.46	11.35	<.0001	
Treatment 1	5.46	2, 196	<.0050	B GAIN	7.81	9.68	<.0022	1, 197
Control				A GAIN	6.71	1.67	<.1973	
Treatment 2	29.73	2, 196	<.0001	B GAIN	29.37	36.42	<.0001	1, 197
Control				A GAIN	81.74	20.40	<.0001	1, 197
Treatment x	1.10	8, 392	<.3634	B GAIN	.67	.83	<.51	4, 197
Grade				A GAIN	5.38	1.34	<.26	

III. Item Information

Item analysis for reliability was done using the GITAP program. This program uses the Hoyt ANOVA procedure for computing a reliability estimate. Table 9 gives the reliability estimates by class and treatment for the pre- and posttests. As can be seen, there is variation in th these values. For example, Grade 1 Control showed very low values, .1705 (pre-) and .1138 (posttest) while K-1 was much higher - .5974 (pre-) and .7059 (posttest). It is almost impossible to tell exactly what the sources of variation might be. Some of the factors identified as class effects in the pre-test analysis may have also contributed to variation here. The items were grouped in clusters which could cause certain patterns of response; ability grouping could cause skewing of results. There just cannot be a definitive statement of cause.

When considering the design, the threats to the internal and external validity as discussed by Campbell and Stanley were considered. The main concern here is with content validity.[14] The measure of difficulty, the p value, was computed by the GITAP program for each item by treatment group. Appendix IV contains the values of p for pretest, posttest and gain. Table 10 is a summary of p gain. Negative values have been omitted for purposes of clarity. As can be seen certain items showed consistent gain, e.g. items 2, 3, 5 were improved upon by nearly all groups. Item one showed nearly no gain at all, but in its case the cause was the ease of the question for the first and second grades. The p value was 1 on both pre-and posttests so there could be no gain.

---

[14] See item content discussion in Chapter IV.

Table 9

Values of the Hoyt Reliability
Estimate Computed by the GITAP Program

Grade	Treatment	Pretest	Posttest
K	1	.5974	.7059
	2	.2383	.3682
	C	.5295	.6397
1	1	.5605	.2869
	2	.4571	.4974
	C	.1705	.1138
2	1	.4829	.3829
	2	.4016	.2823
	C	.4186	.6627

There was a serious problem of invalidity in part B The way the test item was handled, it had validity only for the "treatment 2" groups. The p-gain for these values in 1-2 and 2-2 indicate their success as items for testing treatment measures. The language used in the test question, i.e. deliberately using the term "horizontal," prevented the other groups from applying whatever they had learned about the concepts. Therefore, for the "treatment 1" and control groups these items would be invalid for this test. Handled differently, e.g. using different but equivalent descriptive language in the test questions, they could well remain as items in a test for the concept of horizontality.

Table 10

Summary Table of p Gain by Item and Treatment Group.*

Item	K-1	K-2	K-C	1-1	1-2	1-C	2-1	2-2	2-C
1	.17	.00	.09	.00	.00	.00	.00	.00	.00
2	.67	.65	.09	.15		.35	.03	.11	.00
3	.29	.05	.17	.00	.06	.30	.11	.25	
4	.08	.00		.21	.05	.30	.07	.00	.00
5	.21	.25	.09	.16	.21	.10	.21	.22	.04
6	.33	.25	.13	.16	.26	.40	.08	.15	
7	.25	.00	.17	.26	.16	.20	.09	.11	
8	.04	.15		.31	.02	.00	.07	.07	
9	.08	.15	.17	.11	.00	.15	.00		.33
10	.13	.10	.04	.05	.16	.00	.07	.04	
11	.08	.45	.13	.00	.16	.00	.04	.04	
12	.04	.50	.21	.00	.10	.10		.07	
13		.45	.09			.65	.05	.74	
14			.00	.11	.00	.20		.22	
15	.00	.00			.00	.15		.29	.07
16	.16			.11	.00	.35	.07	.09	

* Negative values are not shown. See Appendix II for full tables of values. Zero values indicate no change.

Chapter VI

DICUSSION

The discovery of the magnitude of class effects influenced the subsequent treatment of the data. That there were such effects was not expected. Before the Stoughton system had been selected, the equivalence of the schools and the classes (with the exception of the kindergarten) had been stated, so that apparently whatever measures the school used for ascertaining this equivalency did not reflect the differences. To overcome these differences, which ran counter to the effects desired, the treatments had to be all the more effective. And the data clearly indicate they were.

Assumptions made of the independence of effects on each individual student in the analyses are clearly not true. The reliability test also assumed independence which was not true of the test situation. Therefore, much of the data can only be presented as reflecting what happened given the population, test measure, etc. There is no reason to expect a substantially different result given a different population, however. Certain effects and variables that could not be predicted or controlled are certainly at work in most classrooms.

The analyses that were used, using the student as if he were independently affected by the treatment, were employed in this way in order to smooth out the "lumps" in distribution caused by the non randomness of the population. While the assumption of independence was violated, these were the best procedures that could have been employed.

The grade x mean gain interaction was substantial. The p-gain values clarify this somewhat. The second grade scored significantly better on the pretest than did either of the lower grades. There was then a kind of "threshhold" effect here causing little room for gain in the second grade. Undoubtedly, in the third grade, with the exception of certain individuals, there would have been almost no gain.

Part B is problematical given the invalidity of the items as discussed above. There was a statistical grade effect "on this part."

Between the two treatment groups there was little difference statistically. The hypothesis that language labels would add efficacy to the treatment must be rejected for that reason.

That there was a significant improvement in gain scores in the treatment groups in relation to those of the control groups was encouraging; from this it could be said that the treatments were "successful." This would be from a relative position, however, for there were no absolute criteria for "passing" or "failing" the test. Table 11 indicates the range of items that were correctly answered by the various classes. Clearly there was a great increase in those who "passed" the test if the minimum score was arbitrarily set as 10 for every grade.

Kindergarten through second grade were chosen as the grades in which to limit the research experiment. Kindergarten children, with an age range from about 5.7 - 6.5 at the time of the year (May) of the testing, would be considered by Piaget to be just coming to some notion of horizontality. Second graders, age range approximately 7.7 - 8.2 should be able to be using horizontality in some instances.

The performance of the kindergarten classes on the tests was much better than had been expected. On their tests they were much more willing to make an effort to answer questions than the first graders. The test instructions had been just to leave blanks if they didn't know the answers. There were fewer blanks in the kindergarten papers than the first grade papers.

Another observed difference of the kindergarteners was that they more actively interacted with their surroundings during the treatment sessions. They did not remember well the words "horizontal" and "vertical," but they did remember the concept through giving instances and by making gestures. They were generally more social among themselves during the treatment session. They seemed to be excited by the classes and were very inquisitive.

A few individuals in the kindergarten classes did answer from 15-16 of the questions correctly. Doing a project with this sort of design, and really dealing with class effects, does prevent the exploration of these exceptional individuals. On the other hand, there were some individuals who did poorly. The "why's" of these behaviors remain unknown.

The kindergarten classes were arranged very informally with more space for the children to move in and contained more, and more varied, objects for them to manipulate. Thus they did have more opportunities to label and test different things as horizontal than did the other classes.

Even though there was a linear progression of improvement from kindergarten to first grade, it was noted that there was less difference between them than there was between the first and second grades. The first

Table 11

Percentages (rounded) of Subjects by Class Giving Correct Responses to Items on the Pre- and Posttests

	PRETEST				POSTTEST			
No. items correct	<5	5-9	10-13	14-16	<5	5-9	10-13	14-16
Treatment →								
Grade K-1	4	75	17	4	0	54	38	8
2	20	75	5	0	0	55	40	5
C	8	44	48	0	0	57	43	0
Grade 1-1	0	37	63	0	0	11	89	0
2	0	50	50	0	0	5	80	15
C	0	53	47	0	0	26	74	0
Grade 2-1	0	19	81	0	0	7	89	4
2	0	19	81	0	0	0	56	44
C	0	7½	85	7½	0	26	67	7

grade was in a different school than the kindergarten and second grades. Although in the factors mentioned before the two schools were alleged to be equivalent, with more sensitive measures of attitudes and behaviors there may be found differences that were not readily apparent. Observed differences in behavior and environment within the first grades (especially the two treatment groups) was most noted in the fact that there was very little vocal response during the treatment sessions, few responses were given without a specific request and few children were volunteering to perform in the sessions. Given time to act in the room during session, these children were much more restricted in their movements and the number of instances of the concept which they were exploring.

Part of the difference in the behavior of these children may have been due to the fact that their rooms were smaller and were structured more formally. The teachers were more authoritarian and tended not to relinquish their positions of authority during the training sessions.

The research could be justifiably criticized if these differences were real. A more thorough project would have been to have had three grade levels in each of three separate schools. But the research was done in a "real" situation, in the same way any curriculum that would be developed would be taught and while not as professional for statistical purposes as it might have been, the experimenter felt the project justified in that it was developed in its appropriate setting.

It was disappointing that there was not a larger difference between the two treatment groups, although in the second grade there was a much larger difference between them. As mentioned before, there had been interest in Brown's and Carroll's discussions of the importance of language labels as something on which to "hang" a concept. This idea of verbal tags giving a set of stimuli discriminality has not been proven; it has only been shown that subjects make varying use of words in mediating discriminations.

On the other hand, this research does not disprove this idea. Carroll significantly modifies the idea by saying, "A label is not particularly useful when it does not readily refer to a well-learned class of experiences."[15] Only in the second grade where we would expect a significantly greater understanding of the concept, does the difference between the groups begin to show with any magnitude (see Table 5). A further study including third grade might be further confirmation (although the increased stage of development might tend to obscure the differences in this instance).

When Piaget has talked of the relation between language and the child's development of logico-mathematical structures; it has been only in the formation of these structures that he has emphasized language is not crucially involved. It is after the structures have been formed that language comes to have a central position.[16]

It would then seem that there may be no contradictions here; the

---

[15] John B. Carroll, op cit., p. 96

[16] Ripple, R. E. and V. N. Rockcastle (eds.) Piaget Rediscovered (Ithaca, New York, 1965), p. 5. "Words are probably not a short-cut to a better understanding . . . the level of understanding seems to modify the language that is used, rather than vice versa Mainly, language serves t translate what is already understood; ... se language may even present a danger if it is used to introduce an idea which is not yet accessible.

project as undertaken would rather seem to be an inadequate vehicle for any definite exploration of the subject.

It must be noted that the most curiosity about the word "horizontal" was generated in the first and second grade treatment 1 groups - the ones that only heard the term during the test. Immediately after the posttest there was a great clamor to know its meaning, and great groans when they found that they really had been given the answer from the beginning.

Especially critical in education is the question of pushing the child, that with specialized training we can get the child to do things that are essentially beyond his capacity. Both Piaget and Montesorri have been concerned with this kind of teaching.[17] Commenting on Bruner's "teaching the structures" (elaborated by him in The Process of Education), Piaget made the following statement:

> The question comes up whether to teach the structure, or to present the child with situations where he is active and creates the structures himself . . . The goal in education is not to increase the amount of knowledge, but to create the possibilities for a child to invent and discover. When we teach too fast, we keep the child from inventing and discovering himself . . . Teaching means creating situations where structures can be discovered; it does not mean transmitting structures which may be assimilated at nothing other than a verbal level.[18]

It was in this spirit that the present research was planned and executed. In a classroom working with children during their school time it was important that they not merely become Ss being manipulated by the E.

---

17 "Piaget and Montessori," (David Elkind), Harvard Education Review, 37:4 (Fall, 1967), 540.

18 R. E. Ripple and V. N. Rockcastle, op cit., p. 3.

That the children who took part were genuinely involved in a learning experience was a secondary goal in the planning of the experiment.

Not really touched has been the argument concerning the role of learning versus the role of development, and it is not presumed that this study contributed to this discussion. It may be argued that all that was done was an exercise in which the children learned to learn this particular concept; or that they were taught merely to use techniques which would enable them to use their abilities more effectively. In that case, the purpose of the experiment was fulfilled. It would be possible to include a training sequence similar to this one in a curriculum designed to effectively teach the concept of horizontality in grades K-2 with the varying success indicated by the data.

REFERENCES

Beilin, H., Kagan, J., and Rabinowitz, R. Effects of Verbal and Perceptual Training on Water Level Representation. In I. E. Sigel and F. H. Hooper (Editors) Logical thinking in children. New York: Holt, Rinehart & Winston, Inc., 1968.

Bever, T. G., Mehler, J., and Epstein, J. "What Children Do in Spite of What They Know." Science, 62, p. 921-924.

Brown, R. Words and Things. Glencoe, Illinois: Free Press (MacMillan), 1958.

Campbell, D. T., and Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Co., 1963.

Carroll, J. B. Language and hought. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1964.

Dodwell, P. C. Children's Understanding of Spatial Concepts. In I. E. Sigel and F. H. Hooper (Editors) Logical thinking in children. New York: Holt. Rinehart & Winston, Inc., 1968.

Flavell, J. H. The developmental psychology of Jean Piaget. Princeton, New Jersey: D. Van Nostrand Company, Inc., 1963.

Piaget, J. On the development of memory and identity. Barre, Massachusetts: Clark University Press, 1968.

Piaget, J. "Quantification, Conservation, and Nativism." Science, 162, 1968, pp. 976-979.

Piaget, J., and Inhelder, B. The child's conception of space. New York: W. W. Norton & Company, Inc., 1967.

Ripple, R. E., and Rockcastle, V. N. (Editors). Piaget Rediscovered: A Report of the Conference of Cognitive Studies and Curriculum Development, March, 1964. Ithaca, New York: School of Education, Cornell University, 1965.

Rivoire, J. L. The development of reference systems in children. Unpublished doctoral dissertation, University of Arizona, 1961.

Shantz, C. U., and Smock, C. D. Development of Distance Conservation and the Spatial Coordinate System. In I. E. Sigel and F. H. Hooper (Editors) Logical thinking in children. New York: Holt, Rinehart & Winston, Inc., 1968.

Smedslund, J. "The Effect of Observation on Children's Representation of the Spatial Orientation of a Water Surface." J. Genetic Psychology, 102, 1963, pp. 195-201.

APPENDICES

Appendix I

THE TEST

A. Natural horizontal

Page 1. Draw the correct waterline.

Page 2. Choosing the proper representation of the waterline.

Page 3. Matching representations of the waterline.

All of these tasks involved three aspects of the property horizontality as indicated by water level:

a. direction in which the bottle was tilted,

b. the correct waterline, either to be drawn by the subject or already indicated, and

c. the placement of liquid in the container, i.e., that the water would always be in what was the bottom portion regardless of the bottle's position.

Page 4. #1. Labeled concept. Pick the instance of "horizontal."

Page 5. Recognize "levelness" as a property of horizontal.

B. Horizontal with rotated axis

Page 4. #2, 3. Recognize that the concept horizontal can be extended to objects with rotated axes.

In all of these items, and in the training procedure, implied conservation of volume was prevented as much as possible from becoming part of the testing procedure. Concentration was entirely on the perceptual line the top of the water made and it was this line that was the sole criteria of correctness on the test.

There were three basic assumptions in the behaviors we expected the children to exhibit:

1. Discrimination of direction. That is, we felt that the children would be able to discriminate the direction in which the bottle was tilted and to match it with the representation on the test paper. Care was taken to hold the bottle in the same direction as the representation the children were viewing so there would be no doer-seer reversals. Direction was always indicated, never labeled, i.e., never called "left" or "right."

2. That the children would be aware of gravitational effects on liquids in containers, and would have no problems understanding that the liquid would be on the "bottom" of the bottle. This is not to say that we were assuming conservation of liquids.

3. That verticality would be somewhat of an established concept, since this seems to be a somewhat easier concept for the child to grasp. It should be pointed out, in retrospect, that this was an assumption made for the convenience of the experiment. Piaget does not clearly state the relationship of the rate of development of these two hypotheses. The concept of verticality was necessarily made use of in the training sessions.

Testing Procedure

The children are to be seated at their desks and equipped with pencils. A copy of the test is handed to each child by the experimenter. The children are instructed to keep the tests face down, and to write their names on the back of the tests, first name only. (This is done so that the pretest of each child can be matched with that child's posttest.) When this was done, the experimenter gives the following introduction:

Equipment: One plastic pint container. The lid is taped on, and the container is approximately half-filled with water colored red by food coloring. At the water level, the waterline is indicated by a strip of dark tape which extends around half of the container.

One plastic liter bottle half-filled with water colored by blue food coloring.

One plastic liter bottle, empty, whose sides were covered by paper and masking tape. The contents are thus completely hidden after it is filled.

The experimenter holds up the pint container. "As you can see I have a plastic container in my hand. It is half full of water that has been colored red with food coloring. You can see very well where there is water because it looks red; and where there is no water, you can see the white color of the plastic. Where the water and the area where there is no water meet, there is a very sharp difference: we call this the "waterline" (pointing to and tracing the line with a finger). We could draw a real line here. (Turning the container around and showing the tape strip.) This line we would also call the "waterline," because it marks the top of the water in the container when we hold it upright. If I wanted to draw a picture of the container to show someone how much water I had in the container (draw a line diagram of the container on the chalkboard) I would draw a straight line from one side of the container to the other. This would be the waterline. It would show him how much water I had.

Sometimes you make waterlines. If you are very, very dirty, and you take a bath, when the water is let out of the tub you leave a ring. This bathtub ring is a waterline, too. It shows how much water was in the tub when you were bathing.

By drawing a waterline like this, (draw a cup with a waterline on the chalkboard) I can show people how much, of say, coffee I wasn't able to drink because I was in too much of a hurry, without their having to see the cup themselves. People understand by this line that it shows how much water or other liquid is in a container.

I have here a plastic bottle (the filled liter bottle) half filled with blue water. Since I want you to draw waterlines where you think they should be, I am going to pour the water from this bottle into one whose sides are covered so that you cannot see the water. You can turn your papers over now. (As the children are doing this, continue to pour the water into the covered liter bottle.) On the first page you can see pictures of this bottle. As I tip the bottle, the way it is tipped in the pictures on the page, I want you to draw the waterline the way you think it should look.

(Holding the bottle straight up:) What do you think the waterline looks like when I hold the bottle like this? This is the first one, the picture of the bottle with the numeral one under it. Draw the waterline. If you can't draw the waterline, just leave it blank."

Continue this process for the following four bottles. They are tipped 0° right, 45° right, 90° right, inverted and 30° left.

Take the test itself and turn the page. "Now turn the page to page 2 (point to the "2" in the corner). Here we have 4 rows of bottles with water in them. As I tip the bottle, I want you to pick the picture in the row that shows the bottle and how the water in it looks. Only one picture will be correct, so mark only one. Mark it either by putting an "X: over it (write an "X" on the chalkboard) or by circling it (draw a circle on the board). In the first row, which picture shows the bottle when I tip it this way? Mark the one correct picture."

Repeat the procedure for the next 3 rows. The correct pictures show the bottle tipped these ways: 45^{o} right, 90^{o} inverted, 45^{o} left, and 0^{o} right.

Taking the test turn the page and instruct the children to do the same. Holding up the test: "Hold the test this way in front of you so that the "3" is here and the arrow points away from you. On this page are three rows of bottles. At the beginning of each row in the box there is a picture of a bottle. Somewhere in the row there is another picture that looks just like it. I want you to mark the picture in the row that looks just like the one in the box. There will only be one that is the same in each row. Do all three rows."

When it appears that most of the children are finished, say: "When you have finished all three rows you may turn the page." Wait until it appears that all of the children have finished. Then turning the page to page 4: "In the box at the top we have a row of 4 boxes. These boxes have stripes in them. I want you to mark the box with the horizontal stripes." (Pause) "If you don't understand the question, just leave the row blank, don't mark anything; it will be all right." Repeat the instructions only once if there are questions.

"In the next row there are pictures of four cups with stripes painted on them. I want you to mark the cup with the horizontal stripes.

Then, turning the page to page 4: "In the box at the top we have a row of 4 boxes. These boxes have strips in them. I want you to mark the box with the horizontal stripes." (Pause) "If you don't understand the question, just leave the row blank, it will be all right." Repeat the instructions only once if there are questions.

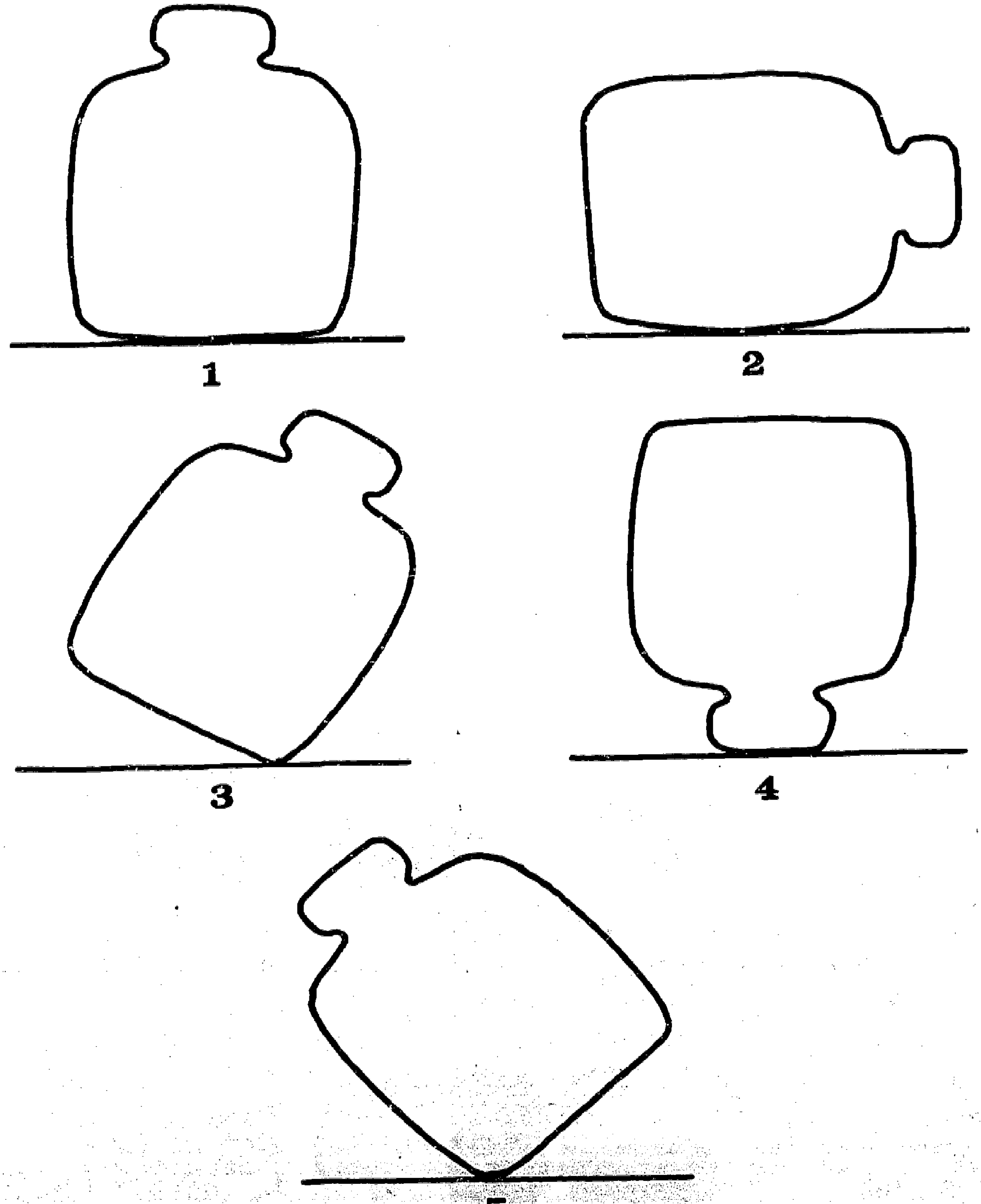
"In the next row there are pictures of four cups with stripes painted on them. I want you to mark the cup with the horizontal stripes painted on it. If you don't understand the question, don't mark any of them." Repeat the instructions once if there are questions.

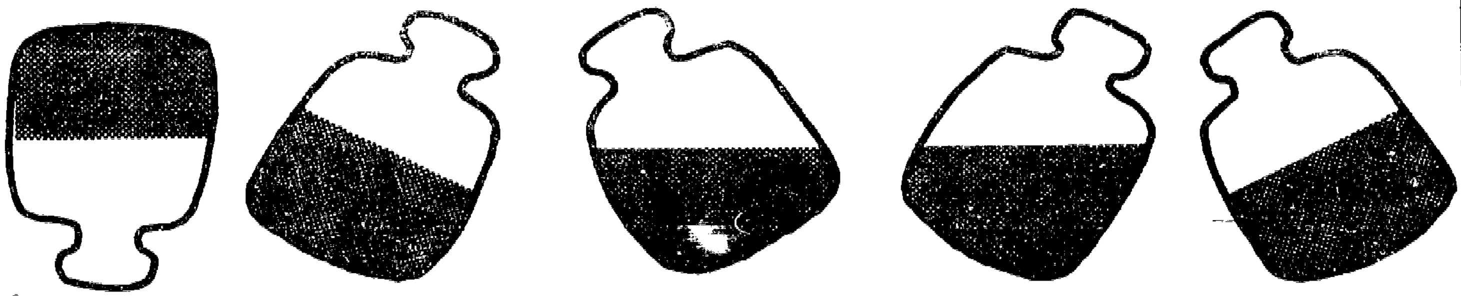
Repeat the next row, using the term "flower pots" instead of cups.

Hold up the test and turn to page 5. "On this page there are pictures of some shapes. Imagine they are wood or plastic or some other kind you would be playing with. If I had a marble and placed it on top of each of these shapes it would roll off all of them but one. Would you mark the one it would not roll off." Repeat the instructions only once if there are questions. When the children are finished, the last child in each row can be instructed to collect the papers.

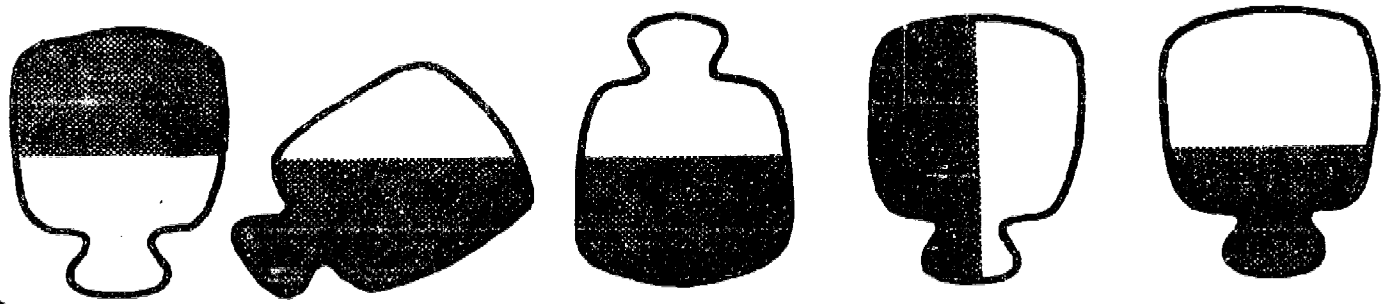
The post-test is identical to the pre-test.

All children are to be given the same test.

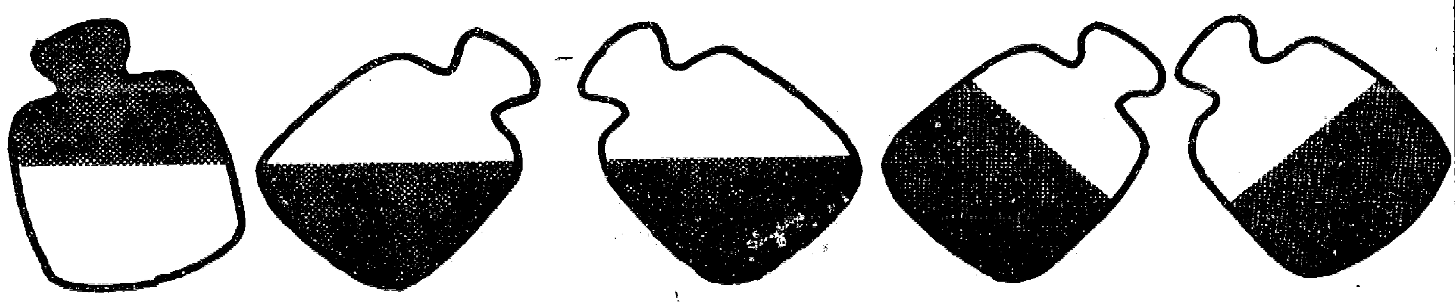

1
2
3
4
5



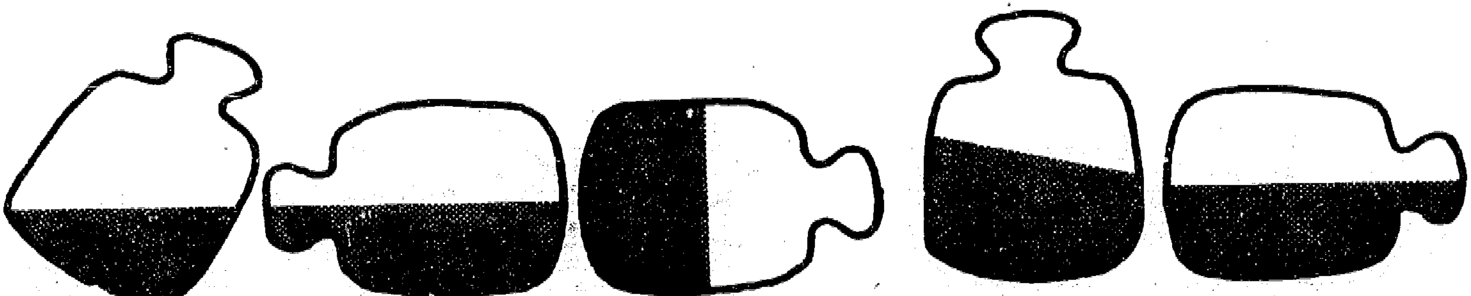
1



2

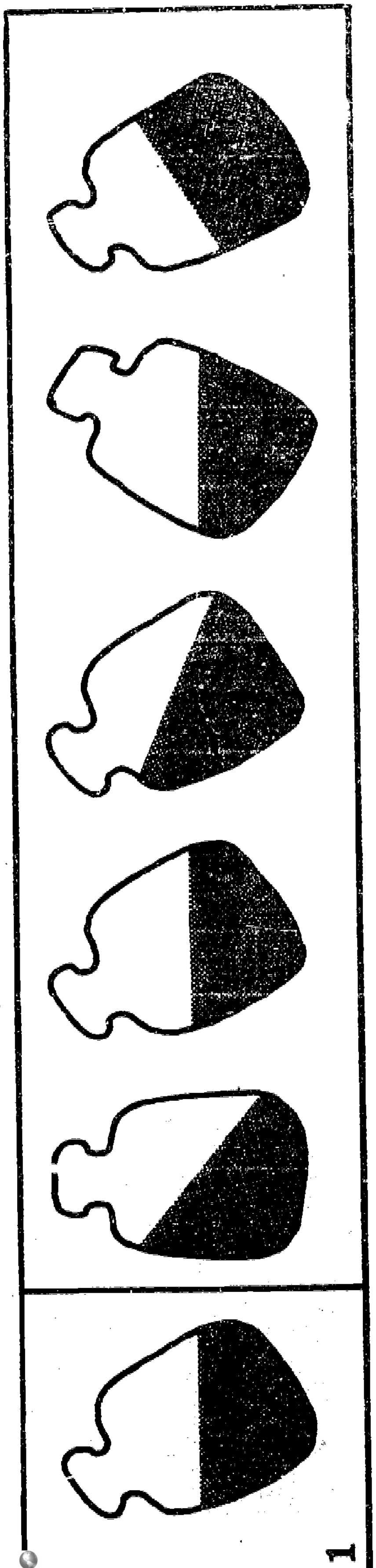


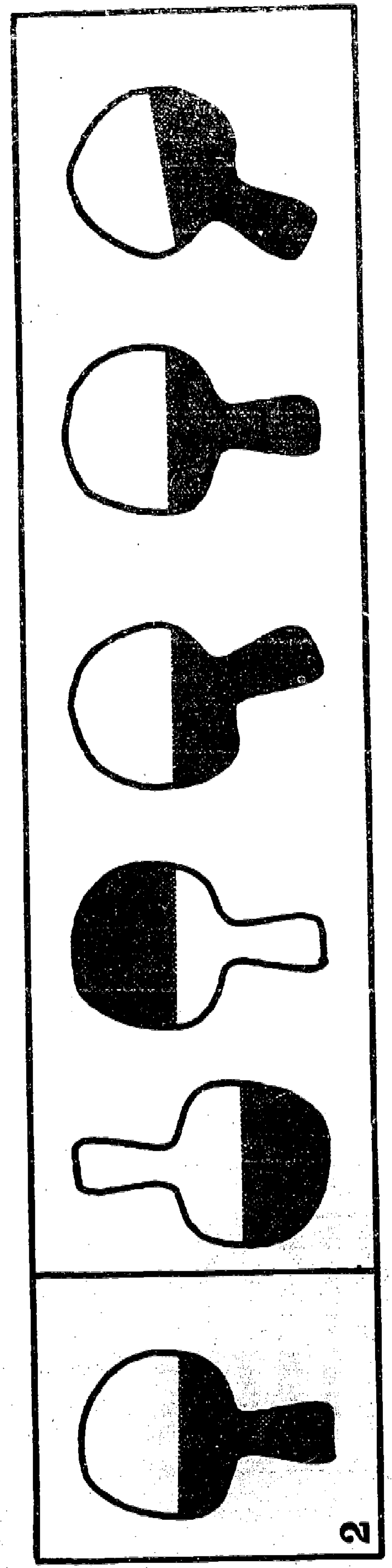
3

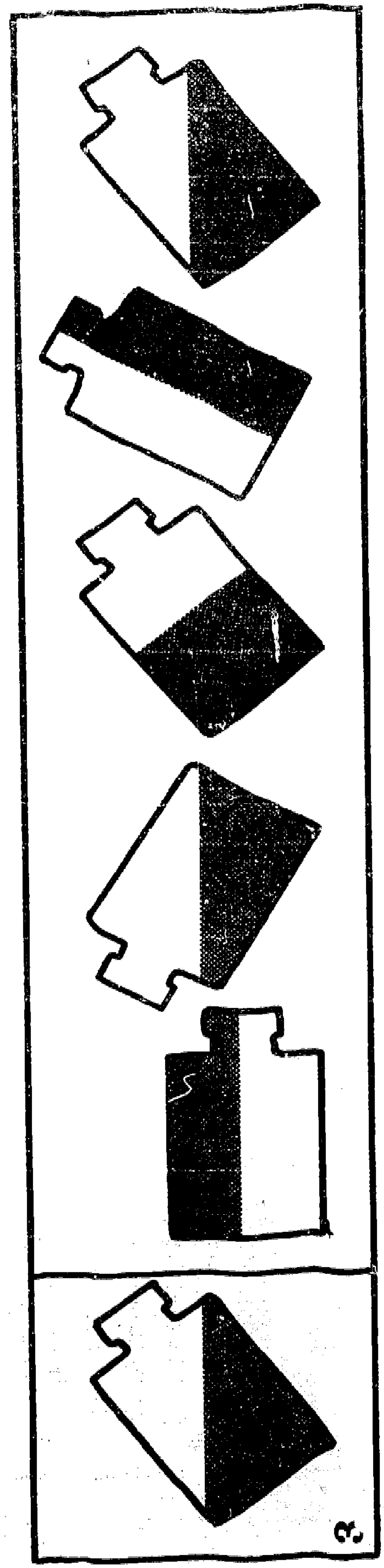


4

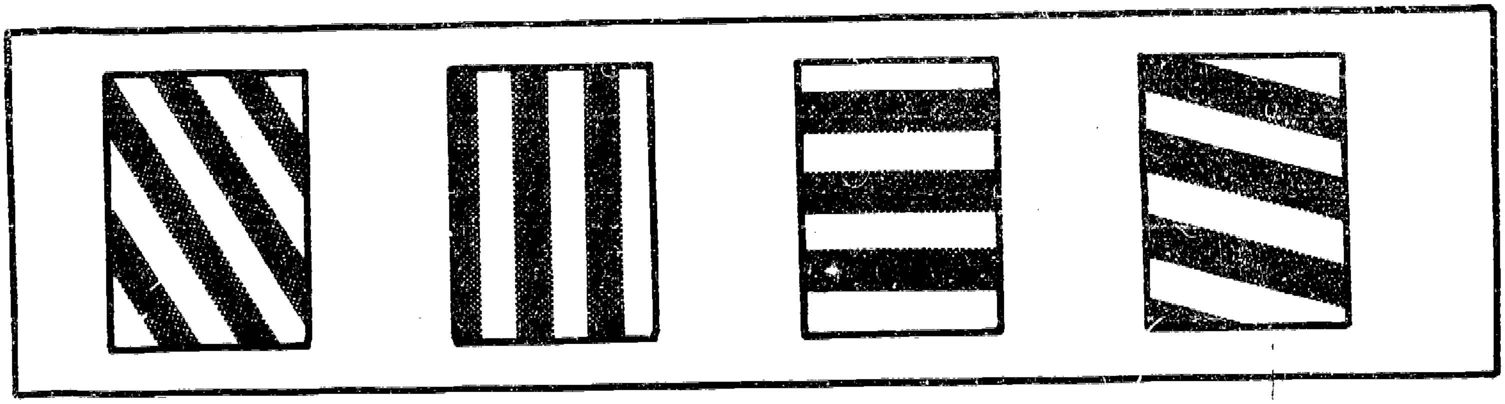
3

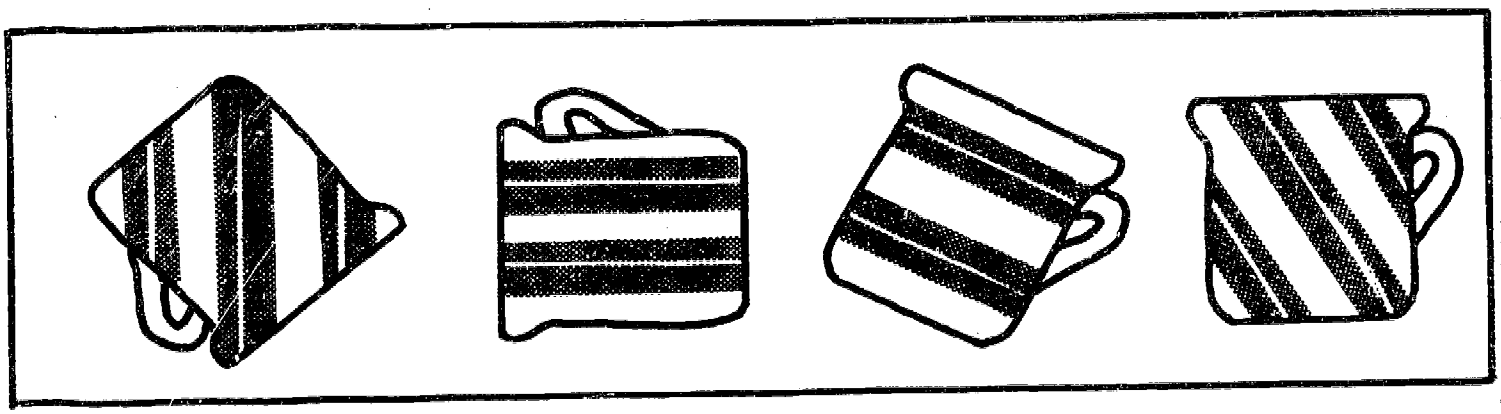


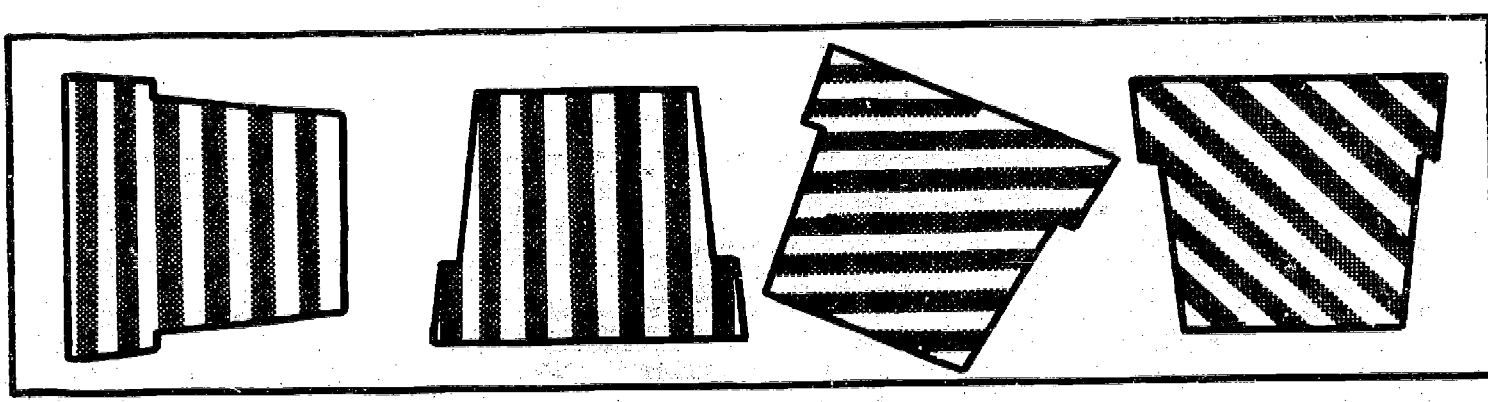




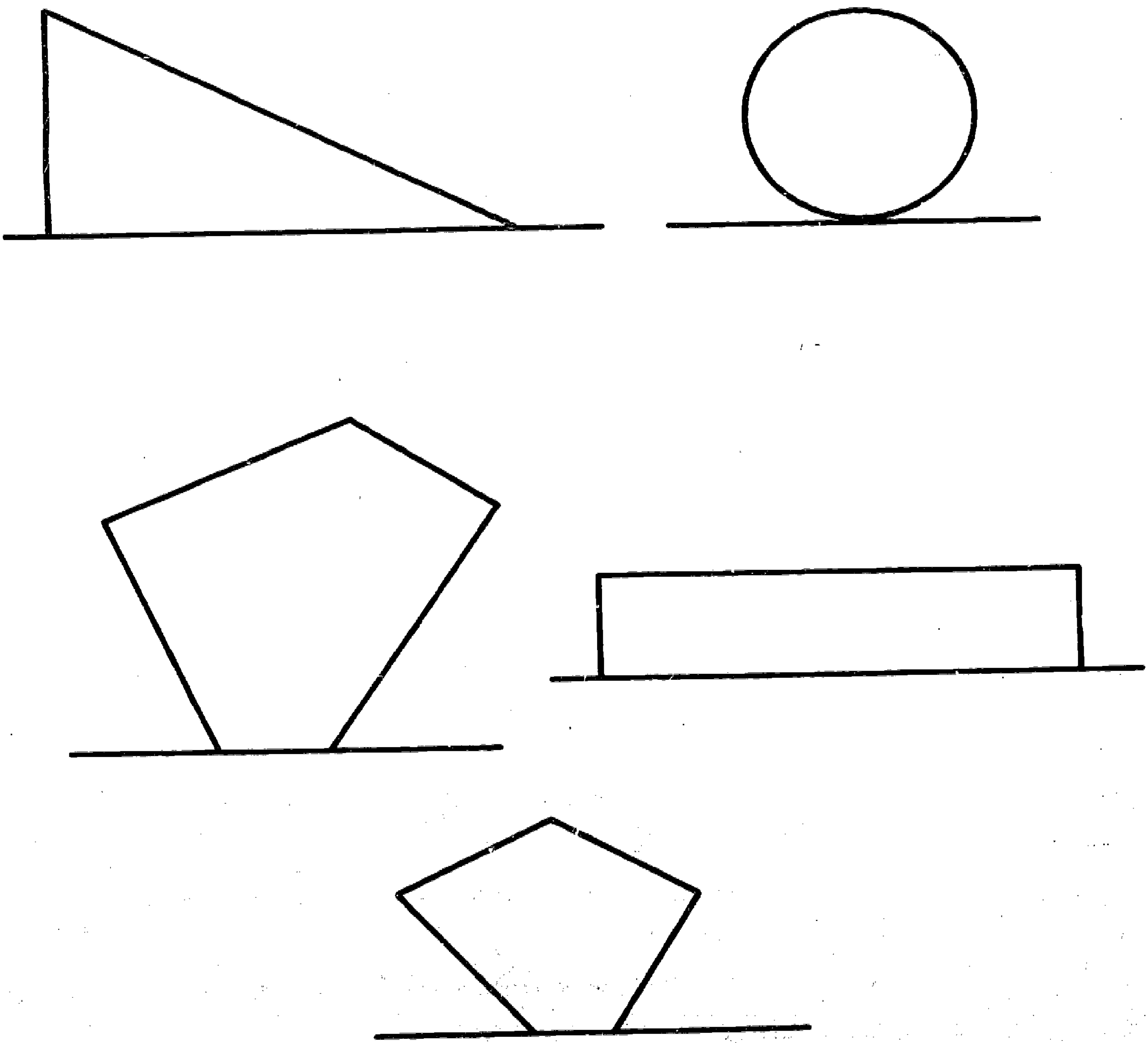
4







5↑



Appendix II

SEQUENCE OF TRAINING TASKS[19]

The purpose of outlining the tasks to be done was two-fold: First, to ensure that the elements of the pre-test administered to the children would be adequately prepared for in the post-test; and secondly, to ensure that the concept was completely and logically developed and that the tasks involved had been separated and analyzed.

I. Concept: Gravitational horizontal.

1. Introduction

   a. Demonstrate gravitational vertical.

   b. Demonstrate gravitational horizontal.

   c. Name the "waterline."

   d. Demonstrate or state that the waterline always crosses or meets the vertical (gravitational) at right angles.

2. Anticipation of correct waterline (group and individual activities).

   a. Draw the correct waterline on empty bottles. Test

---

19 With the help of Mrs. Rochelle Meyer, derived from her work on the geometry task analysis for the project on Prototypic Instructional Systems in Elementary Mathematics, Wisconsin R & D Center.

with a bottle filled with water.

b. Children to be given work sheets with bottles drawn on them on which they can draw the waterline. Test with real bottles.

3. Example of the carpenter's level.

a. Recognize that the bubble in a carpenter's level is always at the top. Demonstrate the same with a nearly full bottle of water.

b. Show the relation between a carpenter's level and the waterline of a nearly full bottle (will appear bubble-like).

c. Use carpenter's level to test levelness.

d. Name level surfaces as "horizontal" (gravitational).

e. Repeat water level demonstrations using the carpenter's level.

f. Use the carpenter's level to find horizontal objects and surfaces in the room.

II. Concept: Horizontal with a non-vertical (gravitational) axis.

a. Match line on paper to horizontal (gravitational) with level.

b. Draw waterline on paper to match base line. Check waterline with level.

c. Name base line as the "paper horizontal" or "horizontal on the paper."

d. Recognize that "horizontal on the paper" does not change when the top-bottom of the paper is not up-down (gravitational).

e. Repeat with other objects.

When the lessons were being planned, there was some revision of this outline in detail, not in substance. It was thought that there were some places where more detail was needed, other places there were changes made to accomodate a better teaching format. There was also some changes due to the equipment used in the teaching procedure.

Training Lessons

The training lessons for the two treatments were identical except for the language used to name the concepts. Treatment II is used to designate the classes that were taught the concepts in the most general terms, i.e., horizontal was described as going across or flat, or a straight line across, usually accompanied by a hand gesture by the teacher. Treatment II had the concepts described using language labels; e.g., the words horizontal and vertical were used to describe the phenomena and the concepts.

In the lessons it was noted that the children designated as straight any line that was not wavy or curved, regardless of direction or inclination of that line. A line drawn vertically on the board was designated as straight as was a line drawn horizontally. Differentiation on the property of direction was necessarily prompted by the teacher.)

The training lessons were each 20 minutes long. Both treatment classes in each grade level were given the same lessons, as was each grade level.

The teacher for the training lessons was Mrs. Caroline Gornowicz, project specialist, Project on Prototypic instructional systems in Elementary Mathematics and an experienced certified elementary teacher.

Although the lessons are written in a straight expository style, Mrs. Gornowicz would ask the children for terms, to describe what was happening, and other details in the lesson plan. From the experimenter's observations, and allowing for the individual differences within and between the classes, she was consistent in her format and the amount of material that was covered with each class.

Training Lesson 1

Equipment: Plastic liter bottles half full of water colored blue with food coloring. There should be one bottle for every two children.

Straw goal posts. This is a construction made of drinking straws that is made in the shape of a football goal post. The upright pieces are anchored on small pieces of modeling clay. The crosspiece should be carefully glued in so that it is horizontal (and tested by a carpenter's level).

Worksheet for each individual child. Each worksheet had 16 different bottle outlines tipped in various positions (including the positions of the bottles given on the pre- and post-test).

Procedure: The teacher has one child stand in front of the room. She names the position of the central axis of the child "vertical," or top-to-bottom. She draws a vertical line on the board saying, "This is a vertical line. It looks the same way ______ looks when he is standing here." She then has the child spread his arms apart, pointing to the line they form from the tips of the fingers through the body and names it "horizontal." She then draws it on the board, perpendicular to the vertical line. She notes that "where they meet they form right angles or a corner" and she picks up a book and matches its corners with the angles of the intersection.

"The horizontal here looks the same as the line where the earth and the sky meet. We call that line the horizon. (She directs the children to look out the window, or to remember where they have seen places where they have seen the horizon.) That's where we get the word horizontal, from horizon." And she writes the word "HORIZON" on the board with the word "HORIZONTAL" immediately under it. All the lines that go across like this (drawing a series of parallel horizontal lines on the board over the vertical), meeting the vertical and forming corners, we call horizontal."

"Why are we talking about horizontal lines? On the papers you did last week with ______(experimenter's name) you worked with a waterline." Teacher stands a liter bottle partially filled with water on a table in front of her. "When the bottle is standing on the table, the water goes across like this," as she runs her finger around the bottle at the waterline.

She draws a line drawing of a bottle on the board. "If we draw a picture of this bottle, we can draw a vertical line through it like this," (she draws a vertical line through the picture), "and when we draw in the waterline," (she draws in the waterline and shades in the water), "we see that it goes across the vertical forming a corner. When we drew this kind of line over here (points to previous drawing) we called it horizontal. The line is the same here. The waterline is a horizontal line."

She picks up one of the goal posts. "This is a goal post, like one used in football or soccer, only this is just a straw model. These poles are vertical, they hold it up. (points) This pole on top, cutting across the vertical ones is horizontal. (She holds it up comparing it to the horizontal and vertical lines already drawn on the board.) When I put the bottle behind the goal post, the waterline and the horizontal line of the post look the same." (She holds up the bottle vertically behind the goal-post.)

Tilting the bottle and holding it up, "When I tilt the bottle, the bottle itself is no longer vertical, but the waterline remains horizontal. We can prove this by matching it with the horizontal of the goal post, showing that the waterline and the horizontal bar look the same. She repeats this procedure, tilting the bottle to an extreme degree in another direction.

"On the paper you were given there are pictures of bottles. Let's look at the first one. (Points) Draw in the waterline the way you think it will look when the bottle is actually tipped in that direction." She draws on the board a bottle that is the same as the first example. She calls on one of the children to come up and draw in the waterline, then has him check his work by comparing the actual waterline in the tilted bottle with the one he has drawn. If he has done it correctly, praise him and procede. If it was incorrect, point it out, and have him redraw the line. If still he is unable to do it, she gets another child to come up and help. When the example is correct, she asks the class to compare that line with theirs, and if they have drawn an incorrect one to cross it out and redraw it. She then repeats this procedure with the second example.

The children are then given bottles and goal posts and asked to finish the page. Because of the time, the children were not expected to finish, so arrangements were made to leave a bottle and a goal post in each of the rooms to allow them to finish if they wished to.

Training Lesson 2

The teacher reviews horizontal and vertical by drawing the lines on the board, asking the class to name them, and by stating again that the waterline is a horizontal line.

"Horizontal and vertical lines are very important. Look around the classroom and see how many horizontal and vertical lines there are just in this one room. (Points out examples) The carpenters who built this building, or any building have to be careful to keep their lines horizontal and vertical. If they don't, the building will tilt more and more until it falls down. They need a tool to help them keep their lines from tilting. So they use what is called a carpenter's level. (She holds one up for the class to see.) There are glass tubes with green-colored liquid in them. On each of the tubes are two black lines. When the bubble in the liquid is between the two lines the carpenter knows his line is straight." (As she talks, the teacher draws a top view, showing the lines and the bubble.) She puts the level against several surfaces, having the children come around to see for themselves.

"When I fill a bottle almost full, the waterline is way up here at the top." She illustrates by filling a bottle nearly to the top. "When I tip the bottle on its side, the waterline looks like this." Holds up bottle. Because the bottle was so full, there is only a bubble showing. She tips the bottle in several positions, each time returning it to a vertical position, and stating that the bubble is the water level indicator. "This bottle is doing the same thing as the carpenter's level. In both of them, the waterline looks like a bubble because the container is so full. In both of them, the bubble is always on top.

When the carpenter's level is horizontal, the bubble in the tube is in the center. If I hold the bottle to match the position of the carpenter's level, its bubble is in the center of the bottle. They are both horizontal now."

Keeping the level in a horizontal position, the teacher holds up a bottle only partially filled with colored water. She matches the waterline with the level and states that the waterline is horizontal. She continues this, tipping the bottle in various angles, each time stating that the waterline is horizontal.

The children are given the constructed water levels and are asked to find horizontal surfaces and lines in the room.

Equipment: Carpenter's level.

Several liter bottles filled with colored water.

Constructed water levels, at least one for every two children. Levels were made by cutting glass tubing into 3 inch lengths and firing the ends. One end was caulked and they were attached to cardboard pieces about 3" x 5" with Elmer's glue. Colored water was poured in them and the other end was then caulked. The children were cautioned to be very careful with them. It would be more advisable to have an apparatus in which the glass could be completely embedded.

Training Lesson 3

Equipment: 1. On standard 8 1/2" x 11" paper was a line picture of bottle which was colored tan. Green stripes, approximately 1/4" wide, were placed horizontally on the bottle. A black line, representing the vertical or "top-to-bottom" line of the bottle was superimposed over the full length of the page. (See Figure 1)

2. Three transparencies of plastic with variously colored strips.

a. The first, on which was a black line representing the vertical or up-and-down.

b. The second had a red stripe on it which represented the horizontal or across and which also designated the waterline on the bottle.

c. The third had a red line on it representing horizontal and was manipulated by the children to show the waterline on an actual bottle.

Procedures: The teacher attached the drawing of the bottle to a board. She held a real bottle half-filled with water that had been colored with blue food coloring. She also drew crossed horizontal and vertical lines on the board. "The stripes on the picture of the bottle go across the bottle, or are called horizontal stripes. They go in the same direction as this line (indicated the drawn horizontal line). If the bottle that I am holding had horizontal stripes on it they would go like this (indicating). The black line in the picture represents the vertical or up and down of that bottle. It goes in the same direction as this vertical line (indicates line on chalkboard). On this bottle (indicates the one in her hand) the up and down or vertical goes in this direction. On this piece of clear plastic I have another vertical line. I'm going to put it over the picture of the bottle. You can see that the vertical here matches the vertical of the bottle.

I have another piece of clear plastic with a red line. This line is a horizontal line. (Matches it with the drawn horizontal line of the board.) It also ir ''cates the waterline. You know that all waterlines are horizontal. (Attached this piece on top of the two others.)

You can see also that the waterline and the stripes on the bottle all run in the same direction. They are all horizontal lines.

Now if I tip the bottle here and the picture of the bottle, what happens? The waterline stays the same, it is still horizontal. And our vertical is still the same. But the up-and-down or vertical of the bottle is no longer vertical to us. But it is for the bottle. The same with the stripes on the bottle--they are no longer horizontal for us, but they are still horizontal on the bottle. (Looks for a child with something horizontal that he is wearing, preferably a shirt.) Look, you can see the stripes on his shirt. They go across, or are horizontal. If he bends over, the stripes are still horizontal on his shirt, but when we look at them they are going in all sorts of directions. If he would lie down, they the stripes would look vertical to us, but to his shirt they would still be horizontal. The same way with our bottle. The stripes no longer look horizontal to us, but to the bottle they still run the same way. (Does this with several examples of clothing if they are available.)

Repeats this twice tipping the bottle at different angles.

This procedure continues, in front of the class, individual students come up, hold up the transparency to show the waterline on the real bottle, then the picture of the bottle, each time the teacher showing the difference between the horizontal of the waterline which the child has indicated and the horizontal of the stripes and explaining that we call both horizontal.

Appendix III

Full values for multivariate analysis of variance on the pretest and posttest and a multivariate analysis of covariance test of posttest/pretest differences are presented in the following tables.

Table A

Multivariate and Univariate Tests of the Pretest and Posttest Means

Source	F-Ratio for Multivariate Test	df	P	Variable	Ms	Univariate F	P	df
Mean	2379.18	4,194	<.0001	Pretest A	17377.01	4791.36	<.0001	1,197
				B	18.06	58.47	<.0001	
				Posttest A	22189.53	7867.29	<.0001	
				B	78.29	137.41	<.0001	
Grade 1-K 2-K	18.30	8,388	<.0001	Pretest A	254.28	70.11	<.0235	2,197
				B	1.16	3.75	<.0001	
				Posttest A	104.74	37.14	<.1507	
				B	1.09	1.91	<.0001	
Treatment 1-Control	5.75	4,194	<.0003	Pretest A	5.13	1.42	<.2355	1,197
				B	.02	.07	<.7917	
				Posttest A	23.58	8.36	<.0043	
				B	8.65	15.19	<.0002	
Treatment 2-Control				Pretest A	22.12	6.10	<.0144	1,197
				B	.25	.82	<.3650	
				Posttest A	18.82	6.67	<.0106	
				B	24.15	42.38	<.0001	
Treatment x Grade	1.72	16,593.32	<.0391	Pretest A	5.18	1.42	<.2262	4,197
				B	.37	1.20	<.3116	
				Posttest A	6.23	2.21	<.0692	
				B	1.36	2.40	<.0513	

Table B

Analysis of Covariance Tests on Posttest Means
Adjusted for Differences in the Pretest

Source	F-Ratio for Multivariate Test	df	P Less Than	Variable	Ms	Univariate F	P Less Than	df
Mean	86.52	2,194	.0001	A B	411.71 2.41	169.12 4.21	.0001 .0415	1,195
Grade 1-K 2-K	4.61	4,388	.0013	A B	19.68 .73	8.08 1.28	.0005 .2795	2,195
Treatment 1-Control	10.72	2,194	.0001	A B	16.65 8.53	6.84 14.92	.0097 .0002	1,195
Treatment 2-Control	28.07	2,194	.0001	A B	33.98 23.96	13.96 41.94	.1003 .0001	1,195
Treatment x Grade	2.01	8,388	.0438	A B	4.80 1.18	1.97 2.07	.1006 .0853	4,195

Appendix IV

The following tables show full values for p (a measure of difficulty) on both pre-and posttests for all items. These tables are presented by treatment group.

1. Kindergarten, Treatment 1

Item	Pretest	Posttest	Gain
1	.8333	1.0000	.1667
2	.1667	.8333	.6667
3	.0417	.3333	.2916
4	.8333	.9167	.0834
5	.1667	.3750	.2083
6	.2917	.6250	.3333
7	.5417	.7917	.2500
8	.6250	.5833	.0417
9	.7083	.7917	.0834
10	.8333	.9583	.1250
11	.8333	.9167	.0834
12	.7917	.8333	.0416
13	.1667	.0833	-.0834
14	.1667	.1250	-.0417
15	.0417	.0417	.0000
16	.6667	.8333	.1666

2. Kindergarten, Treatment 2

Item	Pretest	Posttest	Gain
1	.9500	.9500	.0000
2	.3500	.9000	.6500
3	.0500	.1000	.0500
4	.7500	.7500	.0000
5	.0000	.2500	.2500
6	.3000	.5500	.2500
7	.5000	.5000	.0000
8	.3500	.5000	.1500
9	.4000	.5500	.1500
10	.7500	.8500	.1000
11	.5000	.9500	.4500
12	.3500	.8500	.5000
13	.0500	.5000	.4500
14	.2000	.1500	-.0500
15	.1000	.1000	.0000
16	.9500	.8500	-.1000

3. Kindergarten, Control

Item	Pretest	Posttest	Gain
1	.9130	1.0000	.0970
2	.6522	.7391	.0869
3	.0870	.2609	.1739
4	.9565	.9130	-.0435
5	.3043	.3913	.0870
6	.3913	.5217	.1304
7	.4783	.6522	.1739
8	.5652	.0435	-.5217
9	.5652	.7391	.1739
10	.6957	.7391	.0434
11	.7391	.8696	.1305
12	.6957	.9130	.2173
13	.1739	.2609	.0870
14	.2174	.2174	.0000
15	.2609	.0874	-.1739
16	.8261	.7826	-.0435

4. Grade 1, Treatment 1

Item	Pretest	Posttest	Gain
1	1.0000	1.0000	.0000
2	.8421	1.0000	.1579
3	.2632	.2632	.0000
4	.7368	.9474	.2106
5	.3158	.4737	.1579
6	.6842	.8421	.1579
7	.7368	1.0000	.2632
8	.6842	1.0000	.3158
9	.7895	.8947	.1052
10	.8947	.9474	.0517
11	1.0000	1.0000	.0000
12	.8947	.8947	.0000
13	.2105	.1053	-.1052
14	.1053	.2105	.1052
15	.0526	.0000	-.0526
16	.7368	.8421	.1053

5. Grade 1, Treatment 2

Item	Pretest	Posttest	Gain
1	1.0000	1.0000	.0000
2	.7895	.7368	-.0527
3	.0526	.2105	.0579
4	.8974	.9474	.0527
5	.0526	.2632	.2106
6	.5263	.7895	.2632
7	.6316	.7895	.1579
8	.7895	.8947	.0152
9	.8421	.8421	.0000
10	.8421	1.0000	.1579
11	.8421	1.0000	.1579
12	.8947	1.0000	.1053
13	.1053	.0000	-.1053
14	.0526	.0526	.0000
15	.0000	.0000	.0000
16	.7368	.7368	.0000

6. Grade 1, Control

Item	Pretest	Posttest	Gain
1	1.0000	1.0000	.0000
2	.6500	1.0000	.3500
3	.1500	.4500	.3000
4	.7000	1.0000	.3000
5	.2000	.3000	.1000
6	.5000	.9000	.4000
7	.6500	.8500	.2000
8	.7500	.7500	.0000
9	.8000	.9500	.1500
10	.9000	.9000	.0000
11	1.0000	1.0000	.0000
12	.8500	.9500	.1000
13	.1000	.7500	.6500
14	.0500	.2500	.2000
15	.1500	.3000	.1500
16	.6500	1.0000	.3500

6. Grade 1, Control

Item	Pretest	Posttest	Gain
1	1.0000	1.0000	.0000
2	.6500	1.0000	.3500
3	.1500	.4500	.3000
4	.7000	1.0000	.3000
5	.2000	.3000	.1000
6	.5000	.9000	.4000
7	.6500	.8500	.2000
8	.7500	.7500	.0000
9	.8000	.9500	.1500
10	.9000	.9000	.0000
11	1.0000	1.0000	.0000
12	.8500	.9500	.1000
13	.1000	.7500	.6500
14	.0500	.2500	.2000
15	.1500	.3000	.1500
16	.6500	1.0000	.3500

7. Grade 2, Treatment 1

Item	Pretest	Posttest	Gain
1	1.0000	1.0000	.0000
2	.9630	1.0000	.0370
3	.3704	.4815	.1111
4	.9259	1.0000	.0741
5	.4074	.6296	.2122
6	.7778	.8519	.0841
7	.8148	1.0000	.0852
8	.8889	.9630	.0741
9	.9259	.9259	.0000
10	1.0000	.9259	.0741
11	.9259	.9630	.0381
12	1.0000	1.0000	.0000
13	.0370	.1852	.0482
14	.1111	.0741	-.0370
15	.0370	.1111	-.0741
16	.8148	.8889	.0741

8. Grade 2, Treatment 2

Item	Pretest	Posttest	Gain
1	1.0000	1.0000	.0000
2	.8889	1.0000	.1111
3	.2963	.5556	.2593
4	1.0000	1.0000	.0000
5	.4444	.6667	.2223
6	.7778	.9259	.1481
7	.8148	.9259	.1111
8	.8889	.9630	.0741
9	.9630	.9259	-.0371
10	.9259	.9630	.0371
11	.9259	.9630	.0371
12	.8889	.9630	.0741
13	.1111	.8519	.7408
14	.0370	.2593	.2223
15	.0000	.2963	.2963
16	.7778	.9630	.0852

9. Grade 2, Control

Item	Pretest	Posttest	Gain
1	1.0000	1.0000	.0000
2	.9259	.9259	.0000
3	.4074	.3333	-.0741
4	.9259	.9259	.0000
5	.3704	.4074	.0370
6	.9259	.6296	-.1963
7	.9259	.7778	-.1481
8	.9259	.8148	-.0811
9	1.0000	.6667	.3333
10	1.0000	.9259	-.0741
11	.9630	.8889	-.0741
12	.9630	.9259	-.0371
13	.2222	.1481	-.0741
14	.0370	.1852	-.0482
15	.0370	.1111	.0741
16	.8889	.8148	-.0741