DOCUMENT RESUME

ED 049 920 SE 010 966

TITLE Research Reporting Sections, National Council of Teachers of Mathematics Annual Meeting (49th, Anaheim, California, April 14-17, 1971).

INSTITUTION ERIC Information Analysis Center for Science Education, Columbus, Ohio.

PUB DATE Apr 71

NOTE 68p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29

DESCRIPTORS *Abstracts, College Mathematics, *Conference Reports, Educational Research, Elementary School Mathematics, *Mathematics Education, *Research, Secondary School Mathematics

ABSTRACT

Presented are abstracts of research papers presented at the Research Reporting Sections of the 49th Annual Meeting of the National Council of Teachers of Mathematics. Topics investigated in these research papers include: effects of reward upon children's responses, analysis of Piagetian area concepts, effects of laboratory materials, strategies in learning mathematics, sex differences in attitude and achievement, attitudes and socioeconomic class as predictors of mathematics achievement, effects of participation and motivation on problem solving ability, strategies of teaching mathematical concepts, behavioral objectives, effects of different homework assignments, teaching of non-decimal numeration in elementary school, student teacher's self-assessment, characteristics of mathematics teachers and their students' achievement, factors associated with instructor effectiveness in calculus, methods of teaching logic, readability of mathematical language, factors affecting problem solving in elementary school, measurement of critical thinking, and ability of college students to do proofs in logic. (CT)

RESEARCH REPORTING SECTIONS

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS

49th ANNUAL MEETING

Anaheim Convention Center
Anaheim, California
April 14-17, 1971

TABLE OF CONTENTS

PREFACE

PAPERS PRESENTED AT THE MEETING Page

PREFACE

The ERIC Information Analysis Center for Science and Mathematics Education has compiled abstracts of research papers to be presented at this conference. Some editing was done by the ERIC staff to provide a general format for the abstracts. Special recognition should be given to Dr. F. Joe Crosswhite, Dr. Richard Shumway, Mrs. Marsha Rice and Mrs. Maxine Weingarth who were responsible for compiling and preparing the report.

Many of the papers will be published in journals or be made available through the ERIC system. These will be announced in Research in Education and other publications of of the ERIC system.

April, 1971

Robert W. Howe
Director
ERIC Information Analysis
Center for Science and
Mathematics Education
The Ohio State University
1460 West Lane Avenue
Columbus, Ohio 43221

Sponsored by the Educational Resources Information Center of the United States Office of Education and The Ohio State University.

This publication was prepared pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their judgement in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent official Office of Education position or policy.

RESEARCH REPORTING SECTION 1

April 15, Thursday
1:30 - 2:30 P.M.

Leader: Richard J. Shumway, The Ohio State University, Columbus, Ohio.

Speakers:
1. Edward J. Davis, University of Georgia, Athens, Georgia, "The Effects of Reward on the Responses of Children in Making Transitive Inferences."
2. Joan Rines Needleman, Wayland Schools, Wayland, Massachusetts, "Scalogram Analysis of Certain Area Concepts Proposed by Piaget."
3. Gerald M. Weeks, University of South Florida, Tampa, Florida and William D. McKillip, University of Georgia, Athens, Georgia, "The Effect of Attribute Block Training on Second and Third Graders' Logical and Perceptual Reasoning Abilities."
4. Nicholas A. Branca, Stanford University, Stanford, California, "Strategies in Learning Mathematical Structures."

THE EFFECTS OF REWARD ON THE RESPONSES OF CHILDREN IN MAKING TRANSITIVE INFERENCES

Edward J. Davis and Douglas Owens
University of Georgia
Athens, Georgia

This study was designed to provide evidence on the following two questions: 1) In making transitive inferences, will the response patterns of five year old children change when the prospect of earning a reward enters into the testing situation? 2) If there is a difference in children's response patterns between reward- and non reward-offering testing situations, does the mode of earning the reward affect the performance? In particular do children, when told they must respond correctly to <u>all</u> the questions in order to obtain a reward respond differently than children promised a reward respond differently than children promised a reward for "trying real hard?"

A sample of 42 out of the 51 children attending two kindergartens were selected to participate in the study. Students were eligible for the study if they met two criteria: 1) a chronological age in the range from 65 to 75 months, inclusive; 2) attainment of satisfactory performance level for the relations being studied. These relations were: as many as, more than, and fewer than. In order to meet a performance level for acceptance into the study, the child had to respond correctly to at least four out of six test items on each relation. These items required the students to match sets of objects and to decide which relation held for the particular situation. The reward was the child's choice from a selection of inexpensive items selected to appeal to the age level (e.g. plastic trucks, airplanes, jack, bracelets, spacemen).

Prior to the testing situation all the children had been taught a series of seven lessons concerning the operational definitions and uses of the relations, as many as, more than, and fewer than, by one of the researchers and the children's regular teachers. These lessons were designed by the research staff. Each subject had also been individually tested and classified as either a high or low conserver on the basis of his responses to a series of 18 activities across the three relations.

Each subject was randomly assigned to one of the three relations and to one of the two reward situations. A test of six items dealing with transitive inferences using the assigned relation was individually administered to each child. The perceptual arrangement of the objects for every question was designed to introduce "conflict" into the situation. For example, the following is one of six questions concerning the relation "as many as."

Original arrangement of Blue, Red, and Green checkers on a cardboard sheet (designed for an item dealing with making a transitive inference for the relation "as many as.")

```
       R
           G
       R   G
           G
       R   G
           G
       R   G
       R

       R

     B B
     B B
      B B
```

E: Are there as many Blue checkers as Red checkers?

In every instance the child responded after matching the B's beside the R's. This was to be expected since the seven lessons had included many activities of this type.

> E: Please leave the B's alone and tell me if there are as many R's as G's.

The child in every case transported the R's into a 1-1 correspondence with the G's and then responded.

> E: Removes R's to a position off the cardboard--
> Now tell me, are there more B's than G's (Response).
> Fewer B's than G's? (Response) As many B's as G's? (Response)

The sequence of these last three questions was randomly assigned to each child on each item. Situations involving the transitive inference of the relations had not been a part of the lessons taught to the children.

The six questions asked of each subject were exact replicas of questions asked approximately two days earlier in a testing situation having no external rewards. Each subject had been asked to respond to six items on each of three successive days prior to this experiment. Each set of six questions concerned transitive inferences dealing with the same operation (as many as, more than, or fewer than), and included two items from each of three perceptual situations: conflict, neutral, and screening. In the screening situation, objects corresponding to the B and G checkers were placed in a styrofoam cup before the questions requiring a transitive inference were asked. Steffe and Owens*, in presenting the main results of this experiment report:

1) A significant difference ($p<.05$) in the difficulty of making transitive inferences for the three relations, with "as many as" being the least difficult relation and fewer than the most difficult;

2) A significant effect for level of conservation on the relation "as many as," with high conservers having a higher mean score on this relation than low conservers;

3) No significant differences between the mean scores of high and low conservers on the relations more than and fewer than;

4) A significant effect for perceptual situations with the conflect situation being the most difficult and neutral the least difficult.

The present study, designed to obtain a repeated measure on the six items from the perceptual conflict situation, resulted in precisely the same pattern of significant differences ($p<.05$) in the level of responses with respect to conservation and relations as did those subjects in the Steffe-Owens study. However, the number of correct responses on the repeated measure showed approximately a 20% increase for the reward situation. This gain is significant at the 1% confidence level. Thus, when a reward was present in the testing situation the level of performance increased significantly but the pattern of cognitive responses remained the same. The statistical technique employed was a three-way factorial analysis of variance using the factors conservation level, relation, and reward or non-reward mode.

Since there was a significant gain in achievement in favor of the testing situation characterized by the possibility of earning a reward an investigation into the relative effectiveness of the two reward modes was carried out. Gain scores indicated those children offered the opportunity to earn a reward by getting all the questions correct accounted for 80% of the increase in performance. A paired t-test on the repeated measure for the two different reward modes showed an increase in performance for both types of reward but this increase was only

*A Paper reporting this study was presented to the 1971 AERA Convention by Les Steffe and Doug Owens, Mathematics Education, University of Georgia, Athens, Georgia.

significant for the group required to get all responses correct (p<.01). When the scores for the 21 subjects under the requirement of getting all of the items correct were analyzed the responses with respect to level of conservation and difficulty of the relations were again consistent with those in the Steffe-Owens study. There was no significant difference between the mean gain scores for high and low conservers, nor between the mean gains for each of the three relations.

The results for this sample of children indicate that while performance on questions requiring transitive inference can be significantly increased by offering a reward for perfect performance, the pattern of correct responses with respect to conservation level and difficulty of relation remains unchanged. This result should be of interest to educational researchers investigating young children's logical abilities. The results involving significantly increased performance when a reward is offered for perfect performance may not be readily generalizable to a classroom since the reward was offered in a context of a 1-1 relationship between an adult and a child. This finding could be of interest, however, for tutorial teaching situations, both human and automated, if it is substantiated by further research.

SCALOGRAM ANALYSIS OF CERTAIN AREA CONCEPTS PROPOSED BY PIAGET

Joan Rines Needleman
Wayland Schools
Wayland, Massachusetts

This study was an investigation of Euclidean space and measurement concepts hypothesized, in the developmental theory of Piaget, to be prerequisite to an understanding of area measurement and its computation. The research was designed to provide information relating to the psychological framework within which children operate, from which educators might be able to assess a particular child's intellectual development, aptitudes, and readiness for various kinds of instruction relating to measurement concepts.

Jean Piaget theorized that a child develops concepts of space and geometry slowly and concurrently through stages. He found that specific concepts were acquired at certain ages, approximately, and that different levels of understanding of the same concept were attained by children of different ages, suggesting that concepts may be acquired in a hierarchical order. However, Piaget's empirical findings were based on experiments which failed to test the same children for the same concepts in the same order and in exactly the same language. The present research involved a developmental sequence of hierarchically organized tasks leading to the measurement and computation of area, presented individually to each of sixty-nine subjects in a semi-standardized procedure. The subjects formed a homogenous sample of all upper middle class male subjects from grades three through eight possessing a normal I.Q. (100-120).

Piaget divides a child's mental development into four stages: sensori-motor (birth to two years), preoperational thought (two to seven years), concrete operations (seven to eleven years), and the period of formal operations (about 15 years), the last stage bringing with it the beginning of hypothetical-deductive thought. Piaget identifies cognitive structures for each of these periods from the behaviors and responses of a child during his confrontation with a selected task requiring certain mental processes, and relates the child's cognitive structures so inferred to structures of modern logic.

Since this study developed out of observations that junior high school students experienced great difficulty with the concepts of area and perimeter, and, since Piaget's experiments seemed to indicate that formal thinking was required for the understanding of area, the present study was initiated to provide further knowledge of the evolution of area concepts.

Thirty-one tasks representative of concepts prerequisite to the acquisition of an understanding of rectangular area and its computation were selected from the work of Piaget and others. Each task was administered from a standard form of questionnaire providing for multiple choices to each of the sixty-nine subjects so chosen to be beyond the preoperational stage and not yet at the formal stage. The data consisted of drawings made by the subjects and a record of their responses and behaviors when presented with the series of tasks.

The data were dichotomously categorized as successes or failures based upon Piagetian criteria for attaining the specific concepts. The response patterns of these successes and failures for each individual subject were tabulated in several rows of a matrix, the columns representing the concepts, and each row representing a single subject. The rows and columns were then shuffled to permit a minimum of deviations from the general scalogram which resulted from these patterns. From the scalogram patterns of behaviors and responses for each child, information was obtained as to a particular child's level of attainment in connection with the concepts of measurement, and prediction was made as to his ability to succeed at tasks higher in the conceptual hierarchy.

The scalogram indicated a hierarchy of concepts attained in the following order: (1) seriating and conservation of area, simultaneously, (2) subdivision and counting of units, simultaneously, (3) unit iteration, (4) operational continuity, and (5) the area concept. A Guttman coefficient of reliability of 0.89 and a Green index of consistency of 0.907 were calculated.

All subjects attained the concepts of seriating and conservation of area; 73% of the subjects up to the age of eleven achieved counting of units; 76% of all subjects up to the age of thirteen acquired subdivision. Only 44% of the subjects acquired unit iteration; only 33% acquired the concept of operational continuity, and only 22% acquired the concept of area, based upon doubling the area of a square. Subjects over twelve years of age were shown to be more successful than those under twelve in acquiring concepts of subdivision, counting units, continuity, and area, but not for the concept of unit iteration.

The following conclusions were drawn:

A. A developmental scale of space and measurement concepts prerequisite to the acquisition of an understanding of rectangular area and its computation exists, as hypothesized by Piaget.

B. There is a significant relationship between the acquisition of the concept of operational continuity and that of area.

C. Many students of junior high school age have failed to acquire some of the concepts prerequisite to an understanding of area, as well as the area concept itself.

D. It was found that despite all effort, confusion still existed in the match of examiner's questions with the children's decoding of terminology, particularly with respect to the words: "a piece of chocolate", "points", "line", and "square". The role of experience provided another variable to be considered in the interpretation of the subjects' responses.

Instruments such as that of the research have potential for assisting teachers to identify a child's development stage and to evaluate his readiness to assimilate material. Knowledge obtained from the scalogram patterns can be used to assist grade placement and remedial programs for individual children, and can implement curriculum planning within and across grades.

THE EFFECT OF ATTRIBUTE BLOCK TRAINING ON SECOND AND THIRD GRADERS' LOGICAL AND PERCEPTUAL REASONING ABILITIES

Gerald M. Weeks
University of South Florida
Tampa, Florida

and

William D. McKillip
University of Georgia
Athens, Georgia

The main purpose of this study was to discover whether students exposed to attribute block training improved significantly in recognizing valid verbal conclusions drawn from a given set of verbal premises. A secondary purpose was to investigate the effect this training would have upon students' skill in solving reasoning problems of a nonverbal perceptual nature.

The subjects of this study were selected from the second and third grades at Clayton Early Elementary School, which is an experimental school sponsored jointly by the Research and Development Center of the University of Georgia and the Clayton County Board of Education. Subjects were assigned randomly to either experimental or control groups so that each group contained fifteen second graders and fifteen third graders. Two illnesses and one transfer from the second grade control group reduced the number of subjects in the experiment to a total of fifty-seven. The subjects in the study were slightly above the national average in socio-economic level and verbal maturity.

Two tests of reasoning abilities were employed in this investigation. A test constructed by the investigator, Test L, was employed as a measure of logical reasoning ability and Raven's Coloured Progressive Matrices test was employed as a measure of perceptual reasoning ability. Each test was administered as a pretest and as a posttest to both the experimental and control groups.

Eight weeks of attribute block training, 90 minutes per week, for the experimental group followed immediately after the pretests. The investigator was the only instructor for this training which closely followed guidelines proposed by Zolton Dienes and William Hull for using the attribute blocks. During the training period, the students in the control group met with their regular classroom teachers for additional work in mathematics. Posttests were administered immediately following the instructional period to both groups.

Fifteen null hypotheses were specified for this study and the Lindquist Type 1 design was employed to test these hypotheses. No significant difference in scores on the two tests and subsections was hypothesized for comparisons between the experimental and control groups, between groups at each grade level, and between grade levels in each group.

The results of the analyses indicated that the attribute block training had a strong positive effect at both grade levels, second and third, in the development of logical reasoning ability and perceptual reasoning ability. The effect in logical reasoning ability due to the attribute block training was more apparent in the development of quantificational logic than in the development of sentential logic. The ability to recognize invalid reasoning patterns utilizing the sentential theory of inference and the general theory of inference was also strongly affected. Although the gains on both tests and the various subsections were significant at both grade levels, there was an indication that the second graders profited more from the training than did the third graders.

STRATEGIES IN LEARNING MATHEMATICAL STRUCTURES

Nicholas A. Branca
Stanford University
Stanford, California

The purpose of the study was to investigate and extend the findings of Dienes and Jeeves reported in their monograph Thinking in Structures.[1] Dienes and Jeeves presented two groups of subjects -- adults and children -- with the task of identifying the rules of games embodying mathematical group structures. They classified the evaluations of the subjects into three groups and devised a coding system to identify strategies used. Evidence of the existence of a positive relationship between measured strategies and subjects' evaluations was reported. The present study was undertaken specifically to test their findings concerning strategies in learning mathematical structures and to determine whether subjects are consistent across structures and embodiments in the strategies they use and the evaluations they give.

Each of the one hundred adolescent girls was given three experimental tasks in a game-playing situation. The tasks were presented in three interviews with approximately two weeks between them. In each task the subject manipulated an apparatus that embodied a mathematical structure. The goal was to learn the rules of the game so as to make correct predictions about the outcome of each move. The interviewer kept track of the subject's moves in learning the game, the predictions she made, and the evaluations she gave of how the game worked.

The first task, a Color Game, had been used by Dienes and Jeeves and was based on the Klein group. The second task, a Map Game, had a network structure. The third task, a Light Game, also embodied the Klein group.

For the two group structure tasks, Dienes and Jeeves's scheme was used for classifying subjects' evaluations -- their views of how the game worked. Evaluations fell into three major categories, Operator, Pattern, or Memory, depending on whether subjects saw their moves as operators, looked for patterns in the plays, or merely attempted to memorize the plays. Evaluations of the network structure game were categorized as either Individual Roads or Detour Routes depending on whether the subject focused on parts of the network of considered it as a whole.

---

[1]Zoltan P. Dienes and Malcolm A. Jeeves, Thinking in Structures, (London: Hutchinson Educational, 1965).

The sequence of a subject's moves on each task was taken as a measure of the strategy she was using, and strategy scores were calculated from this sequence. Dienes and Jeeves's system was used for the two group structure tasks. A strategy scoring system was devised for the network structure task.

Distributions and cross-tabulations of the evaluations and the strategy scores confirmed Dienes and Jeeves's findings that the distribution of evaluations is ordered in decreasing frequency of occurrence as Pattern, Memory, Operator, and in decreasing efficiency, as measured by the mean length of play, as Operator, Pattern, Memory. Other findings by Dienes and Jeeves, such as zero correlation between measures of performance and intelligence and the existence of a relationship between evaluations and strategies as measured by their scoring system, were not supported. Consistency across tasks was most pronounced for measures of success and failure. Evaluations showed some consistency across tasks, especially across the two group structure tasks. Strategies tended not to be consistent across tasks, probably owing to inadequacies in the strategy scoring system.

RESEARCH REPORTING SECTION 2

April 16, Friday
10:30-11:30 P.M.

Leader: James M. Moser, University of Wisconsin, Madison, Wisconsin.

Speakers:
1. Lewis R. Aiken, Jr., Guilford College, Greensboro, North Carolina, "Sex Differences in Attitude and Achievement."
2. Andrea Troutman, Hillsborough County, Public School System, Charles M. Bridges, Jr., University of Florida, Gainesville, Florida, Thomas S. Tocco, University of South Florida, Tampa, Florida, "Student Attitudes, Perceived Parental Attitudes and Socio-Economic Status as Predictors of Junior High School Mathematics Achievement."
3. Joseph M. Gangler, University of New Haven, New Haven, Connecticut, "An Experimental Study of the Effects of Participation and Motivation on Problem Solving Ability."
4. Jack Wilkinson, University of Northern Iowa, Cedar Falls, Iowa, "An Experimental Study of a Mathematics Laboratory in Grade Six."

SEX DIFFERENCES IN ATTITUDE AND ACHIEVEMENT IN MATHEMATICS

Lewis R. Aiken, Jr.
Guilford College
Greensboro, North Carolina

The purpose of the research was to assess differences in attitude and achievement of males and females in a cross-sectional study of three age/educational groups: 181 eighth graders, 225 college freshmen, and 124 graduate students in education. The responses of all students to 90-100-item personal data inventories concerned with attitudes, identity, personal history, interests, opinions, and other characteristics which have been hypothesized to be related to achievement in mathematics were analyzed. Extensive analyses of school grades, test scores, and teachers' ratings were also performed on the eighth-grade group, and to a lesser extent on the other two groups. The results are presented in terms of male-female differences and similarities in means and intercorrelations of the cognitive and affective variables, and the wider implications of the data are discussed.

The working hypothesis of this investigation is that, although heredity is important in the formation of general cognitive ability, specific mathematical interest and ability factors are produced by positive transfer among various culturally reinforced, quantitative-type experiences. In line with this hypothesis, sex differences in mathematics attitude and achievement are interpreted in terms of "differential cultural reinforcement" and "masculine role," in contrast to "sex-linked, recessive gene" hereditary explanations that others have proposed. Thus, although teachers and mothers are seen to play an important role in mathematics attitude and ability, the position of the father proved to be highly significant. Peer attitudes, if influential, have primarily a negative effect. In addition, the junior high school years are viewed as a crucial time in the formation of such differences between the attitudes and abilities of the sexes.

The results of these investigations also indicate that attitude toward mathematics is quite broad, involving an active dislike for routine computations, as well as mathematical symbols, terms, and problem-solving tasks in general. Other nonintellective and intellective variables such as perseverance, reflectiveness/impulsiveness, self-confidence, and conformity/obedience were also found to be associated with mathematics attitude and achievement. Finally, the design and preliminary results of other ongoing studies concerned with the development and improvement of attitude and ability in mathematics are discussed briefly, and a comprehensive outline of proposed research on psychological factors in mathematics learning is presented.

STUDENT ATTITUDES, PERCEIVED PARENTAL ATTITUDES, AND SOCIO-ECONOMIC STATUS AS PREDICTORS OF JUNIOR HIGH SCHOOL MATHEMATICS ACHIEVEMENT

Andrea Troutman
Hillsborough County
Public School System

Charles M. Bridges, Jr.
University of Florida
Gainesville, Florida

and Thomas S. Tocco
University of South Florida
Tampa, Florida

The current study was undertaken to assess the relationship between junior high school mathematics achievement and (1) student perception of parental attitude toward mathematics, (2) student attitude toward mathematics and, (3) the socio-economic status of the student's family. A modified version of the Dutton attitude scale was used to measure student attitude toward mathematics. A scale designed for the study was used to measure student perception of parental attitude toward mathematics. The Duncan extension of the National Opinion Research Center scales was used to measure socio-economic status.

Research [Stright, 1960; Stephens, 1960, Murphy and Murphy, 1964, Lyda and Morse, 1963, Fedon, 1958;] tends to indicate that student attitudes play a part in determining success in mathematics. The present study attempts to capitalize on such findings and extend them into the domain of parental attitudes and socio-economic status group membership.

The attitudinal scales were administered to one hundred and fifty students of Thomas Jefferson Junior High School located in Miami, Florida. The students were randomly selected from the total population enrolled in mathematics courses. The sample drawn was checked and found to be representative of the four levels of mathematics instruction used at the school (from remedial to accelerated). The permanent files were utilized to obtain information used to determine the socio-economic status of the family as well as student achievement in mathematics.

Analyses were carried out on the entire group and upon males and females separately. Perceived attitudes of parents were analyzed by overall parental attitude and separately as mother's and father's attitude. Six hypotheses were formulated which led to numerous subhypotheses. The hypothesis which dealt with the direct relationship between student attitude toward mathematics and student achievement in mathematics. The remaining four hypotheses were rejected. The rejection of the hypothesis which dealt with the direct relationship between socio-economic status, attitude and achievement was at least partially attributed to the restricted nature of the range of the socio-economic status measures obtained in the sample.

It was concluded that student attitude toward mathematics directly related to student perception of parental attitude toward mathematics and that student achievement in junior high school mathematics is directly related to student attitude toward mathematics.

AN EXPERIMENTAL STUDY OF THE EFFECTS OF PARTICIPATION AND MOTIVATION ON PROBLEM SOLVING ABILITY

Joseph M. Gangler
University of New Haven
New Haven, Connecticut

Problem solving is a complex activity. Some of the inconsistencies in earlier problem solving research may be due to the lack of investigation of relationships between the variables that affect problem solving behavior and to the wide variety of tasks which have been used for the experimental problem situations. The present study is an experimental investigation of the effects and interactions of certain variables pertaining to both learning and problem solving behavior, using a type of task that possesses numerous advantages.

The subjects were 355 students enrolled in 21 sections of freshman mathematics courses. These subjects participated in a short learning situation in which they learned the use of twelve rules. They were then assigned six different tasks, which could be solved by a sequence of mediating steps obtained through application of the rules. The rules corresponded to certain axioms and theorems that were adapted from the statement calculus of symbolic logic. The tasks corresponded to conditional proofs of symbolic logic, but both the rules and the tasks were presented in the context of a cryptic, military code.

Five experimental variables were chosen: participation in the learning situation, motivation, time of day, intelligence, and mathematical background. The learning situation for the rules was a tape recorded lecture of two types. The first required overt participation on the part of the subject since they were asked to write the answers to a number of examples after applying a specific rule. The second type was the same as the first in all respects except that it involved only covert participation on the part of the subjects since they were asked to respond to the application of the rules to the examples by merely thinking about the answers. In both types there was immediate reinforcement by showing the subjects the correct answer before proceeding with the next example. The type of learning situation was randomly assigned to the 21 sections of the classes.

For the variable of motivation, ten of the sections were told that their performance in the learning and problem solving situations would count toward their mathematics grade in the particular class; the rest were told that it would not count. Eleven of the sections met before noon, and ten met during the afternoon. Time of day was considered to be a subsidiary variable, and its effects were examined in order to widen the basis of inductive inference. The subjects were classified into two categories of intelligence by the upper and lower halves of the distribution of their scores on a scholastic aptitude test. The subjects were also classified into two categories by the amount of high school and college mathematics that each subject had taken. This variable, mathematical background, was considered in an attempt to control the effect of the subject's attitude toward working with abstract symbolic material.

For the learning situation, the dependent variable was a test score used as a measure of how well the subject knew the concepts necessary to apply the rules. In the problem solving situation, three dependent variables were used: the mean number of correct steps per minute, the mean number of errors per minute, and the number of correct solutions. A complete five-way design was used for the statistical analysis of the experiment. The thirty-one main effects and interactions of the experimental variables were tested for significance, using the procedures of analysis of variance or analysis of covariance. Eleven of these effects and interactions were found to be significant for one or more of the four dependent variables. The main effects and the two variable interactions were interpreted by using the means of the dependent variables for the various classifications of subjects. Since four dependent variables were used, the results of the learning and problem situations were also examined for patterns of consistency.

The experimental results indicated that those subjects who participated overtly in the learning situation learned more and solved problems better than those who covertly participated. This effect of participation was significant for the number of errors made in applying the rules. In addition, informing some of the subjects that their work would count toward their class grade over-motivated them, resulting in anxiety that interfered with their learning and problem solving activity. This effect was significant for the learning situation and for the number of correct solutions. Whether the subject's activity took place in the morning or afternoon made the least difference of all of the experimental variables. The effects of the variables of intelligence and mathematical background, which were defined for high and low categories, were significant for all of the dependent variables except for the number of errors. A number of significant interactions were also found. For example, the effect of motivation in producing anxiety was greater for high intelligence subjects than for low intelligence subjects; participation had a greater effect on the over-motivated subjects than on the non-motivated subjects; and the low mathematical background subjects learned more by covertly participating than by overtly participating.

The conclusions of this study emphasize the complexity of both learning and problem solving behavior. Perhaps the most important single conclusion that is of interest to the classroom teacher is that overt participation in the learning situation apparently made the subject more flexible in problem solving behavior, and, consequently, resulted in significantly fewer errors. In addition, the implication of the conclusions of this study is that the type of participation as well as motivating forces play an important part in concept learning and in problem solving and also that the effects of these variables vary according to the type of individual involved in the activity.

AN EXPERIMENTAL STUDY OF A MATHEMATICS LABORATORY IN GRADE SIX

Jack Wilkinson
University of Northern Iowa
Cedar Falls, Iowa

The concept of a mathematics laboratory has become very popular in recent years. Many definitions of what constitutes a mathematics laboratory currently co-exist. For the purposes of the research reported in this paper, the definition of mathematics laboratories was: an approach which gives students experiences designed to enable them to analyze and abstract using an inductive process.

The purpose of the research was to examine the relative effectiveness of three different methods of teaching geometry in selected sixth grade mathematics classes. Three different main effects were identified in an analysis of covariance model. The main effects were: treatment group, sex of student, and I.Q. level of student. In addition to a statistical treatment of the main effects, a purpose of the study was to examine selected first and second order interactions generated by the statistical model. Fifteen null hypotheses dealing with both cognitive and affective measurements were tested.

Nine sixth grade classes taught by three teachers comprised the sample for the study. Each of the three teachers taught each of the three treatments.

Treatment I, (CO), the comparison group received a teacher-textbook instructional strategy. In the teacher-textbook strategy the text provided the content and the teacher used the teaching strategy she would normally use. However, teachers in CO classes were not allowed to use any laboratory instructional strategies.

Treatment II consisted of a laboratory instructional strategy, (LA). The students in these three classes used laboratory units housed in shoe boxes. These shoe boxes contained all the necessary hardware and software for that particular unit. The teachers' role in this treatment was one of a consultant. There was no formal lecture, and the teacher moved about the room answering questions and providing encouragement.

Treatment III was quite similar to LA treatment except that the shoe boxes contained one additional item. Each shoe box in Treatment III, (LC), contained a cassette tape. The tapes contained a verbatim reading of all the directions and questions printed on the worksheets in the respective shoe boxes.

The teacher's role in this treatment was the same as for LA treatment.

The study was conducted over a period of twenty school days.

The measuring instruments consisted of the NLSMA Grade 6 Attitude Scale and a geometry achievement instrument. The achievement test contained items from the grade 6 NLSMA plus all the items from the standardized tests which accompanied the sixth grade text book used in treatment I.

In order to examine interaction effects, the Lorge-Thorndike non-verbal scale was used to stratify the students into three I.Q. levels.

Because treatments were randomly assigned to classes, the classroom is properly defined as the experimental unit. All null hypotheses for this model were found to be tenable. However, there was a tendency for the comparison group, CO treatment, to do slightly better on the geometry achievement test. The IA treatment had slightly better adjusted post-test attitude scores.

Of more interest to this researcher were the first and second order interaction effects. When the classroom was defined to be the experimental unit, the total degrees of freedom was only nine. This made it impossible to test for any interaction effects.

In order to examine questions regarding interaction effects, it was necessary to define the individual student as the experimental unit, ($n = 232$).

Using the student as the experimental unit, the following null hypothesis was rejected at the .05 level of significance ($F = 2.65$): There will be no significant difference in pupil attitude toward mathematics due to the second order interaction of treatments, sex of pupils, and I.Q. level of pupils, when initial differences between pupils have been adjusted with respect to attitude toward mathematics.

Upon careful examination of the data, it became apparent that the above second order interaction effect identified three trends:

1) The attitude score gains for the middle and low I.Q. levels were greater than for the high I.Q. level across all treatments.

2) The greatest gains in attitude score were in the middle and low I.Q. levels in the laboratory treatment, (LA).

3) In both the LA and LC treatments, boys tended to exhibit greater gains than girls in attitude scores.

The relatively small F value (F = 2.65) associated with the above null hypothesis must be treated with caution. The calculated "multiple R^2", which accounts for the percent of variance attributed to each effect, was only 2.04% for the above second order interaction. It may be that we need to look more closely at the "multiple R^2" and less closely at significant F ratios in ANOV models, especially if the significant F is accompanied by a large n.

Three conjectures seem reasonable in light of the statistical findings plus the daily informal meetings between researcher and participating teachers.

1. On the basis of daily meetings with teachers, this researcher was getting glowing reports regarding the attitudes and performance of students in the laboratory group (LA). These anecdotal reports were not supported by the data collected via the attitude scale. It is reasonable to assume that perhaps the teachers were drawing conclusions about a class based on two or three observations. It also seems reasonable to question the effectiveness of a pencil and paper instrument to measure in the affective domain.

2. The middle and low I.Q. groups in the laboratory groups exhibited more gain in attitude scores than high I/Q. groups. This is not too surprising. Students with higher I.Q. scores have understood mathematics in a verbal setting. The physical world is of no real help to them - they don't need it. In fact, sometimes it slows them down.

3. The observation that boys had better attitude gains than girls seems reasonable also. Boys are supposed to be more mechanical--it's expected! Girls are more verbal and hence may not react as positively to physical activities in mathematics.

RESEARCH REPORTING SECTION 3

April 16, Friday
1:30-2:30 P.M.

Leader: Marilyn N. Suydam, The Pennsylvania State University, University Park, Pennsylvania.

Speakers: 1. Jane Allen Gaston, Shoreline Community College, Seattle, Washington, and John R. Kolb, North Carolina State University, Raliegh, North Carolina, "Teaching a Mathematical Concept: A Comparison of Three Strategies."

2. Dan C. Farris, The Williamsport Area Community College, Jersey Shore, Pennsylvania, "A Study of Related Instructional Objectives Classified by Modes of Representation."

3. Robert A. Laing, Western Michigan University, Kalamazoo, Michigan, "Relative Effects of Massed and Distributed Scheduling of Topics on Homework Assignments of Eighth Grade Mathematics Students."

4. Phyllis F. Kavett, Newark State College, Union, New Jersey, "A Study of the Teaching of Non-Decimal Systems of Numeration in the Elementary School."

TEACHING A MATHEMATICAL CONCEPT:

A COMPARISON OF THREE STRATEGIES

Jane Allen Gaston
Shoreline Community College
Seattle, Washington

and

John R. Kolb
North Carolina State University
Raleigh, North Carolina

Henderson (1967) has proposed that instructional dialogue used by teachers to promote the attainment of mathematical concepts can be analyzed and subdivided into functionally different segments. Each segment is called a move and a sequence of such moves comprises an instructional strategy for teaching a concept. Henderson identifies fourteen moves in two categories: characterization moves and exemplification moves.

In this study, three teaching strategies are compared to determine if there are differences in their effectivness in facilitating the acquisition and transfer of a mathematical concept. The strategies selected were adapted from three different sources. Each strategy can be fully described in terms of its component moves using Henderson's designations.

The first strategy, imitating the popular classroom practice of presenting mathematical concepts through definitions followed by several examples, consists of a characterization move which precedes one or more exemplification moves. This strategy is denoted CE.

The sequence of moves in a second strategy patterned after Gagne's (pp. 171-182; 1970) description of concept learning is of the form ECE. This strategy begins with several exemplification moves in which specific instances of the concept are presented along with the term designating the concept. The second stage of the strategy consists of characterization moves intended to isolate each of the concept's relevant attributes. These characterization moves are followed by repetition of the original examples, and then the subject is presented with novel instances which he must classify as either examples or nonexamples of the concept.

A concept formation strategy, based on trial and error procedures often used in psychological studies, contains only exemplification moves and is denoted E. The subject is presented with an example or nonexample of the concept; he must decide whether it is an instance in the concept's referent set. After each response, comparison with the correct answer is made and the subject moves to the next stimulus.

When comparing various instructional strategies incorporating differing concentrations of exemplification and characterization moves it is conceivable that the dependent measures may be biased in favor of one of the treatments. To explore the possibilities of such a bias three dependent measures were used: a vertical transfer test, an exemplification test requiring classification of instances as examples and nonexamples of the concept, and a characterization test requiring the identification of a property as either being or not being an an attribute of the concept. While successful completion of either the exemplification test or the characterization test is sufficient evidence to infer that the concept has been learned, only the vertical transfer test provides a neutral comparison of the three strategies.

The subjects participating in the study were 39 freshmen students enrolled in two College Algebra classes during the Winter Quarter, 1970, at Shoreline Community College. Experimental materials were administered separately to the 18 students in the first class and the 21 students in the second class during one of their regularly scheduled meetings. This sample implied criteria for the selection of the concept to be employed in this study and led to the choice of a partition of a set as the concept.

Experimental materials were incorporated into booklets each of which contained a short review of sets and set operations. The main part of the booklet contained instructional moves peculiar to each of the three instructional strategies. The final portion consisted of the three achievement measures.

Although the students comprising the two classes participating in the study were similar, the experimental conditions surrounding the two administrations proved to be dissimilar. The second administration should not be considered a replication of the first and, for this reason, the two administrations are distinguished as Experiment I and Experiment II.

Since a prediction of the relative order of treatment means on the vertical transfer test was not possible, a one-way analysis of variance was conducted to test the hypothesis of no differences among treatment means at the .05 level of significance. In both experiments this null hypothesis was accepted.

For the exemplification test the bias prediction that the strategy would be superior led to analysis in terms of orthogonal contrasts. The null hypotheses tested were: 1) there is no significant difference between the mean of the E group and the average of the means of the ECE and CE groups, and 2) There is no significant difference between the means of the ECE and CE groups. The F-ratios for the E vs. ECE and CE contrast were significant in both experiments thus substantiating the predicted results.

On the characterization test it was expected that the E strategy group performance would be less than that of the other two groups. The same two orthogonal contrasts and their accompanying null hypotheses stated for the exemplification were also used for the characterization test. No contrast revealed a significant difference on this test.

Rector and Henderson (1970) compared four instructional strategies for teaching eleven mathematical concepts. The sequences of characterization and exemplification moves used were of the forms: C, CE, EC, ECE. On the lowest of three levels of attainment measures, the C strategy group mean was significantly greater than each of the other three treatment means. No other comparisons were significant.

Several hypotheses can be forwarded to explain the apparent discrepancy between the superiority of the C strategy in Rector's study and the greater effectiveness of the E strategy in this study. Obvious differences such as the nature and difficulty of the concepts, number of concepts to be learned, level and ability of the subjects, length of instructional time, and the composition and sequencing of instructional moves in the comparison may account for the discrepancy. Another possible hypothesis is that the mixtures of characterization and exemplification moves are less effective than either exemplification or characterization moves alone.

More subtle biases might include the type of acquisition test used and its similarity to any one of the strategies, or the number of moves used in the strategy. In Rector's study, all strategies contained a constant number of moves - five. In this study, the E strategy contained 38 moves, far greater than the number of moves in the other two strategies. It is conceivable that when few moves are employed, a pure characterization strategy would yield the most information while an E strategy would yield so little information as to be practically useless.

REFERENCES

Gagné, R. M. The Conditions of Learning. New York: Holt, Rine- and Winston, 1970.

Henderson, K. B. A model for teaching mathematical concepts. *Mathematics Teacher*, 1967, *60*, 573-577.

Rector, R. E. and Henderson, K. B. The relative effectiveness of four strategies for teaching mathematical concepts. *Journal for Research in Mathematics Education*, 1970, *1*, 69-75.

A STUDY OF RELATED INSTRUCTIONAL OBJECTIVES CLASSIFIED BY MODES OF REPRESENTATION

Dan C. Farris
The Williamsport Area Community College
Jersey Shore, Pennsylvania

This investigation evolved from the study of learning hierarchies as proposed by Gagné. The hierarchy of learning tasks as revealed by Gagné's "task analysis" procedure served as a starting point for a more detailed consideration of particular learning tasks. A specific task was viewed as a unit of content for which a cluster of objectives was written in behavioral terms.

The objectives were written in three components: given, required performance, and criterion. A more precise analysis of objectives was made possible by classifying each objective into the modes of representation used in the given and required components. The modes of representation employed were those cited by Bruner: enactive (E), iconic (I), and symbolic (S). An objective so classified was represented by an ordered pair of modes, i.e. (M_i, M_o) where M_i is the input and M_o the output mode of representation.

One of the objectives used in this study is offered below as an example. This objective was classified as (S, I), i.e., symbolic input and iconic output.

Given	Required Performance	Criterion
A true statement of the form $\frac{a}{b}$ is equivalent to $\frac{c}{d}$ a,b,c,d natural numbers such that $a<b<25$, $c<d<25$ and four partially shaded rectangles exactly one of which shows the equivalence of $\frac{a}{b}$ and $\frac{c}{d}$	Select the rectangle which suggests that $\frac{a}{b}$ and $\frac{c}{d}$ are equivalent	4 of 5 items in time*

*no time limit set

The problem under investigation involved formulating selected objectives, writing instructional sequences to teach these objectives, then checking to see what effect the mastery of these objectives had upon the mastery of other selected objectives from the same cluster for which no explicit instruction was provided. Specifically, hypotheses 1 and 2 were given the mastery of (I, S) and (S, I) objectives from the same cluster (where at least one had to be acquired as the result of explicit instruction), then without explicit instruction, the subject would also master an (I, I) objective in the case of hypothesis 1, and an (S, S) objective in the case of hypothesis 2. A similar pair of hypotheses (5 and 6) involved the mastery of (E, S) and (I, S) objectives and the effect this had upon the implicit mastery of (I, E) and (E, I) objectives.

Another pair of hypotheses (3.1 and 3.2) predicted no significant effect upon the implicit mastery of (I, I) and (S, S) objectives due to the order of acquisition of (I, S) and (S, I) objectives. Similarly, two additional hypotheses (7.1 and 7.2) predicted no significant effect upon the implicit mastery of (I, E) and (E, I) objectives, due to the order of acquisition of (E, S) and (I, S) objectives.

From approximately seventh fifth grade students, a pool of forty subjects was formed on the basis of a selection test. From this pool a random sample of 24 students was chosen for the study.

Instructional sequences were written for the selected objectives using Coursewriter II, and subsequently, were mediated by an IBM 1500 instructional system. The objectives involved some basic concepts of fractions which these students would have normally studied in the course of their curriculum at this time. Each instructional action was carefully documented using the Ruleg system devised by Evans, Homme and Glaser.

Each subject was tested off-line on the consequent objectives which were to be learned without benefit of explicit instruction. All tests of the antecedent instructional objectives were given on-line.

A criterion referenced mastery level of 80% was used in conjunction with the binomial probability tests to test four hypotheses. The remaining tests involved order of acquisition and the Fisher Exact Probability test was used to make these comparisons.

The results did not support hypotheses 1.0, 2.0, 5.0, or 6.0. That is, mastery of the antecedent objectives apparently did not induce the mastery of consequent objectives for which no explicit instruction was provided. Hypotheses 3.1 and 3.2 were supported. That is, the order of acquisition of antecedent objectives did not significantly affect the implicit mastery of consequent objectives. It was not possible to test hypotheses 4.0, 7.0, or 8.0.

The results of this study imply that, subject to the limitations of the study, when the mastery conditions of (I, S) and (S, I) objectives from the same unit of content were met, mastery of (I, I) and (S, S) objects from the same unit of content cannot be taken for granted.

Furthermore, under the conditions of this study, the order of achievement of antecedent objectives (I, S) and (S, I) have no significant affect upon the achievement of consequent objectives (I, I) and (S, S).

RELATIVE EFFECTS OF MASSED AND DISTRIBUTED SCHEDULING OF TOPICS ON HOMEWORK ASSIGNMENTS OF EIGHTH GRADE MATHEMATICS STUDENTS

Robert A. Laing
Western Michigan University
Kalamazoo, Michigan

Mathematics educators generally agree that the experience a student receives while solving problems involving a given concept both enhances his understanding of that concept and improves its retention. Yet there are many characteristics of the assignment providing this experience about which very little is known. Consider, for example, the following questions:

Do students achieve better under a differentiated assignment technique?

Do problems which require verbalization of concepts by the student heighten his understanding of these concepts and facilitate later verbalization?

Do students achieve better under an assignment schedule which distributes his experience with a given topic over a number of days rather than concentrating this experience on one day's assignment?

Each of these questions has important implications for the textbook author, the classroom teacher and those responsible for his training. It was the purpose of this study to probe the last of the above questions through a comparison of the effects of two schedules for distributing topics on homework assignments in the mathematics classroom.

Since the late 19th century, psychologists have been comparing the relative effects of distributed and massed practice schedules on initial learning and retention of material at the lower end of the cognitive scale. Results of these studies conducted in controlled, laboratory settings reflect the superiority of the distributed practice technique in the area of rote learning. Dodes reported Berman's application of distributed practice schedules to homework assignments in 1953.

Studies conducted by Welborn and English (1937), and Reynolds and Glazer (1964), suggest that only a limited effect on the retention of meaningful material may be accomplished by increasing the number of repetitions during its initial learning.

On the other hand, spaced-review has been shown to facilitate learning and long-term retention in studies conducted within the classroom setting, Schunert (1951). Ausubel reasons that the forgetting that occurs over a period of interpolated activity has a facilitating effect on the learning and retention of meaningful material. In attempting to recall material partially forgotten, the learner becomes aware of areas of instability, ambiguity, and lack of discriminability. Thus, an additional encounter with the material enables him to concentrate his attention on those areas of the material where the confusion occurs and to learn to discriminate between this material and other areas which are interferring with its retention.

The assignment schedules compared in this study were designed analogous to the massed and distributed practice schedules probed in the psychology laboratory, and thus, automatically incorporated review into the daily assignments.

A restricted group of Columbus (Ohio) Junior High Schools was stratified on mathematical achievement and five schools selected to participate in the study. Ten teachers whose teaching assignment included at least two regular sections of eighth grade mathematics administered each treatment to one of their classes for a four month period. This resulted in a sample of 20 classes totaling 526 students.

Homework assignments for the two treatments were designed with identical problem content. The Treatment N schedule massed problems pertaining to a given topic and one concentrated assignment while Treatment D distributed these problems over a number of assignments according to a predetermined schedule.

Three consecutive textbook chapters were utilized for the study. The first, a geometry chapter, was used to accustom the students to the new assignment procedure and researcher-designed examinations in mathematics. The second two chapters, dealing with number concepts, served as material for the experimental phase.

Examinations over the chapters on number concepts provided two measures of initial learning. A third instrument was administered after three weeks of interpolated activity in order to allow for analysis of treatment effects on long-term retention. Reliability estimates of the examinations fell within a range of 0.788 to 0.841. Each of the instruments was divided into subtests on computational skills, comprehension, and applications. Corresponding subtests of each of the two chapter examinations were combined to provide one initial learning measure on each of the components of mathematical achievement.

Treatment groups were compared on each of the initial learning and retention measures. The students in the two groups who had completed at least 80% of the assignments were also compared for differential treatment effects. Data from the standardized testing program provided prior measures of non-verbal I.Q., computational skills, comprehension, and arithmetic application. Ability quartiles were defined for the latter three achievement components in order to test for interaction between treatment and ability level.

Each hypothesis of the study concerning terminal differences between treatment groups was tested using multivariate analysis of covariance. The four measures from the system-wide testing program were utilized as covariates with the level of 0.05 required for significance.

Results of the analyses of total test scores reflected no significant differences between the two treatment groups. However, differences between adjusted means consistently favored the distributed treatment group on each of the initial learning and retention measures with corresponding significance levels falling within a range of 0.084 to 0.157.

Comparisons between Groups N and D on the three components of mathematical achievement suggest a possible interaction between treatment and achievement component. Adjusted means on each of the component subtests favored the distributed treatment group. However, the only component subtest attaining significance was the instrument measuring initial learning of applications. Group differences on the retention subtest on applications approached signiicance at $P<0.103$.

The comparisons of treatment effects on the achievement of students in the various ability and class quartiles revealed no consistent interaction of treatment with ability level.

References

1. This study is detailed in a Ph.D. dissertation with the same title completed at The Ohio State University, 1970.

2. Ausubel, David P. Educational Psychology: A Cognitive View. New York: Holt, Rinehart and Winston, Inc., 1968.

3. Dodes, Irving Allen. "Planned Instruction," The Learning of Mathematics: Its Theory and Practice. PP. 303-334. Twenty-first Yearbook of the National Council of Teachers of Mathematics. Washington, D.C.: NCTM, 1953.

4. McGeoch, J. A. and Irion, A. L. The Psychology of Human Learning. New York: Longman, Green and Co., 1960.

5. Reynolds, J. H. and Glaser R. "Effects of Repetition and Spaced Review Upon Retention of a Complex Learning Task," Journal of Educational Psychology, LV (1964), 297-308.

6. Schunert, J. "The Association of Math Achievement with Certain Factors Resident in the Teacher, in the Pupil, and in the School," Journal of Experimental Education, XIX (1951).

7. Welborn, E. L. and English, H. "Logical Learning and Retention: A General Review of Experiments with Meaningful Verbal Materials," Psychological Bulletin, XXXIV (1937), 1-20.

A STUDY OF THE TEACHING OF NON-DECIMAL SYSTEMS OF NUMERATION IN THE ELEMENTARY SCHOOL

Phyllis F. Kavett
Newark State College
Union, New Jersey

An experiment to evaluate the teaching of non-decimal system of numeration in grades four and six was performed in Roselle and Elizabeth, New Jersey. Eighteen classes consisting of 430 students comprised the sample.

Teachers attended preparatory workshops of five weekly meetings. The student sample was divided into four randomly assigned treatment groups for each grade level. Three classes of each grade studied a unit on non-decimal systems; two classes studied a unit on intuitive geometry (non-computational); two classes studied a unit on the decimal system enriched by means of visual and manipulative aids; and two classes studied the regular program on the decimal system.

Tests were given before a teaching session of five to six weeks, again directly following the teaching period, and once more as retention tests seven weeks after the conclusion of the teaching period.

The following test instruments were employed:

California Test of Mental Maturity - Short Form - Level 2.

Pretest of Vision, Hearing, and Motor Coordination; California Test Bureau.

Stanford Achievement Test - Arithmetic - Level I and II: Forms X and W - Arithmetic Computation and Reasoning.

Non-Decimal Test. Developed by the investigator.

Analysis of variance and covariance were used for the arithmetic computation and reasoning tests. When a difference among group means was observed, the Scheffe Test was used to make comparisons.

A comparison of non-decimal test scores was made for each student's posttest and retention test scores using the Wilcoxon Matched Pairs Signed-Rank Test.

Intercorrelations among scores and other data were made using the Pearson Product-Moment Correlation.

Distributions of scores on the non-decimal tests were compared for the fourth and sixth grades with the Kolmogorov-Smirnov Two-Sample Test.

The immediate effect of the teaching treatments was the lowering of computation test scores for students in the non-decimal and non-computation groups. No significant difference among groups were observed on the retention test of computation.

Equal arithmetic reasoning group mean scores were achieved by all groups on the posttest. The non-decimal groups of both grades had significantly higher group means on the retention test, possibly indicating a positive transfer effect caused by study of the topic.

Retention of knowledge of non-decimal systems of numeration was greater among sixth grade students than among fourth grade students. This knowledge did not improve students' ability to answer questions on place value. Non-decimal groups test scores on the posttest showed significant correlations with arithmetic reasoning and computation test scores. Sixth grade students were generally more successful on the test of non-decimal systems than were the fourth grade students.

The greatest gains on intelligence quotient were achieved by the grade six non-computational treatment group; the least gain by the grade four enriched decimal group.

Intercorrelations of scores were slightly different for boys and girls, similar for different treatment and racial groups, and most dissimilar for groups separated according to teachers' judgements of degree of educational advantage.

This study was made possible through a Small Contract Grant from the Office of Education of the U.S. Department of Health, Education, and Welfare.

CONTRACT OEC-0-8-080008-1884(010)

GRANT NO. Non-Decimal Systems 27-4913

RESEARCH REPORTING SECTION 4

April 17, Saturday
9:00-10:00 A.M.

Leader: Leslie Steffe, University of Georgia, Athens, Georgia.

Speakers:
1. Kenneth J. Travers, University of Illinois at Urbana-Champaign, Urbana, Illinois, and J. Dan Knifong, University of Illinois at Urbana-Champaign, Urbana, Illinois, "Student Teacher Self-Assessment of Teaching Competencies."
2. William E. Geeslin, Stanford University, Stanford, California, "An Investigation of the Relationship between Characteristics of Mathematics Teachers and Student Achievement."
3. Gerardus Vervoort, Lakehead University, Thunder Bay, Ontario, "Report of an Empirical Investigation of Factors Associated with Instructor Effectiveness in Calculus."
4. Ronald D. Dettmers, Wisconsin State University, Whitewater, Wisconsin, "An Experimental Comparison of Three Methods of Teaching Mathematical Logic to Prospective Elementary Teachers."

STUDENT TEACHER SELF-ASSESSMENT OF TEACHING COMPETENCIES

Kenneth J. Travers and J. Dan Knifong
University of Illinois at Urbana-Champaign
Urbana, Illinois

The purpose of this exploratory study was to initiate the development of an inventory of teaching competencies for mathematics students who are enrolled in teaching methods or student-teaching courses. Such a list of competencies might be used by students as a guide to assessing their strengths and weaknesses as they prepare to assume teaching responsibilities. The inventory might also be of use to methods instructors and student-teaching supervisors as they devise appropriate activities for their teachers in training. Such an instrument might also provide useful data in the evaluation of methods or student-teaching courses. At the in-service level, the device might serve to assess professional growth of teachers.

Of all the courses in the teacher preparatory sequence, those relating to teaching methods commonly are the least favored by students and professional critics alike. It often is charged that there is little relationship between the content of the methods courses and those techniques which are most needed by beginning teachers.

In this study, there was a systematic attempt to identify those competencies which are judged by experienced mathematics teachers to be of significance. Once identified, such skills were classified and comparative data obtained from beginners and experienced personnel. Presumably, such empirical evidence would provide useful information to both student teachers and instructors.

The identification of competencies for effective teaching was done with the assistance of doctoral students in a mathematics teacher education seminar. The resulting list was subjected to further screening by consulting teachers from the public schools who were involved in the University's training of teachers of teachers program.

The present exploratory version of the scale asks respondents to rate, anonymously on a five-point scale, their own competence in performing each of 50 activities. Then, on a second scale, they are asked to assess the perceived importance of the same 50 items. This information was sought for two reasons: (1) As a guide to methods instruction. Such ratings are presumably an index of perceived competence, and might be used by instructors as they plan areas of emphasis in their methods courses. (2) And as a source of information concerning what student teachers perceive as being important. It was assumed that students without prior teaching experience have little information on which to make sound judgements as to the importance of certain pedagogical skills.

In the spring of 1970, the 50-item scale was administered to a class of 23 student teachers at the beginning of the "professional semester" (prior to methods instruction and student teaching) and at the end of the semester. Because of missing data, the final sample size was reduced to 15. To provide a source of comparative data, 34-experienced teachers of mathematics, some of whom were supervising teachers for the student teachers, also completed the scale.

Principal axis-factor analyses were performed on the ability data in an attempt to identify dimensions of competence sampled by the instrument.

Five factors for each set of data were found to account for 77 and 68 percent of the variance in the pre- and post-data, respectively. Rotation to simple structure and examination of items with loadings led to the following names for the factors for both sets of data: General pedagogy, Mathematics pedagogy, Managerial, Professional awareness and Computer technology.

Mean scale scores were obtained for student teachers and experienced teachers, and differences between means tested for significance.

Table I reports scale statistics on the ability data for student teachers on both occasions. The associated t statistics for correlated means are all significant beyond the .01 level.

Table II reports scale statistics on the ability data for the experienced teachers, and comparisons between means for the student teachers and the experienced teachers. All statistics comparing the student teachers at the beginning of the semester and the experienced teachers are significant beyond the .01 level and in favor of the experienced teachers with the exception of computer technology. At the end of the semester, none of the differences was significant, except, again, for the computer technology scale which was in favor of the student teachers.

Tables III and IV report statistics on the same scales as in Tables 1 and II, but for the importance data. None of the differences between means is significant for any of the groups, except, again for the computer technology scales.

Taking full account of the limitations of this study (e.g. small sample size and lack of refinement of the items) there is the distinct promise that several dimensions of teaching competencies can be identified and examined in the manner herein described. It is fair to surmise, that the "How Well Could I...." scale, even in this preliminary form, has desirable psychometric properties and holds promise for use in detecting changes in self-assessment of teaching competencies, such as may result from training.

The data here clearly indicate significant increases in student-teacher self-assessment of teaching competencies at the end of the professional semester (comprised of a methods course and student teaching). Such changes could be attributed to training effects of the semester, or to more insightful assessment of abilities, or to a combination of these and other factors. With respect to importance, the evidence suggests that student teachers rate teaching skills in a manner similar to experienced teachers.

Table I.--Ability Scales for Student Teachers

Scale		Pretest	Posttest	$\bar{x}_1 - \bar{x}_2$	S.e.	*t(paired)
General pedagogy (Item N = 18)	$\bar{x}$	2.00	1.37	.63	.15	4.05
	SD	.74	.31			
	rel.	.95	.96			
Mathematics pedagogy (Item N = 8)	$\bar{x}$	2.63	1.59	1.04	.17	6.22
	SD	.75	.38			
	rel.	.72	.78			
Managerial (Item N = 13)	$\bar{x}$	2.03	1.35	.68	.14	4.88
	SD	.63	.32			
	rel.	.87	.86			
Professional awareness (Item N = 8)	$\bar{x}$	3.04	1.98	1.06	.15	6.93
	SD	.80	.56			
	rel.	.93	.76			
Computer technology (Item N = 3)	$\bar{x}$	2.89	1.76	1.13	.22	5.26
	SD	1.15	.51			
	rel.	.83	.85			
Total (Item N = 50)	$\bar{x}$	2.32	1.52	.80	.13	6.37
	SD	.62	.25			
	rel.	.97	.89			

*$P(|t|>2.98) = .01$, df = 14

Table II.--Ability Scales for Student Teachers and Teachers

Scale		Teachers ($\bar{x}_3$)	$\bar{x}_1 - \bar{x}_2$	S.e.	*t	$\bar{x}_2 - \bar{x}_3$	S.e.	*t
General pedagogy (Item N = 18)	$\bar{x}$	1.28	.72	.14	5.14	.09	.08	1.24
	SD	.21						
	rel.	.61						
Mathematics pedagogy (Item N = 8)	$\bar{x}$	1.79	.84	.18	4.74	-.20	.14	1.48
	SD	.46						
	rel.	.73						
Managerial	$\bar{x}$	1.34	.69	.13	5.39	.02	.09	.19
	SD	.26						
(Item N = 13)	rel.	.76						
Professional awareness (Item N = 8)	$\bar{x}$	1.82	1.22	.19	6.53	.19	.16	1.04
	SD	.50						
	rel.	.71						
Computer technology (Item N = 3)	$\bar{x}$	2.66	.23	.37	.62	-.90	.33	-2.73
	SD	1.23						
	rel.	.87						
Total	$\bar{x}$	1.35	.77	.13	6.13	-.02	.08	-.25
	SD	.27						
(Item N = 50)	rel.	.87						

Key: $\bar{x}_1$ = Mean for student teachers at beginning of semester (N = 15)

$\bar{x}_2$ = Mean for student teachers at end of semester (N = 15)

$\bar{x}_3$ = Mean for experienced teachers (N = 34)

*$P(|t| > 2.68) = .01$, df = 47

Table III.--Importance Scales for Student Teachers

Scale		Pretest	Posttest	$\bar{x}_2 - \bar{x}_1$	S.e.	*t(paired)
General pedagogy (Item N = 18)	$\bar{x}$	3.69	3.89	.20	.22	.91
	SD	1.27	1.05			
	rel.	.99	.98			
Mathematics pedagogy (Item N = 8)	$\bar{x}$	3.42	3.41	-.01	.13	-.06
	SD	.80	.87			
	rel.	.86	.89			
Managerial (Item N = 13)	$\bar{x}$	3.38	3.54	.16	.19	.87
	SD	.87	1.00			
	rel.	.97	.98			
Professional awareness (Item N = 8)	$\bar{x}$	3.31	3.22	-.09	.17	-.54
	SD	.85	.97			
	rel.	.82	.93			
Computer technology (Item N = 3)	$\bar{x}$	2.58	2.93	.35	.24	1.50
	SD	.71	.88			
	rel.	.93	.96			
Total (Item N = 50)	$\bar{x}$	3.44	3.55	.11	.16	.75
	SD	.92	.91			
	rel.	.99	.99			

*$P(|t|>2.98) = .01$, df = 14

Table IV.--Importance Scales for Student Teachers and Teachers

Scale		Teachers ($\bar{x}_3$)	$\bar{x}_3 - \bar{x}_1$	S.e.	*t	$\bar{x}_3 - \bar{x}_2$	S.e.	*t
General pedagogy	$\bar{x}$	3.92	.23	.27	.85	.03	.24	.13
	SD	.65						
(Item N = 18)	rel.	.94						
Mathematics pedagogy	$\bar{x}$	3.29	-.13	.19	-.68	-.12	.19	-.61
	SD	.49						
(Item N = 13)	rel.	.57						
Managerial	$\bar{x}$	3.45	.07	.21	.32	-.09	.22	-.44
	SD	.56						
(Item N = 13)	rel.	.89						
Professional awareness	$\bar{x}$	3.16	-.15	.20	-.75	-.06	.21	-.28
	SD	.52						
(Item N = 8)	rel.	.59						
Computer technology	$\bar{x}$	2.12	-.46	.25	-1.83	-.81	.27	-3.00
	SD	.85						
(Item N = 3)	rel	.85						
Total	$\bar{x}$	3.47	.03	.19	.15	-.08	.19	-.48
	SD	.44						
(Item N = 50)	rel.	.94						

Key

$\bar{x}_1$ = mean for student teachers at beginning of semester (N = 15) *P(|t|>2.68) -.01, df = 47

$\bar{x}_2$ = mean for student teachers at end of semester (N = 15)

$\bar{x}_3$ = mean for experienced teachers (N = 34)

AN INVESTIGATION OF THE RELATIONSHIP BETWEEN CHARACTERISTICS OF MATHEMATICS TEACHERS AND STUDENT ACHIEVEMENT

William E. Geeslin
Stanford University
Stanford, California

Past research has shown that student ability and past achievement do not account for all the variation in future student achievement. A portion of the remaining variance in achievement is generally attributed to school environment variables such as textbook, teacher, and school facilities. The present study focused on one of the environmental variables, the teacher. Past research indicates that some teachers are "better" at producing high student achievement. This phenomenon is generally described as teacher effectiveness. The purpose of research in this area is to identify "effective" teachers and to identify and study the characteristics of these teachers. Knowledge of these characteristics would be most useful in teacher selection and training.

Most research in the past has not succeeded in identifying teacher characteristics that are significantly related to student achievement; it was hoped that by using a large national sample of mathematics teachers, definitive information on whether a relationship exists would be obtained. Past research has not investigated the correlations between various measures of teacher effectiveness; the present study investigates both correlations between various measures of teacher effectiveness and the stability of these effectiveness scores from one year to the next. This information should be useful in determining the adequacy of present methodology and measures of teacher effectiveness.

The present study examined the relationship between various teacher characteristics and student achievement. This study utilized the National Longitudinal Study of Mathematical Abilities (NLSMA) data bank as a data source. The sample consisted of all teachers who had returned the Teacher Opinion Questionnaire and their students from Year 1 of the X, Y, Z populations of NLSMA, and from Year 2 of the X and Y populations. The teachers returned these questionnaires on a strictly volunteer basis. This study used NLSMA data from grades 4, 7, and 10 from Year 1, and data from grades 5 and 8 from Year 2. In this sample, 234 teachers participated in both years of NLSMA. Texts were grouped into modern and conventional types. Teachers were grouped both by sex and by the text type they used within each grade level. This resulted in four teacher groups for each grade level; e.g., for each grade, one of the groups would be modern text-male teachers. A homogeneity of regression analysis was performed within each teacher group to determine whether or not boys and girls should be analyzed separately for that group. Two criteria, student computation and student comprehension, were used as measures of teacher effectiveness. Within each teacher group, students were partitioned into above average and below average on the basis of expected performance on each of the criteria. This resulted in either four or eight effectiveness scores for each teacher for each year he was in the sample. For example, if boys and girls were analyzed together within the teacher group, each teacher was assigned an effectiveness score on above average students' computation, below average students' computation, above average students' comprehension, and below average students' comprehension for each year the teacher was in the sample. If boys and girls were analyzed separately, then a teacher was assigned four scores for boys and four scores for girls, independently.

Regression methods were used to determine effectiveness scores. Each of the criteria was regressed on a set of student pre-test scales. The pre-test scales were not the same for each grade level, but included appropriate measures of IQ, mathematical achievement, and personality. A regression equation for each criterion was obtained for each teacher group. For each criterion, the residuals, averaged for each teacher, became the teacher's "effectiveness score." Within each teacher group, each of the teacher effectiveness scores for Year 1 was regressed on the set of teacher characteristics. The regression analysis described the relationship between teacher characteristics and teacher effectiveness. The teacher characteristics used were: number of years teaching, whether or not the teacher had children of his own, whether or not the teacher had a math major or minor, major field, theoretical orientation, attitude toward mathematics, creative approach versus rote, and need for approval. For teachers who were in both Year 1 and Year 2 samples, a correlation analysis was performed on all effectiveness scores to determine the relationship between the various measures of effectiveness and the stability of these scores from Year 1 to Year 2.

This paper presents a summary of the results. A few typical cases are discussed in detail. There was a wide range of effectiveness scores within each teacher group. In the majority of cases, the teacher characteristics were not significantly related to teacher effectiveness. The correlations between various types of effectiveness were low, in general. Also the correlation between Year 1 effectiveness and Year 2 effectiveness was low. The results suggest that present methods of measuring teacher effectiveness are not adequate. The study also suggests that if a relationship exists between teacher characteristics and teacher effectiveness, then there is a need for more refinement in measuring teacher characteristics as well as a need to look at different teacher characteristics

AN INVESTIGATION OF THE RELATIONSHIP BETWEEN CHARACTERISTICS OF MATHEMATICS TEACHERS AND STUDENT ACHIEVEMENT

William E. Geeslin
Stanford University
Stanford, California

Past research has shown that student ability and past achievement do not account for all the variation in future student achievement. A portion of the remaining variance in achievement is generally attributed to school environment variables such as textbook, teacher, and school facilities. The present study focused on one of the environmental variables, the teacher. Past research indicates that some teachers are "better" at producing high student achievement. This phenomenon is generally described as teacher effectiveness. The purpose of research in this area is to identify "effective" teachers and to identify and study the characteristics of these teachers. Knowledge of these characteristics would be most useful in teacher selection and training.

Most research in the past has not succeeded in identifying teacher characteristics that are significantly related to student achievement; it was hoped that by using a large national sample of mathematics teachers, definitive information on whether a relationship exists would be obtained. Past research has not investigated the correlations between various measures of teacher effectiveness; the present study investigates both correlations between various measures of teacher effectiveness and the stability of these effectiveness scores from one year to the next. This information should be useful in determining the adequacy of present methodology and measures of teacher effectiveness.

The present study examined the relationship between various teacher characteristics and student achievement. This study utilized the National Longitudinal Study of Mathematical Abilities (NLSMA) data bank as a data source. The sample consisted of all teachers who had returned the Teacher Opinion Questionnaire and their students from Year 1 of the X, Y, Z populations of NLSMA, and from Year 2 of the X and Y populations. The teachers returned these questionnaires on a strictly volunteer basis. This study used NLSMA data from grades 4, 7, and 10 from Year 1, and data from grades 5 and 8 from Year 2. In this sample, 234 teachers participated in both years of NLSMA. Texts were grouped into modern and conventional types. Teachers were grouped both by sex and by the text type they used within each grade level. This resulted in four teacher groups for each grade level; e.g., for each grade, one of the groups would be modern text-male teachers. A homogeneity of regression analysis was performed within each teacher group to determine

REPORT OF AN EMPIRICAL INVESTIGATION OF FACTORS ASSOCIATED WITH INSTRUCTOR EFFECTIVENESS IN CALCULUS

Gerardus Vervoort
Lakehead University
Thunder Bay, Ontario

The purpose of this study was to investigate whether certain factors are associated with instructor effectiveness in calculus. Achievement of students at the end of the first semester of calculus (or what was considered equivalent, the second semester of the combination analytic geometry and calculus) was measured by a 50 minute, 36 item multiple choice test: 18 questions involved primarily manipulative skills and 18 measured understanding of concepts and knowledge of theorems.

The mean A.C.T. mathematics score for each class was used as the predictor of performance for that class in an analysis of covariance. The resulting adjusted mean class scores were used to compare the relative teaching effectiveness of various classifications of instructors.

A total of 13 hypotheses were tested which related student achievement with the number of graduate hours of the instructor in Mathematics, in Education, and in Science, his highest academic degree, professional interest, number of publications, teaching experience at college and high school levels, number of students per section, and type of institution.

The conclusions were based on data about 73 instructors representing 29 Iowa colleges and universities and 3 institutions outside the State of Iowa.

No significant differences at the .05 level in instructor effectiveness were found with respect to the scores measuring manipulative skills of the students.

With respect to the scores measuring understanding of concepts and knowledge of theorems there was a significant difference at the .05 level in favor of the M.A. + 45 or better as compared to the M.A.; a significant difference at the .05 level in favor of those instructors with 70 or more graduate hours in mathematics as compared to those who had 10-29; a significant difference at the .01 level in favor of those whose interests included research as opposed to those whose professional interest was teaching only. Those with 6-9 years of college teaching experience were more effective at the .05 level of significance than those who had 20 or more years of college teaching. The other differences among the 'concept scores' were not significant.

With respect to the total score a statistically significant difference at the .05 level was found in favor of those classes with 10 or more students as compared to classes that were smaller than that. However other findings within the study indicate that this last result should be treated with some scepticism.

AN EXPERIMENTAL COMPARISON OF THREE METHODS OF TEACHING MATHEMATICAL LOGIC TO PROSPECTIVE ELEMENTARY TEACHERS

Ronald D. Dettmers
Wisconsin State University
Whitewater, Wisconsin

This research project was an investigation undertaken to determine the need for and the effectiveness of teaching concepts of mathematical logic to prospective elementary teachers in a mathematics content course at the college level.

In brief, the major purposes of the project can be considered to be a composite of six inter-related aspects.

Aspect 1: to propose an outline of logical concepts which can be subsumed under the broad heading of mathematical logic for prospective elementary teachers.

Aspect 2: to determine the effectiveness of three distinct methods of teaching logical concepts to pre-service elementary teachers.

Aspect 3: to document the logical concepts which pre-service elementary teachers acquire without any formal instruction of mathematical logic.

Aspect 4: to determine which logical concepts, if any, are learned through formal instruction of mathematical logic.

Aspect 5: to determine if any logical concepts are learned incidentally in a mathematics content course for elementary teachers with no explicit logic taught.

Aspect 6: to determine if instruction in mathematical logic can significantly affect the mathematics achievement of pre-service elementary teachers.

In light of the changing mathematical knowledge requirements of prospective elementary teachers based on the advent of "modern mathematics" in the elementary school curriculum the above research project has significance with respect to formulating answers to the following questions.

(a) Should mathematical logic be included in the pre-service mathematics curriculum for elementary teachers?

(b) What is the logical ability of prospective school teachers prior to receiving instruction in logic and after receiving instruction in logic?

(c) Does training in formal logic increase the prospective elementary teacher's knowledge of logic?

(d) Does the amount of training in formal logic affect the elementary teacher's knowledge of logic?

The experiment involved three treatments of teaching a required mathematics content course for prospective elementary teachers using a unit on mathematical logic developed by the writer. The three treatment groups were differentiated as follows:

Experimental Treatment C:	Students who were not explicitly taught the unit on logic:
Experimental Treatment E1:	Students who were explicitly taught the unit on logic with no further work on logic after completion of the unit.
Experimental Treatment E2:	Students who were taught the unit on logic with additional stress on the concepts of logic implicitly contained in the remainder of the course.

All treatment groups were administered a logic achievement test developed by the investigator both as a pre-test and as a post-test. In addition groups E1 and E2 were administered a technical logic achievement test shortly after the unit on logic had been completed and again as a retention test near the end of the semester. Each treatment group also was administered a mathematics achievement test near the end of the term.

The subjects participating in the experiment were students attending Wisconsin State University, Whitewater, Wisconsin, who were enrolled in the mathematics classes for elementary education majors during the second semester of the 1968-69 academic year.

Logic and mathematics achievement were analyzed by the analyses of variance and covariance using general educational ability as the covariate. The attainment of the various logical concepts was analyzed by the following procedure: If a treatment group correctly answered at least 65% of the items assigned to a particular concept, then the group was judged to have attained that concept; if a treatment group correctly answered less than 35% of the items assigned to a particular concept, then the group was judged to not have attained the concept.

The study indicated that the following logical concepts should be included in a unit of logic for elementary teachers: introductory concepts such as statements, variables, and open sentences plus quantifiers, conjunction, disjunction, negation of simple statement, the conditional, logically equivalent statements and a good discussion of inference and deduction.

The following results indicate the effectiveness of the three methods of teaching logic. It was concluded that the two treatment groups, E1 and E2, which were taught the unit on logic had achieved significantly higher, at the .01 level of significance, than treatment group C. The study also indicated that instruction in the unit of logic with reinforcement throughout the remainder of the course, increased logical achievement over instruction in logic alone, although the difference was not statistically significant.

There was no significant difference in mathematics achievement between those prospective teachers taught logic and those not explicitly taught logic.

The results given in the chart below document the logical concepts acquired with and without instruction in logic.

THE ATTAINMENT AND NON-ATTAINMENT OF LOGICAL CONCEPTS BY THE THREE TREATMENT GROUPS

Group C		Group E1		Group E2	
Concepts Attained	Concepts not Attained	Concepts Attained	Concepts not Attained	Concepts Attained	Concepts not Attained
IC	CD	IC	LOC	IC	LOC
Q	LOC	Q	CCI	Q	
CJ	CCI	CJ		CJ	
DJ	LE	DJ		DJ	
ID		ID		N	
				ID	

KEY:

IC = Introductory Concepts

Q = Quantifiers

CJ = Conjunction

DJ = Disjunction

N = Negation

CD = Conditional

LOC = Language of Conditional

CCI = Converse, Contrapositive, Inverse

ID = Inference & Deduction

On the basis of the preceding chart it was concluded that instruction in logic reduced the number of concepts which were definitely not attained. However, instruction in logic did not result in the attainment of the highly technical concepts of "language of conditional" and "converse-contrapositive-inverse." Thus, attainment of specific highly technical concepts is not as dependent upon instruction in logic as might be expected.

The results dealing with the number of years of high school mathematics instruction and achievement in logic and the results on retention of logical concepts will not be discussed in this summary due to lack of space, but these results will be included in the presentation.

RESEARCH REPORTING SECTION 4

April 17, Saturday
1:30-2:30 P.M.

Leader: James Wilson, University of Georgia, Athens, Georgia.

Speakers: 1. Robert B. Kane, Purdue University, Lafayette, Indiana, Mary Ann Byrne, University of Georgia, Athens, Georgia, and Mary Ann Hater, College of Mt. St. Joseph on the Ohio, Cincinnati, Ohio, "A Readability Formula for the Language of Mathematics."

2. Elton N. Thompson, California State College, San Bernardino, California, "Sex, Readability, Accessory Remarks, and Mental Ability as Factors in the Solution of Verbal Problems in Elementary School Mathematics."

3. Lars C. Jansson, The Pennsylvania State University, University Park, Pennsylvania, "The Development of an Instrument to Assess Critical Thinking Ability in Mathematics."

4. William H. Morgan, St. Andrews Presbyterian College, Laurinburg, North Carolina, "A Study of the Ability of College Mathematics Students in Proof-Related Logic."

A READABILITY FORMULA FOR THE LANGUAGE OF MATHEMATICS

Robert B. Kane
Purdue University
Lafayette, Indiana

Mary Ann Byrne
University of Georgia
Athens, Georgia

and Mary Ann Hater
College of Mt. St. Joseph on the Ohio
Cincinnati, Ohio

At the first research reporting sessions of NCTM a paper was presented which defined ways in which the cloze technique could be adapted to the language of mathematics (Kane and Hater, 1968). In that paper plans for a research program which would lead toward a readability formula tailored to the language of mathematics were outlined. The linguistic differences between ordinary English prose and the language of mathematics were analyzed (Kane, 1968) and the case was made for labeling readability formulas devised for ordinary English as invalid for evaluating textbooks and other materials written in the language of mathematics (Kane, 1970). Cloze scores were established as measures of reading comprehensibility in the language of mathematics (Hater and Kane, 1970). A symposium is scheduled in the program for AERA, Division D in February, 1971 to present progress reports on the measurement aspects of assessing readability in the language of mathematics. As a part of this symposium research based on 4000 Ss will be reported which will establish or reject cloze scores as acceptable measures of reading difficulty in the language of mathematics. Additionally, the vocabulary familiarity of approximately 1400 technical words and symbols from the language of mathematics will be reported. Data for this study were collected in a random sample of the junior high schools in the United States and involved 6000 students.

The goal of this program is the establishment of a procedure which may be used to predict the reading difficulty of instructional materials in mathematics. The opening paragraph of this proposal has documented our progress in establishing the components of such a formula. As a criterion variable in the procedure we will use mean cloze scores of 25 passages from mathematics text materials ranging from upper elementary through the senior high school.* We are quantifying a number of potential predictor variables. These include vocabulary familiarity (both mathematical vocabulary and ordinary English vocabulary), syntactic complexity measures such as mean sentence length and grammatical

*In the event that mean cloze scores are not clearly valid as measures of reading difficulty, then mean comprehension test scores will be substituted for them.

structure, and measures of the relative amounts of non-alphabetic symbolization to alphabetic text. The best linear combination of predictor variables will be selected by means of multiple regression analysis. The potential predictor variables will all be the sort which a textbook selection committee could use with a minimum of difficulty and without recourse to data collection among students.

The regression analysis together with the establishment of a set of directions for using the readability formula will not be complete until the end of February. Thus we will be unable to append the final chapter at the AERA symposium. We would appreciate greatly the opportunity of reporting the readability formula to NCTM.

References

Hater, M. A. & Kane, R. B. The cloze procedure as a measure of the reading comprehensibility and difficulty of mathematical English. Paper presented at annual meeting of AERA, Minneapolis, February, 1970.

Kane, R. B. The readability of mathematical English, Journal of Research in Science Teaching, 5, 296-8, 1968.

Kane, R. B. The readability of mathematics textbooks revisited. The Mathematics Teacher, 63, 579-81, November, 1970.

Kane, R. B., & Hater, M. A. Adapting the cloze technique to study the readability of mathematical English. Paper presented at annual meeting of NCTM, Philadelphia, April, 1968.

Kane, R. B., et al. Measuring readability in mathematical English. Symposium to be presented at annual meeting of AERA, New York; February, 1971. (See Educational Researcher, 21, p. 97, November, 1970 for program.)

SEX, READABILITY, ACCESSORY REMARKS, AND MENTAL ABILITY AS FACTORS IN THE SOLUTION OF VERBAL PROBLEMS IN ELEMENTARY SCHOOL MATHEMATICS

Elton N. Thompson
California State College
San Bernardino, California

The purpose of this inquiry was to study the effects of the factors of sex, readability, accessory remarks, and mental ability, plus their interactions, on verbal problem solving in elementary school mathematics.

The inquiry has significance for persons concerned with mathematics education in the elementary school; teachers, curriculum directors, specialists in mathematics instruction, principals, supervisors, and textbook writers would be able to use the results of the study to improve textbooks containing sets of problem pages, to improve teacher-made tests and sets of problems, and to assist in the general teaching of problem solving.

The factor of mental ability was defined by the intelligence quotient a subject derived from his performance on the California Test of Mental Maturity. Two levels of this factor were defined: high mental ability, embracing IQ's equal to or exceeding 111; and low mental ability, embracing IQ's of 99 or lower. The factor of sex was determined by the sex of the subject (boy-girl). Accessory remarks are common words used casually in advance of a problem solving situation. The two terms, Difficult Puzzles and Easy Arithmetic, were used as part of the heading of various problem pages and were expected to produce different sets in the students reading the headings and solving the problems. Readability was defined as reading ease. A set of ten arithmetic problems, involving the four fundamental operations with whole numbers, were written at two levels of readability. One version was written at the second-to-third-grade level of readability, and a second version of the same problem was written at the eighth-to-ninth-grade level of readability as measured by the Dale-Chall and Spache readability formulas.

The four factors, all defined at two levels, suggest a 2^n factorial design. A factorial design can be used to determine the main effects of the factors as well as their interactions. The analysis of variance of the 2^n design described in Edwards[1] was used in this inquiry.

Figure 1 pictures the model of the research design. Cell GIRP contains subjects (girls) of high mental ability working on problems of eight-to-ninth-grade reading difficulty that are labeled Difficult Puzzles. Cell BIRa contains subjects (boys) of high mental ability

[1]Allen L. Edwards, Experimental Design in Psychological Research (New York: Rinehart and Company, 1958), pp. 229-233.

working on problems of eighth-to-ninth-grade reading difficulty that are labeled Easy Arithmetic. Other subjects and their tasks may be found accordingly.

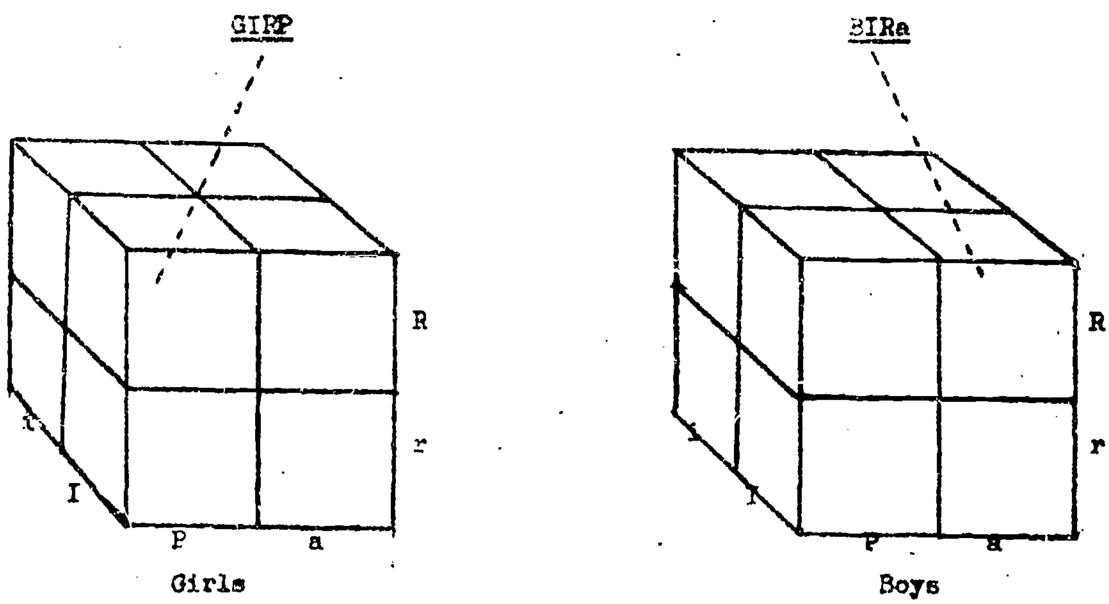

FIGURE 1

MODEL OF THE RESEARCH DESIGN

The two levels of readability and the two levels of accessory remarks combine to produce four types of problem pages: difficult reading-easy arithmetic, difficult reading-difficult puzzles, easy reading-easy arithmetic, and easy reading-difficult puzzles. With the aid of a table of random numbers the different types of problem pages were placed, in random order, in packets of forty; then they were distributed to the schools. Teachers administered the tests with the instructions to distribute the problem pages to the children in the order they found the pages in the packets supplied. Random distribution was later checked statistically, and the results indicated that the distribution of the problem pages was unbiased and random. The subjects were children at the sixth-grade level. Their teachers supplied the data concerning I.Q. and sex.

A total of 467 subjects provided original data. When subjects of 100 to 110 I.Q. were eliminated, a set of random numbers was used to assist in equalizing the subjects in each cell (see Figure 1) at 15 per cell. Thus, 240 subjects contributed to the data used in the study.

A two-part analysis of variance of the scores for the sixteen groups of subjects resulted in an *F* of 10.38 ($p<.001$). Therefore, the sources of these variations, the factors and their interactions, were more closely inspected.

An analysis of the fifteen sources of the variations noted (four main effects, six first-order interactions, four second-order interactions, and one third-order interaction) was made. Mental ability developed an *F* of 132.35 ($p<.001$), readability an *F* of 6.35 ($.01<p<.025$), the interactions of mental ability and readability developed an *F* of 3.76 ($.05<p<.10$). All other *F's* attained lower levels of significance. Generally, these findings are in harmony with other studies dealing with these factors.

Clearly, mental ability (as defined in this study) is the single most important factor affecting problem solving. Readability, as determined by the Dale-Chall and Spache readability formulas, is also a significant factor. These findings suggest that the typical written or verbal problem found in the usual elementary school mathematics text is not suitable for many children. One may hypothesize that children of lower mental ability might profit from work with problems arising out of first hand experiences, as in the Nuffield and similar programs. Readability is a factor at all levels of mental ability but the interaction of this factor with mental ability shows it is felt most heavily at the lower ranges of mental ability. Texts and other sources of written problems should be carefully checked to ensure that the reading ease of these materials more closely fits the reading abilities of the children for whom they are designed. All children have a reading ability level at which they feel comfortable and do much of their independent reading. More study is needed to determine how far the level of reading ease of mathematics materials may go beyond this level of independent reading and not block or seriously impair problem solving.

The two interactions very nearly attained a $p = .05$ level of significance. Girls appear to be somewhat more affected by accessory remarks than are boys. Teachers (and others) should be sensitized to the differential effects of accessory remarks. More study is needed to determine the various accessory remarks which may more forcefully affect problem solving.

THE DEVELOPMENT OF AN INSTRUMENT TO ASSESS CRITICAL THINKING ABILITY IN MATHEMATICS

Lars C. Jansson
The Pennsylvania State University
University Park, Pennsylvania

This study began with a concern for ways of teaching critical thinking within mathematics in the schools. Clearly, if teachers are unable to think critically in mathematics, they will experience difficulty in actively teaching children to do so. The development of this ability in teachers requires assessment of the ability. It was to the construction of such an assessment instrument that this study was directed. More specifically, answers were sought to the following major questions:

1. Can an instrument which possesses construct validity be developed to assess critical thinking ability in mathematics?

2. Given an affirmative answer to Question 1, how well do the following groups think critically in mathematics?

   (a) Prospective Secondary School Mathematics Teachers

   (b) Prospective Elementary School Teachers

   (c) College Freshmen

   (d) High School Seniors

In order to develop an instrument to assess critical thinking in mathematics it was first necessary to clarify two kinds of meanings, viz., (1) mathematics and (2) critical thinking. Ennis (1962) has derived a set of criteria for critically assessing statements. Based on a root notion of critical thinking as the "correct assessing of statements," Ennis' criteria were modified to fit a characterization of school mathematics in terms of certain types of statements made in the classroom. On the basis of logical and practical considerations, the following categories of critical thinking in mathematics were identified:

I. Ambiguity	IV. Definition
II. Assumption	V. Conjecture
III. Deduction	VI. Model Selection

Following the development of this theoretical framework, a pool of items was written on the basis of the detailed definitions of these categories, and submitted to a panel of experts for purposes of content validation. The items were then administered to prospective teacher groups and item analyzed for difficulty and discrimination. Thus on the basis of (1) the judges' agreement, (2) item discrimination, and (3) item difficulty, items were selected for inclusion in the final revised test of 50 items, named the Jansson Assessment of Critical Thinking Ability in mathematics (JACTAM).

The answer to Question 1 was in the affirmative on the basis of answers to four subquestions concerning reliability and validity. First, logical arguments and some examples were presented to show that Ennis' twelve aspects could be modified to six categories suitable for school mathematics. Second, it was argued that the test items were correctly keyed and in appropriate categories. Content validity was considered by submitting the items to a panel of judges and discarding inappropriate ones. Third, reliability was examined by use of Kuder-Richardson Formula 20 reliability coefficients. The overall reliability for all groups combined was .78.

Finally, correlations of the JACTAM scores with those of the Scholastic Aptitude Test in Mathematics (SAT-M) and the Cornell Critical Thinking Test, Level Z (Ennis & Millman, 1961) together with factor and item analyses indicated that the JACTAM is construct valid. The correlations were positive and significant ($p<.05$) as predicted, while the factor analysis showed all six variables (categories) highly loaded on a unitary factor, which was labeled "critical thinking in mathematics." Intercorrelations of categories were positive and significant, but not high. Item difficulty and discrimination indices were all within acceptable ranges.

Means for each of the four groups noted in Question 2, as well as the combined group, were presented. No comparison of group means was undertaken as part of the study. All groups combined yielded a total of 258 subjects for the study.

In addition to these analyses of the test as a whole, two additional questions concerning the individual test categories were considered:

4. How reliable (internally consistent) are each of the six parts of the instrument?

5. Do each of the six parts possess construct validity?

Reliability of the individual test-part categories was less clearly exhibited than for the test as a whole. Most of the individual parts consisted of only 8 items. Further study of abilities within these narrower critical thinking categories appears to be warranted.

Construct validity of the test parts was examined by consideration of the correlation matrix calculated for the factor analysis and by looking again at correlations of scores with the SAT-M and the Cornell Critical Thinking Test. No clear answer could be made to Question 5 on the basis of the data obtained.

The study was naturally subject to various limitations, not the least of which is the content background of the students. This was a first approximation to a suitable definition of critical thinking in mathematics. Clearly further studies involving both instruction and evaluation need to be made.

REFERENCES

Ennis, R. H. A Concept of Critical Thinking. Harvard Educational Review, 1962, 32 (Winter), 81-11.

Ennis, R. H., and Millman, J. Cornell Critical Thinking Test, Level Z (experimental ed.). Ithaca, New York: The Cornell Critical Thinking Project, 1961 (mimeographed).

Note: This report is based on the work done for the author's doctoral dissertation at Temple University.

A STUDY OF THE ABILITY OF COLLEGE MATHEMATICS STUDENTS IN PROOF-RELATED LOGIC

William H. Morgan
St. Andrews Presbyterian College
Laurinburg, North Carolina

Six colleges were randomly selected from eleven state colleges of the University System of Georgia; two from each of the student population categories of 0-2000, 2000-4000, and 4000-6000. Three additional categories were formed according to quarter hours of mathematics completed; 0-15, 16-30, and above 30 hours (30+). Ten students were selected, using stratified random sampling procedures, for each of the nine subcategories defined.

The primary purpose of the study was to determine the extent to which the undergraduate mathematics majors recognized:

1) equivalent forms of the conditional statement (contrapositive and disjunctive forms) and the negation of the conditional,

2) the invalidity of the invalid inference patterns of Denying the Antecedent and Affirming the Consequent, and

3) the starting assumption for a direct proof, contrapositive proof, and a proof by contradition.

The effects of Content (mathematics and real world) and Patterns (see 1 and 2) were ascertained in conjunction with the categorization variables of quarter hours and college size. A fifty-item Logic Test and a nine-item Proof Test were constructed.

An analysis of variance repeated measures design was used to determine the effects of college size, hours of mathematics completed, Content, and Patterns. For the Logic Test (patterns) the main effects of Hours and Size were significant ($p<.01$) with no interaction. Students in the 4000-6000 category scored better ($p<.01$) than the 0-2000 category students. In the three hour levels, students in the 30+ category scored better ($p<.01$) than students in the 16-30 category who scored better ($p<.01$) than students in the 0-15 category. In general, the students performed better on the real world items than the mathematics items and found the Contrapositive of the Conditional and Negation of the Conditional the most difficult of the patterns. The interactions of Patterns with Content ($p<.01$) and Pattern with Hours ($p<.01$) were significant.

A STUDY OF THE ABILITY OF COLLEGE MATHEMATICS STUDENTS IN PROOF-RELATED LOGIC

William H. Morgan
St. Andrews Presbyterian College
Laurinburg, North Carolina

Six colleges were randomly selected from eleven state colleges of the University System of Georgia; two from each of the student population categories of 0-2000, 2000-4000, and 4000-6000. Three additional categories were formed according to quarter hours of mathematics completed; 0-15, 16-30, and above 30 hours (30+). Ten students were selected, using stratified random sampling procedures, for each of the nine subcategories defined.

The primary purpose of the study was to determine the extent to which the undergraduate mathematics majors recognized:

1) equivalent forms of the conditional statement (contrapositive and disjunctive forms) and the negation of the conditional,

2) the invalidity of the invalid inference patterns of Denying the Antecedent and Affirming the Consequent, and

3) the starting assumption for a direct proof, contrapositive proof, and a proof by contradition.

The effects of Content (mathematics and real world) and Patterns (see 1 and 2) were ascertained in conjunction with the categorization variables of quarter hours and college size. A fifty-item Logic Test and a nine-item Proof Test were constructed.

An analysis of variance repeated measures design was used to determine the effects of college size, hours of mathematics completed, Content, and Patterns. For the Logic Test (patterns) the main effects of Hours and Size were significant ($p<.01$) with no interaction. Students in the 4000-6000 category scored better ($p<.01$) than the 0-2000 category students. In the three hour levels, students in the 30+ category scored better ($p<.01$) than students in the 16-30 category who scored better ($p<.01$) than students in the 0-15 category. In general, the students performed better on the real world items than the mathematics items and found the Contrapositive of the Conditional and Negation of the Conditional the most difficult of the patterns. The interactions of Patterns with Content ($p<.01$) and Pattern with Hours ($p<.01$) were significant.

To answer questions raised by 1, 2, and 3, a performance criterion for the pattern subtests was selected as 7 correct responses out of 10 responses, while 3 out of 3 was the criterion for the proof subtests. The following results were obtained:

1) The average percentage meeting criterion on the three patterns involving equivalent statements was only 53 for students in the 30+ category and only 36 for students with fewer than 30 hours.

2) The average percentage meeting criterion on the invalid patterns was 70 for students in the 30+ category and only 37 for students with fewer than 30 hours.

3) On the Direct, Contrapositive, and Contradiction Proof Tests, the percentages meeting criterion were 80, 37 and 33 respectively for the 30+ category and 69, 15 and 5 respectively for the 0-30 category.

Of the 90 students, 18 had some training in logic. For the 72 students with no logic the percentages in 1, 2, and 3 can be replaced respectively with (51,36), (62,33), and (80,18,35), (68,7,3).

It can be concluded for the sample that mathematical experience was not a sufficient condition for learning all the patterns selected for the study nor all the proof strategies selected for the study. It should be noted also that mathematical experience was not necessarily responsible for the performance of the few students who did do well on the test. Indeed, one of the students was observed by the researcher to be using techniques learned in a logic course.

These findings raise serious questions as to the need for explicit teaching of mathematical logic and the logic of proof. Further research is needed to determine the role of ability in logic and proof in the successful completion of an undergraduate program in mathematics. The findings also raise questions as to the adequacy of the preparation of mathematics teachers since the population from which the subjects were taken also included those who were preparing for teaching careers.