R E P O R T   R E S U M E S

ED 011 977   95   AA 000 096

A STUDY OF THE EFFECTS OF EXPERIMENTAL PROGRAMS ON PUPIL ACHIEVEMENT OBSERVED DURING FIRST THREE YEARS OF THE PROJECT, SECONDARY MATHEMATICS EVALUATION PROJECT. INTERIM REPORT.

BY- ERICKSEN, GERALD L.   RYAN, JAMES J.
MINNESOTA STATE DEPT. OF EDUCATION, ST. PAUL

REPORT NUMBER MNL-TR-66-4   PUB DATE   NOV 66
EDRS PRICE MF-$0.09 HC-$2.00   50P.

DESCRIPTORS- *MATHEMATICS INSTRUCTION, *CURRICULUM EVALUATION, *ACADEMIC ACHIEVEMENT, SECONDARY EDUCATION, *TEACHER EXPERIENCE, TEACHER CHARACTERISTICS, *STUDENT ABILITY, ST. PAUL

A REPORT IS GIVEN OF THE FIRST 3 YEARS OF A FIELD STUDY CONDUCTED TO DETERMINE THE EFFECTIVENESS OF SEVERAL PROTOTYPE, SECONDARY MATHEMATICS PROGRAMS THAT WERE PRODUCED BY DIFFERENT CURRICULUM DEVELOPMENT GROUPS. MATHEMATICS TEACHERS WHO HAD NO PREVIOUS EXPERIENCE WITH "MODERN" OR EXPERIMENTAL APPROACHES TO MATHEMATICS TAUGHT A SELECTED GRADE-LEVEL CLASS WITH CONVENTIONAL MATERIALS FOR 1 YEAR, AND IN THE FOLLOWING YEAR TAUGHT TWO CLASSES OF THE SAME GRADE LEVEL, ONE WITH THE CONVENTIONAL METHODS, AND ONE WITH AN EXPERIMENTAL CURRICULUM SELECTED BY THE TEACHER FROM THOSE AVAILABLE. SOME TEACHERS TAUGHT A CONVENTIONAL CLASS AND AN EXPERIMENTAL CLASS USING THE SAME EXPERIMENTAL MATERIALS AT THE SAME GRADE LEVEL FOR A 2D YEAR. PUPILS WERE RANDOMLY ASSIGNED TO THE CLASSES. EFFECTIVENESS OF THE EXPERIMENTAL CURRICULUM WAS MEASURED BY A PRE- AND POST-TEST MEASURE GIVEN AT THE START AND END OF THE YEAR OF STUDY AND AGAIN AT THE START OF THE FOLLOWING YEAR AS A MEANS OF MEASURING RETENTION. RESULTS OF THE STUDY SHOWED THERE WERE FEW STATISTICALLY RELIABLE DIFFERENCES WITH RESPECT TO MATHEMATICS ACHIEVEMENT AND RETENTION BETWEEN STUDENTS INSTRUCTED WITH EACH OF THE EXPERIMENTAL PROGRAMS. SIGNIFICANT TEACHER DIFFERENCES OCCURRED FOR ALL CURRICULUMS. INITIAL PUPIL ABILITY WAS, BY FAR, THE MOST SIGNIFICANT FACTOR INVOLVED IN EITHER THE ACHIEVEMENT OR THE RETENTION STUDIES. (AL)

Effects of Experimental Programs On Pupil Achievement

Interim Report

SECONDARY MATHEMATICS EVALUATION PROJECT

Gerald L. Ericksen
James J. Ryan

Technical Report No. 66-4

Minnesota National Laboratory
Minnesota State Department of Education
St. Paul, Minnesota

November, 1966

SECONDARY MATHEMATICS EVALUATION PROJECT

Interim Report

A Study of the Effects of Experimental Programs on Pupil Achievement Observed During First Three Years of the Project

Gerald L. Ericksen

James J. Ryan

Minnesota National Laboratory
Technical Report No. 66-4

November, 1966

This research was supported by a grant (G-25164) from the National Science Foundation to the Minnesota Academy of Science.

Preface and Acknowledgments

This is one of several technical reports that will be prepared presenting the analysis and results of the Secondary Mathematics Curriculum Evaluation Project. This project was initiated in the fall of 1961 by Paul C. Rosenbloom, then in the Department of Mathematics, Institute of Technology, University of Minnesota, presently in the Department of Mathematics Education, Teachers College, Columbia University. The project had as its most general objective an assessment, primarily in terms of pupil achievement, of several experimental programs in secondary mathematics that had at that time been prepared. Data gathering and processing was carried out for the first several years of the project under the direction of Professor Rosenbloom with the assistance over this time of a fairly large number of individuals in both technical and professional capacities. A final project report would be the most appropriate vehicle for acknowledging these contributions. The authors of this report, who joined the project after it had been under way for several years, would like to acknowledge those who directly contributed to its preparation.

First, we would like to express our appreciation to the participant teachers who provided excellent cooperation at all of the various data gathering stages of this project.

We would like to acknowledge the helpful advice and assistance of our professional colleagues in the Minnesota National Laboratory who had various responsibilities with respect to this project: Paul C. Rosenbloom, Wells Hively, Gerald R. Rising, Ronald A. Weitzman, and Sara Page.

We are especially appreciative of the contribution of Miss Jean Havlish, who had major responsibility for the technical aspects of data organization, processing, and analysis from the outset of the project; of Miss Sharon McDonough, for her role in the data collection and organization as well as her assistance with clerical matters; and of Mrs. Idamae Biebighauser and Mrs. Donna Bourdon for their assistance with clerical and data organization activities. Mr. Eugene Somdahl provided very able assistance with the computer programming.

Gerald L. Ericksen
Statistical Consultant

James J. Ryan
Project Director

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. Introduction

The Secondary Mathematics Evaluation Project was initiated in September, 1961, as a field study of several recently developed experimental programs in secondary mathematics. The primary purpose of the study was to determine the effectiveness of each of several prototype programs produced by different mathematics curriculum development groups in terms of pupil achievement in mathematics.

The revision and development of curriculum materials in a given subject matter area has as its primary purpose greater or more extensive facilitation of achievement of the instructional objectives than the programs and materials previously in use. This project was undertaken to examine whether this purpose had been accomplished for certain experimental materials in secondary mathematics. The adequacy with which this question can be examined in empirical terms, however, depends to the greatest extent upon how adequately or appropriately the achievement indices used represent the instructional objectives of the alternate programs. The achievement measure for which data are reported in this analysis was a test available at the time the project was initiated.

In addition to achievement, which was represented in this study as the performance of the pupil at the end of the specified instructional year, consideration was also given to retention of the material learned during that year as indicated by performance at the beginning of the subsequent school year.

Mathematics classes in grades 7 - 12 from schools in a five state area (Minnesota, Wisconsin, Iowa, North Dakota, South Dakota) participated in the study.

This is a report of the analysis of the pupil achievement and retention data gathered for the first three years of this project.

II. Experimental Objectives

1. To determine the difference in achievement in mathematics between pupils instructed with a given experimental program and those at the same grade level instructed with the conventional materials that were otherwise being used.

2. To determine the differences in achievement for pupils of initially higher and lower mathematics proficiency between those in the experimentally and conventionally instructed classes at the same grade level. That is, it was considered possible that some experimental programs might be relatively more effective for pupils having a higher or a lower level of proficiency as represented by achievement scores obtained at the beginning of the school year.

3. To determine the effects of increased teacher experience with the experimental program on the achievement of pupils instructed with both experimental and conventional materials.

   A teacher's experience with the experimental program during a given year could increase his effectiveness in teaching the program during subsequent years.

It is also possible that exposure to a new instructional program might have a general effect upon the teacher's instructional competence which could influence the performance of the conventional classes being instructed during the same year or subsequent years.

4. To determine the differences in achievement among classes of pupils instructed with alternate experimental programs at the same grade level.

5. To examine each of the above questions in terms of retention of acquired knowledge as well as in terms of end-of-year achievement. That is, the experimental programs gave emphasis to instructional factors which could contribute to better retention of the material than with a conventional program.

III. Method

A. Experimental Materials

The experimental mathematics program materials that were included in this study were those developed for the secondary level under the auspices of:

a) The School Mathematics Study Group (SMSG).

b) The Ball State, Indiana Teachers College Mathematics Program (BSP).

c) The University of Illinois Committee on School Mathematics (UICSM).

d) The University of Maryland Mathematics Project (UMMaP).

Within each of these curriculum development projects, materials were developed which were appropriate for various grade levels over the range of secondary grades. Table 1 shows the programs that were available when the project was initiated and were included in the study.

Table 1

Experimental Program Materials Included at Each Grade Level

	PROGRAM			
GRADE	SMSG	UICSM	BALL STATE	UMMaP
7	X		X	X
8	X		X	X
9	X	X	X	
10	X	X	X	
11	X	X	X	
12	X	X		

X - indicates program available and introduced as experimental material.

Data were gathered and analyses carried out for classes instructed with materials for each of the available programs indicated for each grade level.

A list of the instructional materials (i.e. textbooks) used in the experimental classes has also been compiled and is included as Appendix B.

B. Participation Procedures and Sample

Teachers in all secondary schools in the five state area were invited to submit applications to participate via their school administrators. From among the applicants for the 1961-62 school year, teachers were selected to participate whose primary teaching responsibilities were in the area of mathematics and who had current responsibilities for instruction in several mathematics classes. Participation was restricted to teachers who had not had any previous in-service or institute instruction in any of the "modern" or experimental approaches to mathematics.

The teachers' first year of participation (1961-62) consisted of providing instruction with conventional mathematics materials at the grade level at which they would be using the experimental materials during the subsequent year. Prior to the second year, teachers were asked to indicate which experimental program they would like to use among those available at their grade level. These materials were then provided for one experimental class for each teacher for his second year of participation. During the second year of participation, teachers taught two classes at the same grade level, one with the experimental program materials and the other with the conventional materials the teacher would otherwise have used. Principals and teachers were asked to randomly assign pupils to these two alternate classes and procedural instructions for doing so were provided. Following the second year of participation, teachers were requested to continue to participate for another year by teaching the same experimental program in one class and the conventional program in another at the same grade level. A certain proportion of teachers continued to participate for a third year (the second experimental year) in this way.[1]

During each subsequent year of the project, a new set of applications for participation was distributed and an additional group of teachers (Phase 2) began participating in the project. For teachers entering the project during each successive year, the same procedure was followed as outlined above for the initial year participants, i.e. administration of tests only to a conventional class the first year, introduction of experimental materials to one class the second year with an additional conventionally instructed class at the same grade level serving as a control class.

1 A small number of teachers continued in the project by teaching the same class at the next higher grade level using materials from the same experimental program appropriate for that grade level. The data for the second experimental year for these teachers were not included in this analysis.

A mathematics achievement test was administered to pupils in each of the participating classes (including the first year conventional class) at the beginning (September) and end of the year (May) and at the beginning of the subsequent year. The pre-measure served as a control for the pupil's <u>initial level</u> of achievement or proficiency, the end of year as the measure of <u>achievement</u>, and the test at the beginning of the subsequent year as a measure of <u>retention</u> of material learned during the previous year.

The participation procedure that was followed provided for two conventional comparison classes for each experimental class taught by the teacher; i.e., the previous year conventional class and the same-year conventional class (termed the control class). This, in effect, represents three treatment conditions for each teacher. Information concerning the previous and same-year conventional classes provided a control for and an assessment of the possible effects of instructional experience with the experimental program on the instruction given to the control class.

Teacher participation in the manner outlined above represented the minimum condition necessary for the classes of a teacher to be included in the analysis.[2] In addition, as will be noted below, the sample of classes included in the analysis was further restricted when classes for a given teacher were examined for non-random assignment of pupils.

This analysis is, then, concerned with the classes of teachers who initiated their participation during the 1961-62 school year (Phase 1) or the 1962-63 school year (Phase 2). The period of participation for these groups covers a period of three school years, 1961-62 through 1963-64.

Classes for teachers in both phases 1 and 2 provided data concerning the effects of the experimental programs for the teachers' first year of experience with the experimental program. The second year classes for Phase 1 teachers who continued to participate provided data concerning the effects when teachers have had one year's experience with the experimental program.

Therefore, for purposes of this analysis, the Phase 1 teachers fall into two participation categories: those who participated for one year with the experimental program and those who participated for two years. In this interim report only the first year of participation for the Phase 2 teachers was included in the analysis.

2 Among the teachers indicating willingness to participate and receiving experimental materials to do so, a number were not able to provide the necessary beginning and/or end of year achievement test data for the specific classes required. Consequently, the data that were obtained for the classes of these teachers could not be included in the analyses.

IV. Experimental Variables

A. Dependent Variables

Measure of mathematics proficiency.

The measure of mathematics proficiency used in this study to assess achievement, retention, and as a premeasure control was the mathematics section of the Sequential Tests of Educational Progress (Educational Testing Service, 1957). This test reflects the instructional objectives being emphasized by educators at that time prior to the extensive logical evaluation of secondary mathematics curricula which provided the impetus for the development of the experimental programs being studied in this project. Consequently, the test may not reflect, at least to the same degree, certain of the instructional objectives that are given emphasis in the experimental programs. Nonetheless, in contrast to other achievement tests available at the time, the STEP mathematics tests represent an attempt to measure to a greater extent the understanding and application of more general mathematics concepts and skills by using problem solving tasks that rely less on rote memory and specific skills. Reviewers generally concede that the STEP mathematics tests were quite successful in this regard. The STEP publishers recommend that level 1 of the test by used for grades 13 and 14, level 2 for grades 10-12 and level 3 for grades 7-9. However, since these tests were developed to measure knowledge in a given subject matter area over a series of grades rather than subject or grade specific content and because previous experience and expert judgment suggested that at the higher grades for each recommended level the tests might not be sufficiently difficult to be sensitive to instructional effects, the separate levels of the tests were used for the following grades in this study:

Grade	Test Level
12	1
9-11	2
7-8	3

There was no indication in the data that the change in test level for grades 9 and 12 resulted in the tests being too difficult for pupils in these grades.

Alternate forms of the tests for a given level were administered at the beginning and end of the school year. The form used at the beginning of the year for a given grade was also used as the retention test at the beginning of the subsequent year.

Both converted and raw STEP scores were considered as dependent variables in this study (4). However, in general, converted STEP scores are not well suited to serve as a basis of curricula comparisons. The difficulty may be seen by examining the non-linear relationship between converted and raw scores at the lower end of the STEP scale. Fig. 1 gives the general form of the relationship. This irregularity was also reflected in the skewed frequency distributions of the converted scores referred to in Supplementary Report 100A.

The decision to abandon converted scores as a basis of comparison agrees with Stecklein who states in the Sixth Mental Measurements Yearbook (1):

> "The reviewer agrees with the criticism concerning the questionable utility of the converted scores and the need for them."

For these reasons all the analysis was based on raw STEP score data.

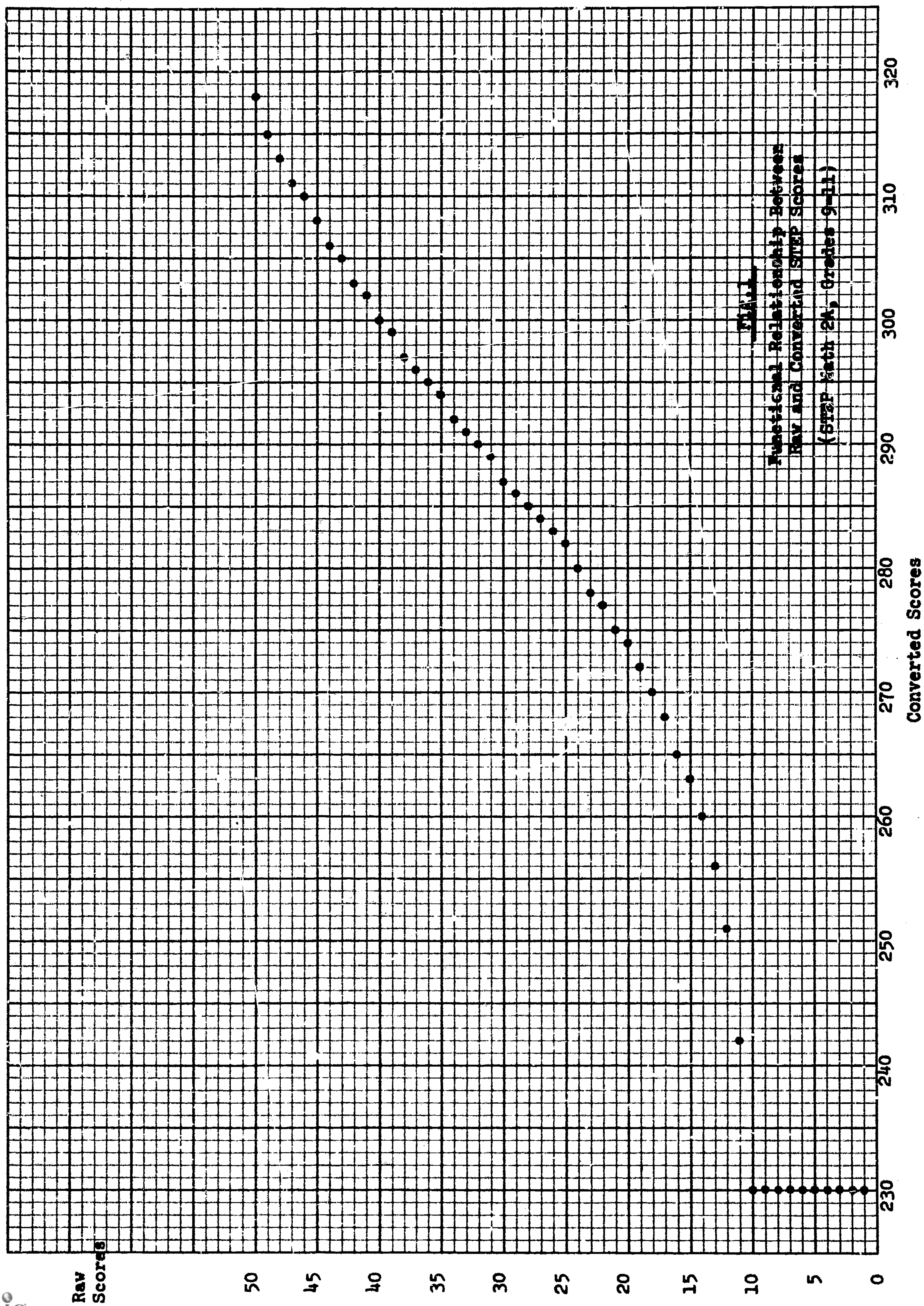

Fig. 1
Functional Relationship Between Raw and Converted STEP Scores
(STEP Math 2A, Grades 9-11)

B. Independent Variables

The following independent variables were considered in the study:

Curricula - Each experimental program (SMSG, BSP, UICSM and UMMaP) was analyzed in conjunction with conventional classes.

Teachers - Separate samples of teachers were analyzed as random factors for each experimental situation.

Pupil Ability - The prior ability levels of the pupils were established by STEP pre-testing in September for both the achievement and the retention studies.

Teacher Experience - Separate analyses were made of the effect of teacher experience with the experimental programs on pupil achievement scores.

Pupil Sex - A pilot study of the sex variable was also undertaken. The results indicated that the reduction in the error sum-of-squares was not sufficient, in general, to off-set the loss in degrees of freedom. Thus the experimental error was not substantially reduced and the analyses were conducted with the sex variable pooled together. The small sex differences resulted from a tendency for males to score higher than females as shown in Table 23.

V. Specification of the Experimental Design

A. Characteristics of Achievement Test (STEP) Data

The following are some relevant characteristics of the STEP data collected in the Secondary Mathematics Evaluation Project:

1. There was some evidence of inequality of regression coefficients between the Curricula x Teacher cells based on the regression of post-test on pre-test STEP scores. See Supplementary Report 100-I and reference 3.

2. STEP raw scores were approximately normally distributed within each experimental situation. See Supplementary Report 100-A.

3. Some teachers gave evidence of a lack of random assignment of pupils to classes. This initial bias is evident in the frequency diagrams of Supplementary Report 100-A. An attempt to remove these teachers from further analysis was made by both inspection and a χ^2 test. It is, of course, apparent that in this type of experimentation such a lack of randomization can seriously distort the analysis.

B. Analysis-of-Variance Models

The above data characteristics, in part, dictated the selection of an experimental design.

Initially, scatter diagrams were prepared for each experimental situation pre-test and post-test scores. Few specific conclusions could be drawn except, perhaps, that higher ability pupils had higher post-test scores and the relationship was generally linear.

Both gain-score-analysis and the analysis-of-covariance techniques were discarded because of previous evidence of inequality of linear regression adjustments for pupil ability across the curricula and teacher variables. See Supplementary Report 100-I.

The ability level of the pupil was then treated as an independent variable in the model rather than serving as a covariate. The basic model adapted was a three-way analysis-of-variance mixed model involving curriculum, teacher and pupil ability. Thus an individual pupil score was represented as follows:

$$Y_{ijkl} = \mu + a_i + \beta_j + \gamma_k + d_{ij} + e_{ik} + \phi_{jk} + f_{ijk} + \epsilon_{ijkl}$$

where teacher ($i = 1, 2, 3, \ldots, p$) was considered as a random factor whereas curriculum ($j = 1, 2, 3$) and pupil ability ($k = 1, 2$) were considered as fixed factors.

This model was followed for both the comparisons between experimental and conventional classes and the teacher trend analysis. For the comparisons between experimental programs teachers were, of course, nested within a given curriculum.

At the time of initial analysis the available computer facilities utilized a weighted means program (UMSTAT 61 of the University of Minnesota Computer Library). Comparisons between experimental and conventional curricula utilized this program. Later an unweighted means program was prepared by the Minnesota National Laboratory which was used for comparisons between the experimental curricula and for analyzing teacher trends. At the time of this report significant results of the weighted means analysis of achievement scores were also analyzed by the unweighted means program and included in this report. Generally, the two types of analysis were in close agreement. However, the unweighted analysis had the advantage of presenting the data in a manner consistent with a later interpretation of individual comparisons by the Newman-Keuls method. There is also some logic to the argument that, in this study, the unequal cell frequencies were not an inherent part of the experimental design. Thus for those experimental situations analyzed in greater detail, following the detection of a significant curriculum effect, the method of unweighted means was followed.

C. Data Sources

As described in section III, the following experimental situations were considered for the comparisons between the experimental and conventional curricula:

Second Year of Teacher Participation

The three curricula compared were the conventional class (C_1), the conventional class taught the second year (C_2), control class, and the experimental class (E_1) which was initiated at the start of the second year of the teacher's participation in the project. Some teachers did not continue in the project past the second year. These teachers are designated as Group 1a in Table 2. Those teachers continuing are designated as Group 1b.

Third Year of Teacher Participation

The three curricula compared were the conventional class taught the second year (C_2), the conventional class taught the third year (C_3) and the experimental class taught for the second time (E_2).

These results are summarized in Table 2.

Table 2

Data Sources for Comparisons Between Experimental and Conventional Curricula

Teacher Participation Category	Year of Project: 1961-1962	1962-1963	1963-1964
1a	C_1	C_2 , E_1	
1b	C_1	C_2 , E_1	C_3 , E_2
2		C_1	C_2 , E_1

For purposes of this study teacher data were pooled across comparable stages of the project. This interim report incorporates data from 1961 through 1964.

D. Data Coding

Table 3 lists the particular experimental situations analyzed in this study. The retention samples show some shrinkage since the tests were administered after the summer vacation and not all pupils were available in September of the following academic year. In one situation, retention data were available for a teacher where the achievement data were missing (Teacher 377 of the experimental situation involving teachers with no previous experience with the ninth grade SMSG materials).

The number of teachers listed in Table 3 for the achievement study is less than the total number of participating teachers for various reasons. Chiefly responsible for this shrinkage was the lack of randomization in placing pupils in the classes. Table 4 summarizes the losses incurred for this reason. Supplementary Report 100-A includes the details of the χ^2 tests of randomization together with the frequency distributions of pre-test scores.

Table 3

List of Experimental Situations

Experimental Situation			Number of Teachers		Number of Pupils									
					C_1		C_2		C_3		E_1		E_2	
Participation Year	Grade	Program	Achievement Study (A)	Retention Study (R)	A	R	A	R	A	R	A	R	A	R
1	7	BSP	4	4	102	93					102	94		
1	7	UMMaP	3	3	74	69	81	67			80	74		
1	7	SMSG	4	4	99	91	80	89			109	95		
1	8	BSP	10	10	271	260	280	246			276	250		
1	9	BSP	10	8	248	181	247	174			259	183		
1	9	UICSM	7	7	172	171	148	138			182	169		
1	9	SMSG	8	7	211	183	195	151			193	157		
1	10	BSP	8	8	222	204	195	164			205	178		
1	10	SMSG	10	8	224	178	243	160			244	176		
1	11	BSP	6	6	161	145	104	76			114	91		
1	11	SMSG	6	-	151	-	133	-			126	-		
1	12	SMSG	2	-	38	-	50	-			52	-		
2	7	SMSG	2	-			54	-	54	-			56	-
2	8	BSP	6	5			163	120	172	133			166	139
2	9	BSP	4	4			78	67	80	77			75	69
2	9	UICSM	6	4			129	89	149	88			142	92
2	9	SMSG	4	2			97	41	77	34			101	53
2	10	BSP	2	2			58	53	59	49			59	46
2	10	SMSG	6	4			133	74	142	92			143	81
2	11	BSP	4	3			70	50	78	58			73	56
2	11	SMSG	4	2			86	36	87	34			76	25
2	12	SMSG	2	-			50	-	42	-			43	-

Table 4

Number of Classes Selected for Analysis

xperimental Situation			Total Number of Participating Classes	Identification of Eliminated Classes	Number of Classes Analyzed
Year	Grade	Program			
1	7	BSP	4	None	4
1	7	UMMaP	4	455	3
1	7	SMSG	7	607, 631, 635	4
1	8	BSP	11	586	10
1	9	BSP	11	608*	10
1	9	UICSM	8	737	7
1	9	SMSG	14	265, 327, 331, 377, 677, 648	8
1	10	BSP	10	588, 649*	8
1	10	SMSG	18	279, 573**	16
1	11	BSP	7	659	6
1	11	SMSG	7	293	6
1	12	SMSG	2	None	2
2	7	SMSG	2	None	2
2	8	BSP	6	None	6
2	9	BSP	4	None	4
2	9	UICSM	6	None	6
2	9	SMSG	4	None	4
2	10	BSP	2	None	2
2	10	SMSG	6	None	6
2	11	BSP	4	None	4
2	11	SMSG	4	None	4
2	12	SMSG	2	None	2

* Due to computer difficulties present at the time of analysis of these experimental situations one randomly selected teacher was dropped.

** Due to programming limitations present at the time of analysis the experimental situation was divided into one group of 10 classes and one of 6 classes.

VI. Experimental Results

A. Comparisons Between Experimental and Conventional Curricula

This section presents a summary and interpretation of the statistical analysis carried out on the STEP data. Complete details are available in the Supplementary Reports listed in Appendix A.

A three-way analysis-of-variance was used for each experimental situation indicated in Table 3. The specific tests made were dictated by the expected mean squares given in Table 5 that follow from the adopted mathematical model discussed above for p teachers, q curricula and r ability levels. The harmonic mean is denoted as $\bar{n}$.

Table 5

Expected Mean Squares for Curricula Comparisons

Source of Variation	Expected Mean Square
Teacher (A)	$\sigma_\varepsilon^2 + \bar{n}qr\,\sigma_A^2$
Curriculum (B)	$\sigma_\varepsilon^2 + \bar{n}r\,\sigma_{AB}^2 + \frac{\bar{n}pr\sum_j \beta_j^2}{q-1}$
Pupil Ability (C)	$\sigma_\varepsilon^2 + \bar{n}q\,\sigma_{AC}^2 + \frac{\bar{n}pq\sum_k \gamma_k^2}{r-1}$
Teacher x Curriculum (AB)	$\sigma_\varepsilon^2 + \bar{n}r\,\sigma_{AB}^2$
Teacher x Pupil Ability (AC)	$\sigma_\varepsilon^2 + \bar{n}q\,\sigma_{AC}^2$
Curriculum x Pupil Ability (BC)	$\sigma_\varepsilon^2 + \bar{n}\,\sigma_{ABC}^2 + \frac{\bar{n}p\sum\sum_{jk} \phi_{jk}^2}{(q-1)(r-1)}$
Teacher x Curriculum x Pupil Ability (ABC)	$\sigma_\varepsilon^2 + \bar{n}\,\sigma_{ABC}^2$
Experimental Error	σ_ε^2

Table 6 summarizes the analysis-of-variance results for both the achievement and the retention data. It is seen that there are four significant experimental situations, all of which involve the teacher's first experience with the experimental materials. These significant results are the ninth grade SMSG achievement and retention studies, the tenth grade SMSG achievement study and the eleventh grade BSP study. Therefore these experimental situations were analyzed in greater detail.

Table 6

A Summary of F Test Comparisons of Achievement and Retention STEP Scores Between Experimental and Conventional Curricula

Experimental Situation			Achievement: Significance Level of F Test			Retention: Significance Level of F Test		
Participation Year	Grade	Program	Not Significant	1%	5%	Not Significant	1%	5%
1	7	BSP	*			*		
1	7	UMMaP	*			*		
1	7	SMSG	*			*		
1	8	BSP	*			*		
1	9	BSP	*			*		
1	9	UICSM	*			*		
1	9	SMSG		*				*
1	10	BSP	*			*		
1	10	SMSG			*	*		
1	11	BSP	*					*
1	11	SMSG	*			*		
1	12	SMSG	*			No Data		
2	7	SMSG	*			No Data		
2	8	BSP	*			*		
2	9	BSP	*			*		
2	9	UICSM	*			*		
2	9	SMSG	*			*		
2	10	BSP	*			*		
2	10	SMSG	*			*		
2	11	BSP	*			*		
2	11	SMSG	*			*		
2	12	SMSG	*			No Data		

An interpretation of the significant F ratios of Table 6 follows:

Experimental Situation - First Year of Teacher Participation - Ninth Grade - SMSG

The F ratios were significant both for the achievement and the retention test data. Furthermore, the highest mean was obtained from the experimental class. Therefore a more complete inspection of the data was made. Complete computational details may be seen in Supplementary Reports 100-D and 100-E.

a. Achievement Data Analysis - An analysis-of-variance table, a tabular presentation of means, diagrams of significant interactions and a Newman-Keuls analysis of mean differences were used in interpreting the results. These results are summarized in Tables 7, 8, 9 and 10 together with Fig. 2.

The teacher x ability interaction indicates how the relative achievement of higher and lower ability pupils varies among different teachers.

As seen in Fig. 2 the significant teacher x ability level interaction is most apparent for Teacher 687. This teacher ranked highest with higher ability pupils and lowest with lower ability pupils.

Table 7

Analysis-of-Variance for Achievement Data Following a Teacher's First Experience with the Ninth Grade SMSG Materials (1-9-SMSG)

Source of Variation	S.S.	d.f.	M.S.	F Test
Teacher	289.45	7	41.35	
Curriculum	425.19	2	212.60	*
Ability Level	6,874.03	1	6,874.03	*
Teacher x Curriculum	275.14	14	19.65	
Teacher x Ability Level	787.17	7	112.45	*
Curriculum x Ability Level	.44	2	.22	
Teacher x Curriculum x Ability Level	510.69	14	36.48	
Residual	12,956.27	549	23.60	

* Significant at .01 level

Table 8

Summary of Curriculum Means for the 1-9-SMSG Achievement Data

Curriculum	Pupils Below Pre-Test Median	Pupils Above Pre-Test Median	Unweighted Class Mean
Conventional	26.21	33.46	29.84
Control	25.54	32.74	29.14
Experimental	27.73	34.84	31.29

Table 9

Summary of Teacher x Ability Level Means
For the 1-9-SMSG Achievement Data

Teacher	Pupils Below Pre-Test Median	Pupils Above Pre-Test Median
255	25.54	32.66
326	28.02	31.67
359	26.18	34.42
648	25.98	33.16
667	26.34	34.33
685	28.71	33.98
687	24.72	37.13
703	26.35	32.09

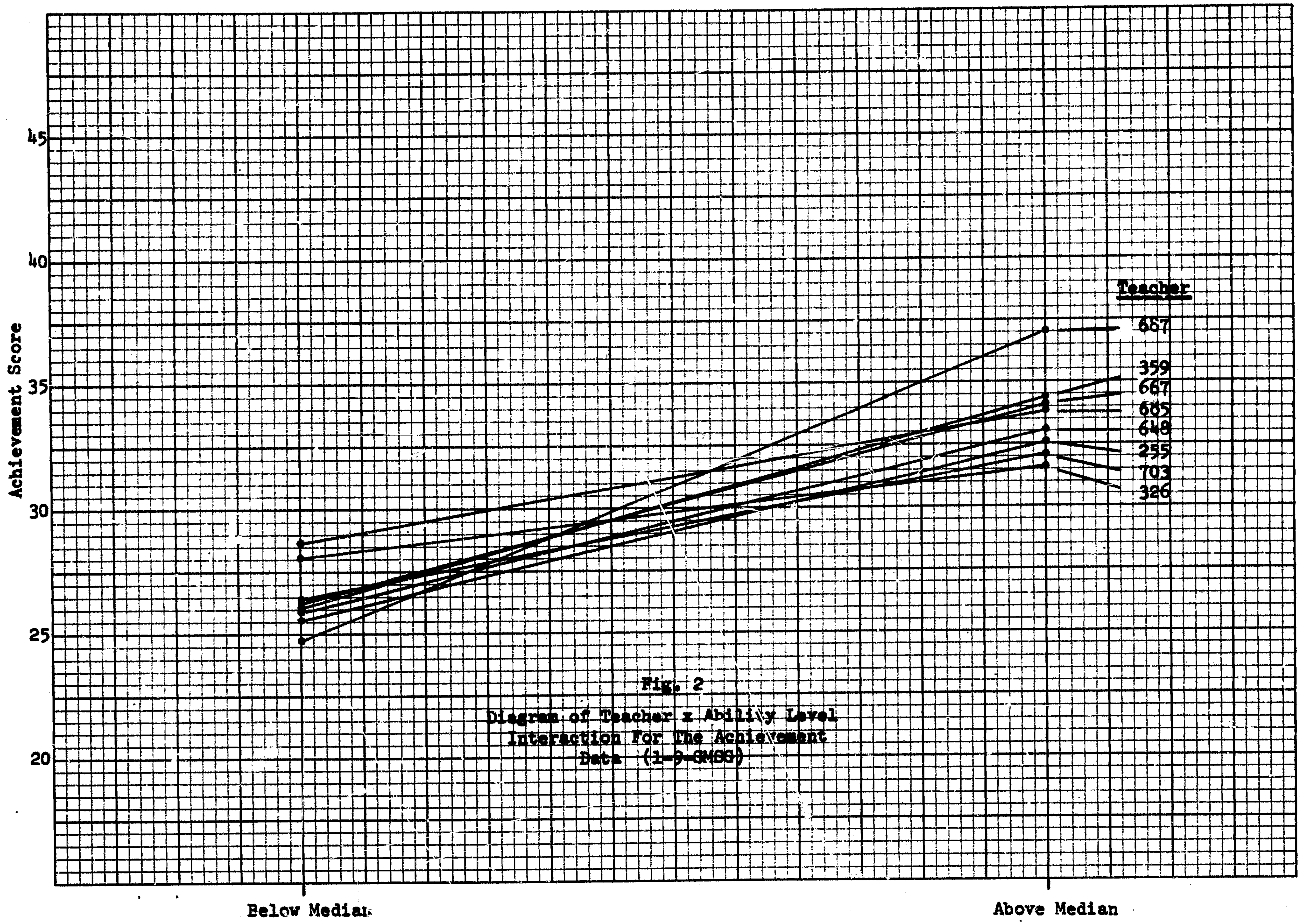

Fig. 2

Diagram of Teacher x Ability Level Interaction For The Achievement Data (1-9-SMSG)

The Newman-Keuls method of analysis was followed in interpreting individual mean differences between the curricula. This analysis is outlined in Table 10.

Table 10

Newman-Keuls Analysis of the (1-9-SMSG) Achievement Data

		$\bar{C}_2$	$\bar{C}_1$	$\bar{E}_1$
Ordered Means		29.14	29.84	31.29
Mean Differences	$\bar{C}_2$	-	.70	2.15
	$\bar{C}_1$		-	1.45

$$S_{\bar{x}} = \sqrt{MS_{TXC} / \bar{n} \cdot 16} = .332$$

		$\bar{C}_2$	$\bar{C}_1$	$\bar{E}_1$
Critical Value (.99)			1.40	1.62
	$\bar{C}_2$		-	*
	$\bar{C}_1$			-
Critical Value (.95)			1.01	1.23
	$\bar{C}_2$		-	*
	$\bar{C}_1$			*

On the basis of the Newman-Keuls analysis, it was concluded that the experimental curriculum showed some superiority over the conventional curriculum.

b. Retention Data Analysis - The same detailed information was prepared for the retention data as for the achievement data described above. The results are summarized in Tables 11, 12, 13 and 14 as well as Fig. 3.

Table 11

Analysis-of-Variance for the (1-9-SMSG) Retention Data

Source of Variation	S.S.	d.f.	M.S.	F Test
Teacher	975.09	6	162.52	*
Curriculum	391.53	2	195.77	**
Ability Level	8,568.41	1	8,568.41	*
Teacher x Curriculum	535.40	12	44.62	
Teacher x Ability Level	494.99	6	82.50	*
Curriculum x Ability Level	3.75	2	1.88	
Teacher x Curriculum x Ability Level	572.67	12	47.72	
Residual	12,604.70	449	28.07	

* Significant at .01 level
** Significant at .05 level

Table 12

Summary of Curriculum Means for the (1-9-SMSG) Retention Data

Curriculum	Pupils Below Pre-Test Median	Pupils Above Pre-Test Median	Unweighted Class Mean
Conventional	24.42	33.70	29.06
Control	24.03	32.99	28.51
Experimental	26.39	35.24	30.81

Table 13

Summary of Teacher x Ability Level Means for the (1-9-SMSG) Retention Data

Teacher	Pupils Below Pre-Test Median	Pupils Above Pre-Test Median
255	25.78	32.90
326	23.09	31.89
377	23.63	32.52
648	24.89	32.01
667	24.30	35.17
685	27.64	34.78
687	25.30	38.56

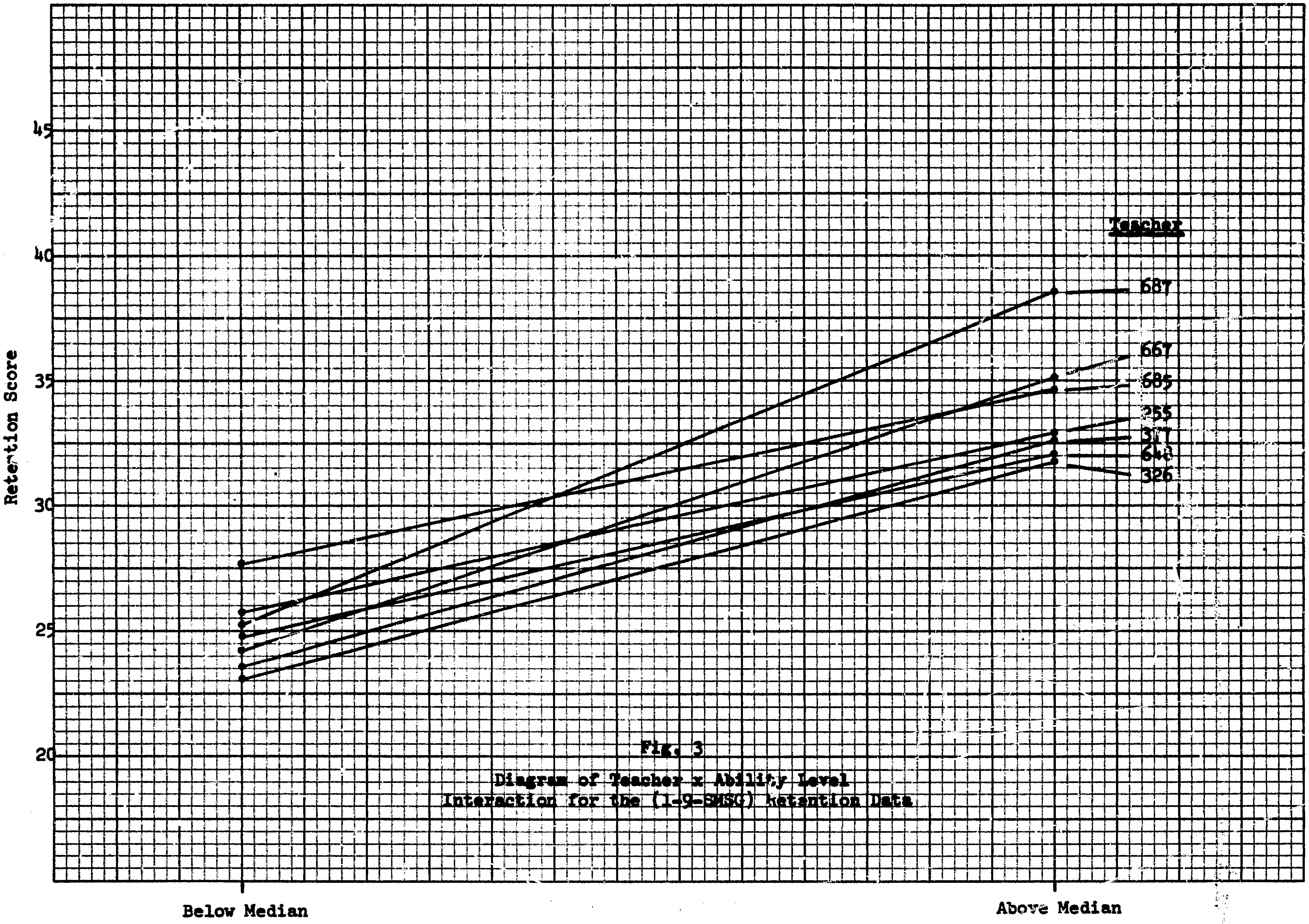

Fig. 3

Diagram of Teacher x Ability Level
Interaction for the (1-9-SMSG) Retention Data

Here again a significant teacher x ability effect was observed indicating a variation among teachers for the relative performances of the high and low ability pupils. Unlike the achievement analysis, no single teacher stood out in the interaction analysis for the retention data. Thus the interaction pattern was not consistent between achievement and retention testing and no specific generalizations concerning the nature of the interaction appeared warranted from the data.

The significant curriculum effect was analyzed by the Newman-Keuls method and is presented in Table 14.

Table 14

Newman-Keuls Analysis of the (1-9-SMSG) Retention Data

		$\bar{C}_2$	$\bar{C}_1$	$\bar{E}_1$
Ordered Means		28.51	29.06	30.81
Mean Differences	$\bar{C}_2$	-	.55	2.30
	$\bar{C}_1$		-	1.75

$$S_{\bar{x}} = \sqrt{MS_{TXC} / \bar{n} \cdot 14} = .511$$

		$\bar{C}_2$	$\bar{C}_1$	$\bar{E}_1$
Critical Value (.99)			2.20	2.57
	$\bar{C}_2$		-	-
	$\bar{C}_1$			-

		$\bar{C}_2$	$\bar{C}_1$	$\bar{E}_1$
Critical Value (.95)			1.57	1.92
	$\bar{C}_2$		-	*
	$\bar{C}_1$			-

On the basis of this analysis, it was concluded that the experimental curriculum did not continue to show superiority over the conventional curriculum. That is, the differences observed at the end of the year were not maintained over the intervening period.

However, this specific result is to be viewed with caution since the mean differences were small and one or two significant results could occur by chance in a series of separate analyses such as those carried out in this project.

Experimental Situation - First Year of Teacher Participation - Tenth Grade - SMSG

The F ratios were significant at the 1% level for the achievement study. The computational details may be seen in Reports 100-D and 100-E.

a. Achievement Data Analysis - The results for the achievement data are summarized in Tables 15, 16, together with the Newman-Keuls analysis of Table 17.

Table 15

Analysis-of-Variance for the (1-10-SMSG) Achievement Data

Source of Variation	S.S.	d.f.	M.S.	F Test
Teacher	488.09	9	54.23	**
Curriculum	264.36	2	132.18	**
Ability Level	8,586.22	1	8,586.22	*
Teacher x Curriculum	434.04	18	24.11	
Teacher x Ability Level	240.79	9	26.75	
Curriculum x Ability Level	20.83	2	10.41	
Teacher x Curriculum x Ability Level	345.74	18	19.21	
Residual	15,293.93	650	23.53	

* Significant at .01 level
** Significant at .05 level

Table 16

Summary of Curriculum Means for the (1-10-SMSG) Achievement Data

Curriculum	Pupils Below Pre-Test Median	Pupils Above Pre-Test Median	Unweighted Class Mean
Conventional	27.61	35.61	31.61
Control	28.59	35.70	32.14
Experimental	29.49	36.90	33.20

Table 17

Newman-Keuls Analysis of the (1-10-SMSG) Achievement Data

		$\bar{C}_1$	$\bar{C}_2$	$\bar{F}_1$
Ordered Means		31.61	32.14	33.20
Mean Differences	$\bar{C}_1$	-	.53	1.59
	$\bar{C}_2$		-	1.06

$$S_{\bar{x}} = \sqrt{MS_{TXC} / \bar{n} \cdot 20} = .344$$

Critical Value (.99)			1.38	1.60
		$\bar{C}_1$	$\bar{C}_2$	$\bar{E}_1$
	$\bar{C}_1$		-	-
	$\bar{C}_2$			-
Critical Value (.95)			1.01	1.23
		$\bar{C}_1$	$\bar{C}_2$	$\bar{E}_1$
	$\bar{C}_1$		-	*
	$\bar{C}_2$			-

Since, in this situation, the experimental class did not differ significantly from the control class, a further comparison was made. The conventional class average ($\bar{C}_1$) was compared with the combined experimental ($\bar{E}_1$) and control class ($\bar{C}_2$) averages. This was intended to provide an indication of "general" improvement of both the conventional and experimental classes. The general form utilized was as follows (5):

$$MS_{Comparison} = \frac{pr\bar{n}\ (\bar{C}_2 + \bar{E}_1 - 2\bar{C}_1)^2}{\sum_{i=1}^{3} C_i^2}$$

Where p = number of teachers

r = number of ability levels

$$\sum_{i=1}^{3} C_i^2 = 1^2 + 1^2 + 2^2 = 6$$

The following F ratio was computed:

$$F = \frac{MS_{Comparison}}{MS_{Curriculum \times Teacher}}$$

The a posteriori critical value was taken as $(q-1)\ F_{1-\alpha}\ [\ (q-1)\ ;\ (q-1)(p-1)\ ]$ where q = number of curricula. For this experimental situation $MS_{Comparison} = 152.07$, $F = 6.31$ and the 5% critical value = 7.10. Thus as a group, $\bar{C}_2$ and $\bar{E}_1$ were not significantly different from $\bar{C}_1$.

It was concluded that the significant F ratio could be explained by the difference between the experimental and conventional class. Little evidence was present to indicate a clear superiority of the experimental curriculum, however, since it did not differ significantly from the control class.

Experimental Situation - First Stage of Teacher Participation - Eleventh Grade - BSP

The F ratio was significant at the 5% level for the retention study. However the class means were:

$C_1 = 36.92$

$C_2 = 35.44$

$E_1 = 35.26$

Since the means for both the control and experimental classes were lower than that for the conventional class a comparison was made to see if there was a significant difference when comparing $\bar{C}_1$ against $\bar{C}_2$ and $\bar{E}_1$ combined. The comparison followed the same form as for experimental situation (1-10-SMSG). The results were as follows:

$MS_{Comparison}$ = 152.21, F = 6.31 and the 5% critical value = 8.20

Therefore the analysis was terminated with no evidence of experimental class superiority and also no significant evidence of a general (i.e. combined control and experimental classes) second year decline in retention scores.

From the point-of-view of practical significance, the lack of large curricular differences is evident in Table 18 which presents the largest mean differences observed in each of the experimental situations. The complete list of individual curriculum means is found in Supplementary Reports 100-C and 100-D.

Table 18

Summary of Sample Sizes, Largest Mean Differences and Error Variances for Comparisons Between Experimental and Conventional Curricula

Experimental Situation			ACHIEVEMENT							RETENTION						
			N	N	Class Means			Largest Mean	Error	N	N	Class Means			Largest Mean	Error
Yr.	Gr.	Pr.	Tchrs.	Pupils	CONV.	CONT.	EXP.	Diff.[1]	Var.	Tchr.	Pupils	CONV.	CONT.	EXP.	Diff.[1]	Var.
1	7	BSP	4	228	30.30	29.20	29.46	1.10	32.01	4	279	30.00	29.29	29.77	.72	28.87
1	7	UMMaP	3	235	31.96	33.77	31.74	2.03	34.85	3	210	30.03	30.86	28.80	2.06	31.86
1	7	SMSG	4	288	32.64	30.71	31.34	1.93	33.84	4	275	29.60	28.24	28.57	1.36	33.49
1	8	BSP	10	827	33.09	34.64	33.42	1.55	31.43	10	756	32.24	33.06	32.40	.82	27.52
1	9	BSP	10	754	29.00	27.92	28.11	1.08	27.81	8	538	28.15	27.64	28.41	.77	29.43
1	9	UICSM	7	502	27.74	26.64	27.51	1.10	29.58	7	478	26.95	26.45	26.77	.50	32.77
1	9	SMSG	8	599	29.84	29.14	31.29	2.15	23.59	7	491	29.06	28.51	30.81	2.30	28.08
1	10	BSP	8	622	29.32	29.21	30.15	.94	29.97	8	546	30.54	29.91	30.48	.64	27.50
1	10	SMSG	10	711	31.61	32.14	33.20	1.63	23.75	8	514	31.66	31.70	33.13	1.47	27.85
1	11	BSP	6	379	35.90	36.56	36.69	.79	20.15	6	312	36.92	35.44	35.26	1.66	25.98
1	11	SMSG	6	410	34.79	34.47	33.90	.89	25.17	6	369	36.44	35.99	34.44	2.00	
1	12	SMSG	2	140	30.45	29.43	30.99	1.56	21.74	No Data						
2	7	SMSG	2	164	30.95	27.44	31.45	4.01	37.90	No Data						
2	8	BSP	6	501	36.26	35.48	36.35	.87	27.05	5	392	34.73	33.43	35.04	1.61	25.71
2	9	BSP	4	233	25.84	25.57	27.28	1.71	30.55	4	213	26.78	24.73	27.16	2.43	38.04
2	9	UICSM	6	420	25.87	26.46	26.26	.59	33.34	4	269	27.70	25.42	25.86	2.28	41.35
2	9	SMSG	4	275	28.76	27.13	28.67	1.63	27.20	2	128	28.19	27.00	29.58	2.58	26.89
2	10	BSP	2	176	28.97	29.94	30.08	1.11	29.73	2	148	28.64	28.35	28.42	.29	36.19
2	10	SMSG	6	418	31.69	32.50	31.86	.81	27.34	4	247	31.02	32.92	32.67	1.90	29.62
2	11	BSP	4	221	35.52	34.54	36.02	1.48	28.53	3	164	36.12	35.93	38.03	2.10	22.08
2	11	SMSG	4	249	34.11	34.94	36.41	2.30	30.63	2	95	35.74	38.10	37.70	2.36	32.40
2	12	SMSG	2	135	29.76	33.20	33.40	3.64	22.55	No Data						

[1] Largest Absolute Mean Differences Among the Three Curricula

In an attempt to provide further evidence of the practical significance of the findings, rough estimates of the relative contribution of each experimental factor to the total variance were made for the significant results of the achievement study. See Hays (2). For teacher considered as a random factor and curriculum and pupil ability as fixed factors the estimates were as follows:

$$\text{Total variance, } \sigma_y^2 = E[Y_{ijkl} - \mu]^2$$

$$\sigma_y^2 = \sigma_A^2 + \frac{\sum_j \beta_j^2}{q} + \frac{\sum_k \gamma_k^2}{r} + \sigma_{AB}^2 + \sigma_{AC}^2 + \frac{\sum\sum_{jk} \phi_{jk}^2}{qr} + \sigma_{ABC}^2 + \sigma\varepsilon^2$$

The estimates of each of these terms are, from Table 5.

$$\sigma_A^2 = \frac{MS_A - MS_{Error}}{\bar{n}qr}$$

$$\frac{\sum_j \beta_j^2}{q} = \left[\frac{MS_B - MS_{AB}}{\bar{n}pr}\right] \frac{q-1}{q}$$

$$\frac{\sum_k \gamma_k^2}{r} = \left[\frac{MS_C - MS_{AC}}{\bar{n}pq}\right] \frac{r-1}{r}$$

$$\sigma_{AB}^2 = \frac{MS_{AB} - MS_{Error}}{\bar{n}r}$$

$$\sigma_{AC}^2 = \frac{MS_{AC} - MS_{Error}}{\bar{n}q}$$

$$\frac{\sum\sum_{jk} \phi_{jk}^2}{qr} = \left[\frac{MS_{BC} - MS_{ABC}}{\bar{n}p}\right] \frac{(q-1)\ (r-1)}{qr}$$

$$\sigma_{ABC}^2 = \frac{MS_{ABC} - MS_{Error}}{\bar{n}}$$

$$\sigma_\varepsilon^2 = MS_{Error}$$

Two significant results of interest are for the teacher's first experience with ninth and tenth grade SMSG materials. The contribution of each factor to the total variance is estimated in Table 19 for these two situations.

Table 19

Percentage of Variance Accounted for by Each Experimental Factor in Experimental Situation (1-9-SMSG) and (1-10-SMSG)

	Percentage of Variance Accounted For in the Experimental Situation	
Source of Variation	(1-9-SMSG)	(1-10-SMSG)
Teacher	.65	1.30
Curriculum	1.77	.93
Ability Level	30.87	36.41
Teacher x Curriculum	0	.07
Teacher x Ability Level	6.49	.27
Curriculum x Ability Level	0	0
Teacher x Curriculum x Ability Level	2.82	0
Residual	57.37	61.01

In summary, it is apparent that not only are the absolute magnitudes of the curriculum differences small but also the curriculum differences relative to pupil ability and the error term are unimpressive.

B. Comparisons Among Experimental Curricula

Direct comparisons were made among experimental curricula. This was done by examining each grade level where more than one experimental program was taught. A single pre-test median was computed for all pupils in these experimental programs. The data were then analyzed under a three-way analysis-of-variance model with teachers nested within curricula.

Table 20 summarizes these results. Complete details are given in Report 100-F.

Table 20

Summary of Comparisons Among Experimental Curricula

Experimental Situation	No. of Exp. Classes	Curriculum Comparisons: Not Significant	5%	1%
(1-7-BSP) vs (1-7-UMMaP) vs (1-7-SMSG)	9	*		
(1-9-BSP) vs (1-9-UICSM) vs (1-9-SMSG)	21		*	
(1-10-BSP) vs (1-10-SMSG)	16	*		
(1-11-BSP) vs (1-11-SMSG)	12	*		
(2-9-BSP) vs (2-9-UICSM) vs (2-9-SMSG)	12	*		
(2-10-BSP) vs (2-10-SMSG)	4	*		
(2-11-BSP) vs (2-11-SMSG)	8	*		

From Table 22 it is seen that the only significant curriculum differences appearedin the ninth grade in the comparison of BSP, UICSM and SMSG programs following a teacher's initial experience with these programs. The means were BSP = 27.92, UICSM = 28.04 and SMSG = 30.62.

Of course, extreme care must be taken in interpreting this particular result since no control was made on the teachers and they are completely nested within the curricula. Also, this result is not independent of the previous significant result involving SMSG since both analyses utilized the same classes.

C. Effects of Increased Teacher Experience with Experimental Programs on Achievement Trends for Conventional Classes

A study of teacher trends of pupil achievement scores was made for continued teacher experience with experimental programs over a two year period. This section presents the results for teacher trends in the conventional curricula.

Table 21 summarizes the curricular results of a three-way analysis-of-variance for conventional curricula. A single pre-test median was computed for each experimental situation. Complete details are found in Supplementary Report 100-H.

Table 21

Summary of Teacher Trend Analysis For Pupil Achievement Scores in Conventional Curricula

Experimental Situation				Teacher Means for Conventional Achievement Scores					
				Year of Participation			Curricular Results		
Yr.	Gr.	Pr.	No. of Tchrs.	1961-62	1962-63	1963-64	Not Significant	5%	1%
1,2,3	7	SMSG	2	33.83	31.51	28.50	*		
1,2,3	8	BSP	6	34.12	35.88	35.27	*		
1,2,3	9	BSP	4	27.35	25.84	25.57	*		
1,2,3	9	UICSM	6	28.36	26.68	27.42	*		
1,2,3	9	SMSG	4	30.15	29.38	28.54	*		
1,2,3	10	BSP	2	30.89	28.97	29.94	*		
1,2,3	10	SMSG	6	30.58	31.78	32.53	*		
1,2,3	11	BSP	4	34.94	35.48	34.54	*		
1,2,3	11	SMSG	4	34.89	34.11	34.94	*		
1,2,3	12	SMSG	2	31.08	29.76	33.20	*		

It was concluded that, as far as the achievement scores reflected the teaching of conventional curricula, teachers did not show a significantly increasing trend with additional experience.

D. Effects of Increased Teacher Experience with Experimental Programs on Achievement Trends for Experimental Classes

This section presents the teacher trend analysis for the experimental curricula. Table 22 summarizes the analysis-of-variance results. Complete details are presented in Supplementary Report 100-G.

Table 22

Summary of Teacher Trend Analysis for Pupil Achievement Scores in Experimental Curricula

Experimental Situation			No. of	Teacher Means For Experimental Achievement Scores — Year Of Participation		Curricular Results		
Yr.	Gr.	Pr.	Tchrs.	1962-63	1963-64	Not Significant	5%	1%
1,2	7	SMSG	2	34.07	32.34	*		
1,2	8	BSP	6	34.92	36.04	*		
1,2	9	BSP	4	25.88	27.28	*		
1,2	9	UICSM	6	27.68	27.60	*		
1,2	9	SMSG	4	31.48	29.44			*
1,2	10	BSP	2	31.38	30.08	*		
1,2	10	SMSG	6	31.91	31.86	*		
1,2	11	BSP	4	35.49	36.02	*		
1,2	11	SMSG	4	34.62	36.41	*		
1,2	12	SMSG	2	30.53	33.40	*		

It is seen from Table 22 that the only significant result was for experimental situation (1st and 2nd - 9-SMSG) and that this was due to a drop in the class means for the second year of teaching the experimental class. Conventional class averages also dropped during this time. However the experimental class (E_1) was significantly higher than the conventional class (C_1) hence little evidence is present for inferring a general decline in pupil achievement as a result of teacher participation in the experimental program.

A brief study of the pupil sex was also undertaken and the results are summarized in Table 23.

Table 23

Class Averages for STEP Achievement Scores
Classified By Curriculum and Sex

Experimental Situation			C_1		C_2		E_1	
Yr.	Gr.	Pr.	Male	Female	Male	Female	Male	Female
1	7	BSP	31.75	29.19	28.44	27.36	29.96	28.59
1	7	UMMaP	32.07	30.46	34.59	32.53	29.56	31.43
1	7	SMSG	32.72	30.49	29.75	27.52	30.18	30.42
1	8	BSP	33.20	31.76	34.86	33.25	33.33	32.37
1	9	BSP	30.31	27.22	29.39	26.71	29.21	27.58
1	9	UICSM	29.88	27.00	26.26	22.11	27.57	26.35
1	9	SMSG	29.67	29.68	30.78	27.52	32.62	29.57
1	10	BSP	32.41	29.41	30.77	29.15	31.78	29.92
1	10	SMSG	32.03	29.33	32.13	30.75	34.90	31.37
1	11	BSP	35.63	33.83	36.26	33.45	37.79	35.02

It is seen in Table 23 that the boys generally were higher than the girls on achievement scores. However an analysis requiring a distinction to be made between sexes would greatly reduce the number of degrees of freedom within each cell and in some cases eliminate a class entirely from the analysis. Since this further subdivision would not necessarily result in a greater precision for the experimental comparisons, sex was not treated as a separate factor in the experimental design.

VII. Discussion

The comparisons between pupils instructed with the conventional programs and those instructed with each of the experimental programs, within the separate grades, revealed relatively few statistically reliable differences with respect to mathematics achievement and retention. Among the differences observed, four patterns of average class differences are of interest:

1. Experimental class greater than either the control or conventional classes -

   This pattern was observed 20 times with three of the results being statistically significant. The clearest difference observed was for the ninth grade SMSG program where the teachers had no previous experience with the experimental program. Pupils in the experimental program performed significantly better on both the achievement ($P<.01$) and retention ($P<.05$) measures than pupils in either the same or previous year conventional classes. It appears that instruction in ninth grade algebra with the SMSG program had a measurable effect upon pupil performance. Nonetheless it also appears that this effect was not maintained for the subsequent year of the teachers' experience with this program, at least for the teachers who continued in the project with this program.

   For the tenth grade SMSG geometry program comparison, the experimental class pupils performed significantly better than the previous year conventional class pupils on measures of achievement ($P<.05$). These pupils did not, however, differ reliably from those in the control class instructed with conventional materials during the same year and the latter did not differ from those in the previous year conventional classes, their average achievement being between that of the other two treatment groups. Among other reasons, this pattern of differences might have been a result of an experimental instruction effect which in turn facilitated, to a lesser extent, the instruction in the control class possibly by sensitizing a teacher to certain instructional limitations of his more routine conventional class presentation. Here, again, none of the differences were manifested for the teachers second year of experience with the program.

2. Experimental class less than either the control or the conventional classes -

   This pattern was observed five times with one result being statistically significant. The pre-experimental year, conventional class pupils scored significantly higher ($P<.05$) on the measure of retention than the experimental class pupils for the teacher's first year with the new program. The conventionally instructed same-year control class pupils had an average score between the other two instructional conditions and did not differ reliably from either. This difference occurred however only for the retention test scores, not for the achievement test scores. A difference of this nature - a drop in performance for the experimental class relative to the previous year conventional class would, in general, suggest two possibilities: that the experimental program is generally less effective than the control program, or that the teacher's instructional effectiveness is somewhat restricted by having to become familiar with a new program. The fact that the second year conventional class also showed a drop in performance (although not significant) is more in keeping with the latter possibility, as is the fact that these differences were not observed during the teacher's second year of experience with the program. Since this difference was observed only for retention scores and did not reach a very high level of reliability, it can only be considered suggestive of a possible difference rather than as a basis for drawing any firm conclusion concerning the differential effectiveness of this experimental program.

3. Conventional class greater than either the experimental or control classes -

   This pattern of differences was observed 14 times with one of the results being statistically significant. This result is the same as discussed under pattern 2. As pointed out, this result was perhaps due to the teacher's effectiveness being reduced by the additional burden of teaching the experimental class. In any event, it is not an effect associated with the experimental program alone.

4. Conventional class less than either the experimental or control classes -

   This pattern was observed 12 times with none of the results being statistically significant.

In addition certain other findings were evident:

No significant increasing trends of pupil achievement scores were found for continued teacher experience with experimental programs over a two or three year period. This finding held for trends both in conventional and experimental curricula.

Significant teacher differences consistently occurred for all curricula. This finding held for both pupil achievement and pupil retention scores.

Initial pupil ability was, overwhelmingly, the most significant factor involved in either the achievement or the retention studies.

The relative magnitude of the ability factor emphasizes the need for participating teachers in curriculum field studies to make a clear distinction between random and haphazard placement of pupils into the contrasted curricula.

Male pupils generally scored higher than female pupils in all mathematics classes. However a pilot analysis indicated that the error term would not be significantly reduced by separating the pupils by sex and hence males and females were pooled.

Three points should be noted concerning the instructional differences that have been observed above:

1. These differences may have been due to factors other than differences associated with the instructional conditions. Although the comparisons were made controlling for differences in pupils' initial ability, there may have been systematic or chance differences in pupil motivation and attitude or in teacher attitudes which could possibly have contributed to the observed effects.

2. As was evident in Table 19 even the largest and most reliable instructional program differences were relatively quite small in proportion to the differences associated with other variables in the model, particularly initial pupil ability.

   The instructional program differences that were observed appear to indicate primarily that the analysis was sufficiently sensitive to detect quite small but reliable performance differences. The actual differences are probably not large enough to be in themselves a sufficient basis for instructional program decisions.

3. Even though few differences were observed, the clearest and most reliable differences were in favor of the E rather than the C programs. As stated above, three of the four significant results were of this form.

One effect that was not observed in the data that is somewhat puzzling was the occurrence of any continued improvement in pupil performance after teachers had had more experience with the E programs. In the two instances where the clearest differences favoring the E programs were observed (ninth and tenth grade SMSG) this effect was observed only for the teacher's first year with the E program but not for the subsequent years.

There are several factors that singly or in combination could account for the lack of any clearcut instructional program differences even though the E programs may have had some definite effect on pupil learning. Of primary concern in the fact that the measure of achievement (STEP) may in itself not have been a very sensitive measure of either within grade instructional effects in general and/or of the instructional outcomes specific to the E programs. It appears that this test might be more capable of reflecting gross changes in mathematics proficiency that occur over a several year period and consequently is not very sensitive to single year changes in proficiency. It may also be reflecting to a greater extent than would be necessary a general mathematics ability rather than acquired proficiency. This suggests using a more sensitive instrument to measure achievement gains.

Another significant factor is the lack of control over or knowledge of the teacher's actual instructional input and consequently the degree of instructional difference that really existed for pupils in the E and C classes. Provision of a new text in itself and even the teacher's intent to teach a "new" or "different" program may not be sufficient to bring about an instructional difference of any magnitude for a certain number of teachers. On the other hand, some teachers who did become familiar with or adopt the approach and content appropriate for the E text and teacher manual may have inadvertently "carried-over" some of the instructional change to the C classes. Of course, if the latter had occurred to any great extent, it would have been as readily detected as would an E program effect because of the comparison with the performance of the previous year conventional class.

The question of the degree to which instructional differences may have existed between E and C classes is in itself part of a more general question of the extent to which the overall instructional input or, more exactly, the actual knowledge pupils acquire is a function of the text materials used as it is of the teacher's presentation. There is quite likely variation among teachers as to the amount the pupil needs to rely on the teacher or the textbook to obtain whatever is learned. It may well be that for a large proportion of classes a different textbook does not alter to any extent the instructional content the pupils actually do learn, i.e. that the latter is much more a function of the teacher than of the text.

VIII. References

1. Buros, O. Sixth Yearbook of Mental Measurement. New Brunswick, N.J.: Rutgers University Press, 1966.

2. Hays, W. L. Statistics for Psychologists. New York: Holt, Rinehart and Winston Inc., 1963.

3. Kraft, C. H. Report of Investigations of the Cumulative Effects of Teacher and Student Experience with SMSG Texts, A Report by the Minnesota National Laboratory, Minnesota State Department of Education, 1962.

4. Sequential Tests of Educational Progress (STEP) Test Manual. Princeton, N.J.: Educational Testing Service, 1957.

5. Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.

Appendix A

Supplementary Report List

1. Frequency Diagrams of Pre and Post Test STEP Scores for each Teacher. Report 100-A

2. Scatterdiagrams of Pre and Post Test STEP Scores for all Teachers Within a Given Experimental Situation. Report 100-B

3. Weighted Means Analysis of Achievement Scores. Report 100-C

4. Weighted Means Analysis of Retention Scores. Report 100-D

5. Unweighted Means Analysis of Achievement Scores. Report 100-E

6. Unweighted Means Analysis of Comparisons Among Experimental Curricula. Report 100-F

7. Unweighted Means Analysis of Teacher Trends of Pupil Achievement in Experimental Curricula. Report 100-G

8. Unweighted Means Analysis of Teacher Trends of Pupil Achievement in Conventional Curricula. Report 100-H

9. Analysis of Covariance of Achievement Scores. Report 100-I

Appendix B (Corrected - this is to replace Appendix B as originally printed which was incomplete.)

List of the Instructional Materials Used in the Experimental Classes

Author: School Mathematics Study Group (SMSG), Publisher: Yale University Press.

Grade 7	Math for Junior High School - Volume I, 1962
Grade 8	Math for Junior High School - Volume II, 1962[1]
Grade 9	First Course in Algebra, 1962
Grade 10	Geometry, and Geometry With Coordinates, 1962
Grade 11	Int[illegible]mediate Mathematics, 1962
Grade 12	Elementary Functions and Introduction to Matrix Algebra, 1962

Author: Brumfiel-Eicholz-Shanks (Ball State), Publisher: Addison-Wesley.

Grade 7	Arithmetic, Concepts and Skills, 1963
Grade 8	Introduction to Mathematics, 1961
Grade 9	Algebra I, 1961
Grade 10	Geometry, 1960
Grade 11	Algebra II, 1962

Author: University of Maryland Mathematics Project (UMMaP), Publisher: University of Maryland.

Grade 7	Math for Junior High School - 1st Book, 1959
Grade 8	Math for Senior High School - 2nd Book, 1961[1]

Author: University of Illinois Committee on School Mathematics (UICSM), Publisher: University of Illinois Press[2].

Grade 9	High School Mathematics Units 1-2-3-4, 1960
Grade 10	High School Mathematics Units 5 and 6, 1960[1]

[1] Although materials for these programs were available and distributed for use in a few classes, there was not a sufficient number of classes providing the necessary test data to be included in the analysis.

[2] Table 1 in the text incorrectly indicates that 11th and 12th grade UICSM materials were available and included in the study. Only 9th and 10th grade UICSM materials were available and included in the study.

Appendix B

List of the Instructional Materials Used in the Experimental Classes

	AUTHOR	TITLE	PUBLISHER	DATE
th Grade				
ARYLAND		Math for Jr. High School - 1st Book	Univ. of Maryland	1959
MSG	School Mathematics Study Group	Math for Jr. High School - Vol. I	Cushing-Malloy, Inc.	1962
th Grade				
ALL STATE	Brumfiel-Eicholz-Shanks	Introduction to Mathematics	Addison-Wesley	1961
th Grade				
ALL STATE	Brumfiel-Eicholz-Shanks	Algebra I	Addison-Wesley	1961
LLINOIS	University Illinois Committee on School Mathematics	High School Mathematics Units 1-2-3-4	Univ. Illinois Press	1960
MSG	School Mathematics Study Group	First Course in Algebra	Cushing-Malloy, Inc.	1962
0th Grade				
ALL STATE	Brumfiel-Eicholz-Shanks	Geometry	Addison-Wesley	1960
MSG	School Mathematics Study Group	Geometry	Cushing-Malloy, Inc.	1962
MSG	School Mathematics Study Group	Geometry With Coordinates	Cushing-Malloy, Inc.	1962
1th Grade				
ALL STATE	Brumfiel-Eicholz-Shanks	Algebra II	Addison-Wesley	1962
MSG	School Mathematics Study Group	Intermediate Mathematics	Cushing-Malloy, Inc.	1962
2th Grade				
MSG	School Mathematics Study Group	Elementary Functions & Introduction to Matrix Algebra	Cushing-Malloy, Inc.	1962