DOCUMENT RESUME

ED 034 686 24 SE 007 688

AUTHOR Gordon, C. Wayne
TITLE Preliminary Evaluation Report on the Los Angeles City Schools, SB 28 Demonstration Program in Mathematics.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-6-1646
PUB DATE Oct 69
NOTE 66p.

EDRS PRICE EDRS Price MF-$0.50 HC-$3.40
DESCRIPTORS *Educational Research, *Experimental Programs, Grade 7, Grade 8, Instruction, Instructional Materials, Material Development, *Mathematics, *Program Evaluation

ABSTRACT

The purpose of this preliminary report is to describe and evaluate the Los Angeles Model Mathematics Project (LAMMP). The objectives of this project include the improvement of mathematical skills and understanding of mathematical concepts, the improvement of students' self-image, the development of instructional materials and the assessment of commercially developed instructional materials. The report describes the manner by which participants in the project were chosen and the instructional programs which were employed. An evaluation of measurement instruments used, a summary and interpretation of evaluation data, and recommendations for improving the project are also included. Only a preliminary analysis of data is contained in this report and no attempt is made to make statistical inferences. The report concludes that in many cases subjects were not randomly assigned to experimental and comparison groups and thus no valid inferences about the effects of instructional variables based on differences between experimental and comparison groups can be made. (FL)

# PRELIMINARY EVALUATION REPORT ON THE LOS ANGELES CITY SCHOOLS SB 28 DEMONSTRATION PROGRAM IN MATHEMATICS

C. Wayne Gordon

CSE Working Paper #1
October 1969



Center for the Study of Evaluation
UCLA Graduate School of Education

# PRELIMINARY EVALUATION REPORT ON THE LOS ANGELES CITY SCHOOL'S SB 28 DEMONSTRATION PROGRAM IN MATHEMATICS

C. Wayne Gordon

CSE Working Paper
October 1969

Center for the Study of Evaluation
UCLA Graduate School of Education

Table of Contents

List of Tables

## List of Figures

I. Objectives

The objectives of the Los Angeles Model Mathematics Project, hereafter referred to as LAMMP, were stated by the administration of the project (L. A. City Schools publication, April 12, 1967)

OBJECTIVES

To improve mathematical skills and understandings of mathematical concepts

To improve the pupils' self-image

To identify specific assets and limitations relating to the learning process

To develop and use special instructional materials and programs and to assess their values

To select and use appropriate commercially developed equipment, instructional materials, and programs and to assess their values.

II. Participants - Experimental

A. Characteristics

1. The participants were 7th and 8th grade students. The distribution of students by grade level is shown in Table 1.

Table 1

Grade Levels of Participants

| | Fall, 1967 | | | Spring, 1968 | | | |
|---|---|---|---|---|---|---|---|
| Belvedere | B7 | A7 | Total | B7 | A7 | B8 | Total |
| No. of classes | 6 | 2 | 8 | 0 | 6 | 2 | 8 |
| No. of students | 115 | 37 | 152 | 0 | 111 | 37 | 148 |
| Edison | | | | | | | |
| No. of classes | 6 | 0 | 6 | 2 | 4 | 0 | 6 |
| No. of students | 88 | 0 | 88 | 40 | 75 | 0 | 115 |
| Pacoima | | | | | | | |
| No. of classes | 6 | 2 | 8 | 0 | 6 | 2 | 8 |
| No. of students | 116 | 38 | 154 | 0 | 106 | 35 | 141 |

| | No. of students, all schools | | | |
|---|---|---|---|---|
| | B7 | A7 | B8 | Total |
| Fall, 1967 | 319 | 75 | -- | 394 |
| Spring, 1968 | 40 | 292 | 72 | 404 |

2. The distribution of participants by sex is shown in Table 2.

Table 2

Sex of Participants

| School | Fall, 1967 | | Spring, 1968 | |
|---|---|---|---|---|
| | % Girls | % Boys | % Girls | % Boys |
| Belvedere | 53 | 47 | 52 | 48 |
| Edison | 52 | 48 | 47 | 53 |
| Pacoima | 46 | 54 | 45 | 55 |

3. The distribution of participants by ethnic groups is shown in Table 3.

Table 3

Ethnic Groups Among Participants

| School | % Caucasian | % Negro | % Mexican-American | % Other |
|---|---|---|---|---|
| Belvedere | 2 | 2 | 95 | 1 |
| Edison | 0 | 95 | 5 | 0 |
| Pacoima | 58 | 2 | 39 | 1 |

4. An estimate of the ability levels of participants was available from intelligence quotients computed from a fifth grade administration of the California Test of Mental Maturity. Information from this test is summarized in Table 4.

Table 4

Mean I.Q. by School for Participants on the California Test of Mental Maturity (1957 Short Form)

| School | Verbal Mean | Non-Verbal Mean | Total Mean |
|---|---|---|---|
| Belvedere | 93.6 | 94.1 | 94.0 |
| Edison | 92.6 | 86.2 | 89.6 |
| Pacoima | 95.6 | 93.2 | 94.6 |

E. Procedures used to select students. The target group for the program was students of average ability who were achieving in mathematics a year or more below grade level. The procedures used to select students for the program were determined by the counselor consultant for the project and the head counselor in each of the three schools. The criteria used to identify eligible students were described by the counselor consultant as follows:

1. Used test information CTMM and CAT scores from pupils' 1965 A5 evaluations.
2. An effort was made to consider current mid-term mathematic grades.
3. Consideration was given to the fact that standardized group tests do not always reflect these children's full potential.
4. When current group IQ tests did not indicate average ability, scores from individual tests or earlier non-verbal tests were used.
5. Average intelligence was interpreted to be that which represents average intelligence for that school.
6. When unusual score patterns were found, counselor and teacher consultation was sought.

The procedures followed in selecting students for the program were reported as follows by the project counselor.

Edison Junior High School The B7 students identified as eligible were scheduled into classes of 40. The first day of school an experimental and a comparison teacher "randomly" divided each of these classes. The rate of attrition at this school resulted in small experimental and large comparison classes by the end of the first semester; new students were scheduled into the comparison classes.

At the beginning of the second semester, the original six experimental classes were "randomly" combined into four classes and two new B7 classes of 40 were formed and divided in the same manner as described above.

Pacoima Junior High School The students identified as eligible were scheduled into six B7 and two A7 experimental classes. Four comparison classes (three B7 and one A7) were formed from among other students considered to have "average" academic capability.

The same experimental students were "randomly" reassigned to experimental classes the second semester. A number of comparison students were assigned to "average" level math classes not designated as comparison classes and their place filled by other students considered to have average academic capability.

Belvedere Junior High School A6 students identified as eligible were randomly scheduled into experimental and comparison B7 classes by the head counselor. The A7 classes were formed by the counselor consultant, who used a table of random numbers to assign eligible students to experimental and comparison groups.

Aside from the few students who transferred during the year, experimental and comparison classes were maintained intact throughout the year.

C. Diagnosis of mathematics problems. Any diagnosis of students' difficulties in learning mathematics were carried out by each teacher for his own students. No systematic diagnosis procedures were followed by all project personnel.

Students' total scores from the Iowa Tests of Arithmetic and Vocabulary and the LAMMP Diagnostic Test, which were administered by the evaluation team in September, were made available to the teachers. Teachers used the test information as they wished in diagnosis of students' mathematics skills.

Although the project plans called for a counselor consultant to assist teachers in the diagnosis of individual students' needs, this assistance was not in fact available to the teachers until February 1968. The first counselor employed left the project before school opened in September 1967, and the position was not filled again until the beginning of the second semester.

III. Methods

A. Instructional methods. The focus for each of the three experimental centers was described as follows. (L. A. City publication, April 12, 1967)

Operational Objectives of the Programmed Learning Center at Edison Jr. High:

To select and use commercially developed instructional materials and programs, and to assess their values
To develop and use special instructional materials and assess their value

Operational Objectives of the Closed-Circuit Television Retrieval Center at Pacoima Jr. High:

To measure the changes in student and teacher behavior which occur due to availability and use of a retrieval system
To provide teachers and other school personnel an awareness of improvements of teaching, for introduction of new methodology and technology
To identify specific skills and to train teachers to utilize effectively information retrieval services
To provide a variety of supplementary instructional materials to students and teachers
To produce appropriate lessons on video tape not available commercially
To provide teachers with the skills needed to design appropriate individual study programs utilizing all instructional resources

Operational Objectives of the Mathematics Laboratory at Belvedere Jr. High:

To study the effects of a multi-media, multi-sensory environment on the mathematical achievement of disadvantaged pupils
To develop learning materials which are suitable for use with pupils in target areas

In general, the instructional program at each center was oriented to the focus described, except at Pacoima. There it was not possible to carry out this year a program based on a closed-circuit television retrieval system because the necessary equipment was not installed. The teachers at the Pacoima center

developed lessons to be used later in the television system. During the second semester some of these lessons were tested with one television camera and a videotape recorder.

Observations of classrooms by the evaluation team indicated that experimental classrooms were characterized by:

flexible classroom organization for instruction;

use of a wide variety of instructional techniques and materials;

emphasis on developing students' understanding of basic mathematical principles;

teacher efforts to develop positive concepts in students and positive attitudes toward school.

One objective of the program was to develop new instructional materials. Each teacher was originally assigned only two classes per day in order to allow time for work on materials. (At mid-year, however, the shift of one of the teachers from Pacoima to central office staff made it necessary to increase the load of two teachers there to 3 classes per day.) About half of the time of the illustrators who were employed by the project was devoted to helping teachers prepare materials.

A mathematics consultant was also assigned to the central project staff to work with teachers in all three centers. Observations over the school year, however, indicated that, in fact, the mathematics consultant spent at least half (often more) of his time in administrative duties which did not assist teachers in the preparation of instructional materials.

Although some materials were created, it is impossible to evaluate their effectiveness at this time. Samples of the materials developed and used during each semester were not submitted to the evaluation team until near the end of each semester. Therefore, it was not possible to carry out any kind of evaluation of the effectiveness of these materials apart from the rest of the instructional program. Two units of materials were field tested by the project in a few other schools in the district during spring semester, but complete results of these field tests are not yet available.

B. Non-instructional activities. Three major types of non-instructional activities were carried out in the project: 1) initial training workshops for the project teachers; 2) inservice training activities for project teachers throughout the year; 3) production of public relations materials about the project.

1. Los Angeles City training workshop. When the project teachers were first brought together in February, 1967, they participated in a four-week workshop designed specifically as an orientation to this project. The workshop program included:

a) lectures and discussions on teaching mathematics

b) demonstrations of teaching mathematics

c) lectures and discussions on working with disadvantaged youth

d) field trips to resource centers

e) lectures and practice sessions in developing objectives for the program

f) discusssions with community groups

g) one-day teaching in experimental schools

h) sensitivity training

Workshop on formative evaluation. During the summer the evaluation staff directed a workshop for the LAMMP teachers, in which Dr. Benjamin Bloom was brought in to discuss and demonstrate his theory of formative evaluation.

2. Inservice training. These activities throughout the year included field trips to resource centers, observation and discussion of demonstration lessons by visiting mathematics consultants, and working with consultants on materials which the project teachers were developing. About a dozen outside consultants in the teaching of mathematics were brought in for 1 to 5-day periods during the year. About 15-20 district personnel also provided input at various times for project teachers.

Sensitivity training was also continued to some extent after the initial workshop experience. During the year the trainers worked with each center rather than with the whole project group together.

3. Production of public relations materials. Approximately one fifth to one-half the time of the illustrators who were employed on the project was spent in preparing explanatory brochures and charts about the project. Three brochures describing the program at each center were prepared for distribution in the community. Other materials described the organization and objectives of the total project. Visibility for the project was also sought through some television coverage and various kinds of publicity within the district. Perhaps the greatest effort was made in bringing visitors to the centers. A very large number of people--from within the district and from outside--visited the three centers this year.

## IV. Evaluation

### A. Plan for evaluation

#### 1. Measurement instruments used

a) Standardized

The Iowa Test of Basic Skills was used to collect some data for evaluation. Two subtests, the Verbal and the Arithmetic (sixth Grade Level) were administered in September 1967 to students in experimental and comparison classes. The Arithmetic Subtest was given again in May, 1968.

b) Non-Standardized

Four types of non-standardized instruments were administered as a part of the evaluation.

(1) The LAMMP Diagnostic Test, a measure of achievement in a variety of mathematical skills, was constructed from items judged to be pertinent to the instructional goals of the LAMMP project. The development of this instrument originated in recognition on the part of the evaluation staff that standardized achievement tests may be generally insensitive to relatively short-term learning experiences because their items are mainly irrelevant to the particular instructional program under study.

Selection of the items was based on a two-dimensional classification system, shown in Figure 1, which organizes test items into classes of operational objectives defined by the intersection of content and process dimensions.

Figure 1

CLASSIFICATION SYSTEM FOR SELECTION OF ITEMS FOR THE LAMP DIAGNOSTIC TEST

| CONTENT | Perception | Recognition | Computation | Conservation | Classification | Seriation | Relations | Application |
|---|---|---|---|---|---|---|---|---|
| Integers | 1 | * 13 | 25 | 37 | * 49 | * 61 | * 73 | 85 |
| Rational Numbers | * 2 | 14 | 26 | * 38 | 50 | * 62 | * 74 | 86 |
| Measurement | * 3 | * 15 | * 27 | * 39 | 51 | * 63 | * 75 | 87 |
| Algebra | 4 | 16 | 28 | 40 | 52 | 64 | * 76 | 88 |
| Geometry | * 5 | 17 | * 29 | 41 | * 53 | 65 | * 77 | * 89 |
| Numerals and Place Value | 6 | * 18 | 30 | * 42 | 54 | 66 | * 78 | * 90 |
| Number Theory | 7 | * 19 | * 31 | * 43 | * 55 | * 67 | * 79 | * 91 |
| Sets | * 8 | 20 | * 32 | 44 | * 56 | 68 | 80 | 92 |
| Field Axioms and Principles | * 9 | * 21 | * 33 | * 45 | * 57 | * 69 | * 81 | 93 |
| Statistics | * 10 | 22 | * 34 | 46 | 58 | * 70 | 82 | 94 |
| Word Problems | 11 | 23 | 35 | * 47 | 59 | * 71 | * 83 | 95 |
| Facts | 12 | * 24 | 36 | 48 | 60 | 72 | 84 | * 96 |

For example, one such intersection is "integers" (content) and "classification" (process). An item from this cell is,

Which numeral has been left out?

5 10 ___ 20 25 30

a. 11
b. 15
c. 19
d. 100

It was hoped that reference to this organization of test content would result in an instrument useful for comparing groups of students with respect to specific areas of achievement in mathematics. The broad coverage of the LAMMP Diagnostic Test in terms of types of mathematical skills contrasts with the typical standardized ahievement test. To maximize the accuracy of statements about the performance of individual students, the latter often contain many items of the same type.

Two forms (Form A and Form B) of the LAMMP Diagnostic Test were constructed by randomly assigning members of item pairs selected from the same content-process category. The 49 starred cells of the content-process matrix presented in Figure 1 are represented by items on the LAMMP Diagnostic Test. The two forms of the test were constructed so as to reduce the amount of testing time and still provide for replication of particular item types. Forms A and B were randomly assigned to classes within experimental and comparison groups.

Items of the LAMMP Diagnostic Test were organized into three sections. The first, containing items 1-40, is a varied set of item types drawn from a relatively large number of cells of the matrix. The second section, items 41-52 on both forms, contains computational problems in addition, subtraction, multiplication, and division. These questions required students to produce the answers rather than select them from several alternatives. The first two sections were administered on a pretest-posttest basis to all subjects. The final section, consisting of an additional fifteen items, contains additional questions selected from the matrix after examination of instructional materials which were developed after the pretesting. This last section of the test was of course administered only on a posttest basis.

(2) A number of measures of level of cognitive development were administered to samples of experimental and comparison students at each school. These measures will be described only briefly in the preliminary report, as their analysis is presently underway. The individually administered problem situations, derived mainly from the work of Piaget and Bruner, provide measures of generalized cognitive functions rather than indications of level of school achievement as would be inferred from the LAMMP Diagnostic Test or the Iowa Test of Basic Skills. As such, the cognitive measures

reflect stages in intellectual growth as represented in developmental theory rather than stages of achievement in school subject matter as reflected in grade norms.

There were two purposes in administering the cognitive tests. First, performance on such measures will make it possible to describe the LAMMP and comparison groups with regard to a number of generalized cognitive skills presumably applicable to a variety of instructional content. Performance of LAMMP students will be compared with the approximate level of cognitive performance which, under typical circumstances, would be anticipated at the seventh grade level. Second, it may be possible to discover interactions between cognitive variables and the effects of instruction. Such information may help determine whether or not students with certain patterns of cognitive skills learn more or less than students characterized by other patterns. We might expect, for example, that students who have not developed generalized formal reasoning skill, as defined by the measures listed below, will be poorly equipped to deal with abstract mathematical content. In general, the cognitive measures have been administered in the hope of learning whether the instructional programs under study are appropriate for the entire target population or only for sub-groups of that population.

The cognitive measures can be divided into four types: Measures of the correspondence of ordinal and cardinal numbers are based on the student's ability to understand that only when each element of a series is combined with the preceding ones can its position be determined, and only their position differentiates the units, which in other respects are equivalent. Measures of conservation indicate the extent to which the student's grasp of the idea of quantity is invariant with respect to observable physical transformations. The measure of formal reasoning has to do with the ability to generate abstract rules or principles that explain observed events. Such rules are not simply the summary of experience, but rather represent a formal understanding of the principles underlying a set of events. For example, the law of specific gravity is a generalized explanation of how it is that any object will sink or float in a liquid. A probability statement is a rule for predicting the frequency or likelihood of specific events. Both of these types of rules are arrived at through what Piaget has termed formal reasoning. Finally, the measures of grouping reflect the levels at which people organize or categorize objects and events.

Brief descriptions of the instruments used are given below:

## Seriation and cardination

Seriation and cardination: (Cards) The student is presented with ten cards lettered from A-J, representing units from one to ten respectively. The student is asked how many units different cards represent, first while in order, then again after they are disarranged.

## Conservation

Conservation of area: The student is asked to compare the areas of two fields of grass (sheets of green paper), which are acutally the same size. On one field houses are placed close together in rows. On the other field the houses are spread about over the whole area.

Conservation of weight: The student is asked to compare the weights of two equal balls of clay, one of which is flattened and the other of which is left in the shape of a ball.

Conservation of volume: (Displacement of Volume) The student observes how far the water in a cup rises when a ball of clay is dropped into it. He is then asked to predict whether the water will rise the same amount when the ball of clay is flattened before it is dropped into the cup.

Formal Reasoning

Combinatorial problem: From squares of railroad board in six different colors the student is asked to make all the possible color combinations.

Floating bodies: The student is presented with various objects and asked whether or not they will float. He is then instructed to put them in a bowl of water, after which he is asked to formulate a rule stating why things float.

Grouping

Grouping: The student is presented with two different pictures and is asked how they are alike. A third picture is added to the previous two and he is again asked how all three are alike.

Classification: The student is given a sheet of paper which contains geometric figures of varying size and color. He is asked to group them according to their attributes.

(3) The Social Survey Instrument included items on:
students' school social history
attitudes toward school, self and world outlook
educational materials in the home
background characteristics of self, family and language behavior
educational aspirations of self and parents

(4) The purpose of this test is to assess children's and adolescents' attitudes toward various facets of school life as well as to other

intellectual and cultural activities. The test has been designed specifically so that these attitudes can be assessed indirectly and non-verbally

The Attitude Toward School Test consists of 40 pictures representing situations which are familiar to children and adolescents. The 40 pictures are divided into four series of 10 pictures each to measure attitudes toward a variety of activities and situation. Specifically:

1. Attitudes toward school in general.
   This series includes pictures of activities which occur in school such as assembly, the library, a cafeteria scene, etc.

2. Attitudes toward Social Studies.
   Pictures representing situations which occur in social studies classes such as looking at a map, an historical mural, etc.

3. Attitudes toward Mathematics.
   In this series pictures are included which represent classroom situations in which mathematical activities are presented, e.g., writing mathematical problems on the board.

4. Attitudes toward General Intellectual Activities. These pictures depict activities which take place outside the school, e.g., going to a museum, a concert, an athletic event.

In each picture boys and girls are depicted engaging in the various activities described above. The figures have purposely been drawn only with general outlines, the features obscured and specific cues limited, to reduce the possibility of the students' responding to irrelevant aspects of the pictures.

The students are told that they will look at a series of pictures depicting girls and boys engaging in different activities.

They are asked to indicate on a seven point scale (1) how the people in the picture feel and (2) how they would feel if they were in the situation depicted in the picture.

Prior to administering the test, students are instructed that there are no right or wrong answers, and to respond the way they really feel.

In addition to the data collected by the instruments listed above, information has been obtained from two other sources. First, regular observations of classrooms were made by the evaluation team throughout the school year. Second, the evaluation team has maintained frequent contacts with teachers and central office project personnel from the beginning of the project in February 1967. Through interviews and informal discussions information has been collected about teachers' and supervisors' perceptions of the program.

2. Pertinent dates

   a) The instructional program began September 5, 1967

   b) Evaluation data were collected with the instruments described above in a pretest period from September 26 to November 30 and in a posttest period from April 23 to May 23. Collection of information from interviews and informal contacts with project personnel began in February 1967 and has continued to date.

3. Comparison groups.

The procedures for selection of comparison classes were described in section IV, B above. Comparison classes met during the same class periods as experimental classes.

Characteristics of comparison classes are shown in Tables 5 through 8.

Table 5

Sex of Comparison Classes

| School | Fall, 1967 | | Spring, 1968 | |
|---|---|---|---|---|
| | % Girls | % Boys | % Girls | % Boys |
| Belvedere | 53 | 47 | 53 | 47 |
| Edison | 49 | 51 | 45 | 55 |
| Pacoima | 55 | 45 | 59 | 41 |

Table 6

Ethnic Groups Among Comparison Classes

| School | % Caucasian | % Negro | % Mexican-American | % Other |
|---|---|---|---|---|
| Belvedere | 1 | 2 | 94 | 3 |
| Edison | 1 | 92 | 6 | 1 |
| Pacoima | 67 | 5 | 25 | 3 |

Table 7

Grade Levels of Comparison Classes

| | Fall, 1967 | | | Spring, 1968 | | | |
|---|---|---|---|---|---|---|---|
| Belvedere | B7 | A7 | Total | B7 | A7 | B8 | Total |
| No. of Classes | 3 | 1 | 4 | | 3 | 1 | 4 |
| No. of Students | 55 | 26 | 81 | | 49 | 24 | 73 |
| Edison | | | | | | | |
| No. of Classes | 6 | -- | 6 | 2 | 6 | -- | 8 |
| No. of Students | 101 | -- | 101 | 33 | 141 | -- | 174 |
| Pacoima | | | | | | | |
| No. of Classes | 3 | 1 | 4 | | 3 | 1 | 4 |
| No. of Students | 86 | 33 | 119 | | 110 | 35 | 145 |

No. of students, all schools

| | B7 | A7 | B8 | Total |
|---|---|---|---|---|
| Fall, 1967 | 242 | 59 | -- | 301 |
| Spring, 1968 | 33 | 300 | 59 | 392 |

Table 8

Mean I.Q. by School for Comparison Classes on the California Test of Mental Maturity (1957 Short Form), Fifth Grade

| School | Verbal Mean | Non-Verbal Mean | Total Mean |
|---|---|---|---|
| Belvedere | 91.0 | 89.0 | 90.2 |
| Edison | 93.1 | 86.8 | 90.1 |
| Pacoima | 101.5 | 99.4 | 100.5 |

4. This report contains only the preliminary analysis of data from the Iowa Test of Basic Skills, the LAMMP Diagnostic Test and the Social Survey Instrument. The final report will include further analysis of data from these instruments, plus analysis of data from the measurement of cognitive development and attitudes toward school. The final report will also include some assessments based on information obtained from discussions with project personnel.

   For this report only summary statistics, i.e., mean and correlations have been used. There is no attempt at this point in the analysis to make statistical inferences.

B. Summary and Interpretation of Evaluation Data

1. Assignment of students to experimental and comparison groups

The research design adopted in the original LAMSP proposal required the identification of a target group of students within each of the three schools, followed by random assignment into experimental and comparison groups. This procedure was adopted in order to provide assurance that systematic differences did not exist at the beginning of instruction between experimental and comparison subjects. Differences in achievement observed later on thus cannot be attributed to differences that existed prior to the beginning of the program.

An example of the type of problem caused by failure to achieve random assignment is provided by the following. Assume that in one school the comparison subjects are initially higher in mathematics achievement. For a given test item further assume that 70% of the comparison subjects and 40% of the experimental subjects answered correctly at the pretest. This means that a gain of only 30% is possible for the comparison group as compared to a possible 60% for the experimental group. This kind of situation makes it relatively easy for the experimental group to look superior with respect to gain, simply because initial performance was lower.

In such a case time might as well have been saved by not bothering with a comparison group, since contrasts between experimental and comparison subjects are mainly meaningless or uninterpretable.

Were experimental and comparison groups in the three schools approximately equivalent at the beginning of the program?

Assigmment of students to experimental or comparison classes was conducted by staff members of the school involved in the project. This is usually an unsatisfactory procedure unless explicit rules are provided and adhered to by the schools. Random behavior is simply not in the nature of man. That this principle applies in the present case is evident from the data in tables 9-12.

Table 9 presents means and percentiles on the verbal subtest of the Iowa Test of Basic Skills (ITBS). While no differences in verbal ability at pretest are evident for the Belvedere students, the comparison students at Edison rank 6 percentile points above the experimental students. At Pacoima the difference is 12 percentile points, again in favor of the comparison subjects. Performance on the Arithmetic subtest of the ITBS, summarized in Table 10 shows the comparison subjects to be above the experimental subjects in all three of the

Table 9

PRETEST PERFORMANCE BY
SCHOOL FOR EXPERIMENTAL AND COMPARISON
SUBJECTS ON VERBAL SUBTEST OF IOWA TEST OF BASIC SKILLS*

| School | Pretest | |
|---|---|---|
| | Experimental | Comparison |
| Belvedere | | |
| Mean | 17.05 | 17.43 |
| %ile rank* | 23 | 23 |
| Edison | | |
| Mean | 14.66 | 16.67 |
| %ile rank | 17 | 23 |
| Pacoima | | |
| Mean | 19.83 | 24.68 |
| %ile rank | 31 | 43 |
| Total | | |
| Mean | 17.57 | 20.0 |
| %ile rank | 26 | 31 |

*Percentile ranks refer to national ITBS norms for end of sixth grade.

Table 10

PRE AND POSTTEST PERFORMANCE BY
SCHOOL FOR EXPERIMENTAL AND COMPARISON
SUBJECTS ON ARITHMETIC SUBTEST OF IOWA TEST OF BASIC SKILLS

| School | Pretest | | Posttest | |
|---|---|---|---|---|
| | Experimental | Comparison | Experimental | Comparison |
| Belvedere | | | | |
| Mean | 16.02 | 16.8 | 18.3 | 20.44 |
| %ile rank | 17 | 21 | 26 | 32 |
| Edison | | | | |
| Mean | 14.26 | 15.32 | 16.04 | 16.93 |
| %ile rank | 10 | 14 | 17 | 21 |
| Pacoima | | | | |
| Mean | 16.49 | 21.19 | 19.27 | 25.17 |
| %ile rank | 17 | 35 | 30 | 49 |
| Total | | | | |
| Mean | 15.8 | 18.1 | 17.99 | 20.89 |
| %ile rank | 17 | 26 | 26 | 35 |

schools. The superiority of the comparison subjects is only 4 percentile points at Belvedere and Edison, but it is there. At Pacoima the difference is very large indeed.

Mean scores on the LAMMP Diagnostic Test are provided in Tables 11 and 12. Table 11 is based on the first forty items of the LAMMP Diagnostic Test while Table 12 includes the 12 additional items measuring computational skills. Only in the case of Belvedere does it appear that the two groups were at the same level of initial mathematics achievement. Small but consistent differences favor comparison over experimental subjects at Edison. As in the case of the ITBS, the LAMMP Diagnostic Test at Pacoima reveals large differences in favor of comparison subjects.

Initial differences between experimental and comparison groups also appear in the data from the Social Survey Instrument. Some differences, small but consistent, are seen in all three schools. (Table 13)

Experimental students in all three schools showed higher school mobility than comparison students. Educational aspirations of experimental students themselves and their parents' aspirations for them were lower than the same aspirations for comparison students. Other differences show up when the schools are considered one

Table 11

PRE AND POSTTEST PERFORMANCE BY SCHOOL FOR EXPERIMENTAL AND COMPARISON SUBJECTS ON LAMMP DIAGNOSTIC TEST (ITEMS 1-40)*

| School | Pretest | | | | Posttest | | | |
|---|---|---|---|---|---|---|---|---|
| | Experimental | | Comparison | | Experimental | | Comparison | |
| | Fm A | Fm B | Fm A | Fm B | Fm A | Fm B | Fm A | Fm B |
| Belvedere | | | | | | | | |
| mean | 23.1 | 24.2 | 24.7 | 23.4 | 25.7 | 27.6 | 31.3 | 25.9 |
| sigma | 6.9 | 5.6 | 5.8 | 3.5 | 5.8 | 5.3 | 3.8 | 4.7 |
| N | 60 | 63 | 28 | 31 | 60 | 63 | 27 | 32 |
| Edison | | | | | | | | |
| mean | 21.9 | 21.2 | 24.7 | 21.7 | 25.3 | 26.4 | 26.5 | 28.9 |
| sigma | 4.1 | 4.8 | 4.4 | 5.6 | 4.8 | 5.3 | 5.2 | 3.9 |
| N | 21 | 32 | 32 | 24 | 24 | 34 | 33 | 18 |
| Pacoima | | | | | | | | |
| mean | 26.8 | 24.8 | 30.5 | 29.3 | 27.7 | 27.4 | 31.6 | 32.4 |
| sigma | 4.9 | 5.2 | 4.4 | 4.9 | 5.0 | 4.4 | 4.5 | 4.6 |
| N | 56 | 53 | 41 | 47 | 57 | 52 | 41 | 47 |

*The data reported in Table 11 are based only on those students who were in school for the entire 1967-1968 year.

Table 12

PRE AND POSTTEST PERFORMANCE BY SCHOOL
FOR EXPERIMENTAL AND COMPARISON SUBJECTS
ON LAMP DIAGNOSTIC TEST (ITEMS 1-52)*

| School | Pretest | | | | Posttest | | | |
|---|---|---|---|---|---|---|---|---|
| | Experimental | | Comparison | | Experimental | | Comparison | |
| | Fm A | Fm B | Fm A | Fm B | Fm A | Fm B | Fm A | Fm B |
| Belvedere | | | | | | | | |
| mean | 29.3 | 31.6 | 32.1 | 31.1 | 32.1 | 34.9 | 39.6 | 33.9 |
| sigma | 8.1 | 6.7 | 6.3 | 4.3 | 6.7 | 6.5 | 5.3 | 5.5 |
| N | 60 | 63 | 28 | 31 | 60 | 63 | 27 | 32 |
| Edison | | | | | | | | |
| mean | 27.7 | 27.5 | 31.4 | 28.5 | 32.1 | 34.3 | 33.9 | 37.4 |
| sigma | 4.5 | 6.1 | 5.0 | 7.2 | 5.5 | 6.1 | 6.4 | 5.2 |
| N | 21 | 32 | 32 | 24 | 24 | 34 | 33 | 18 |
| Pacoima | | | | | | | | |
| mean | 33.1 | 32.2 | 38.0 | 37.6 | 34.5 | 35.4 | 39.2 | 40.6 |
| sigma | 5.6 | 6.2 | 5.2 | 5.5 | 5.5 | 5.3 | 6.2 | 5.5 |
| N | 56 | 53 | 41 | 47 | 57 | 52 | 41 | 47 |

*The data reported in Table are based only on those students who were in school for the entire 1967-1968 year.

at a time. At Edison the experimental students expressed less liking of school, less trust of people and more feeling of powerlessness than comparison students (Table 14). (It is interesting to think of how these characteristics might have been unconsciously taken into account by the teachers who assigned students to experimental and comparison classes.) At Pacoima experimental students had more Spanish-speaking friends than comparison students (67% experimental, 53% comparison). Even at Belvedere the experimental students expressed less trust of people than comparison students (50% experimental, 65% comparison), and more experimental students came from homes where Spanish rather than English was spoken (64% experimental, 47% comparison). Like the data from the mathematics tests, the results of the Social Survey Instrument tend to show the comparison students in a more favorable position than experimental students as far as an academic stiuation is concerned.

The implication of these data is that subjects were not randomly assigned to experimental and comparison conditions at Edison and Pacoima. This renders invalid any inferences about the effects of instructional variables based upon differences between experimental and comparison subjects.

Table 13

PERCENTAGE OF STUDENTS EXPRESSING AGREEMENT WITH SELECTED ITEMS ON SOCIAL SURVEY INSTRUMENT, PRETEST

| Item | Belvedere | | Edison | | Pacoima | |
|---|---|---|---|---|---|---|
| | Exp. | Comp. | Exp. | Comp. | Exp. | Comp. |
| Have attended 3 or more schools | 56 | 48 | 67 | 57 | 55 | 46 |
| I would like to go to college | 63 | 75 | 76 | 90 | 62 | 74 |
| Parents would like me to go to college | 68 | 78 | 81 | 91 | 70 | 80 |

Table 14

PERCENTAGE OF STUDENTS AT EDISON EXPRESSING AGREEMENT WITH SELECTED ITEMS ON SOCIAL SURVEY INSTRUMENT, PRETEST

| Item | Experimental | Comparison |
|---|---|---|
| Like school | 63 | 80 |
| People can be trusted | 42 | 55 |
| There isn't much people can do about the way things are going to turn out | 77 | 61 |

2. Measurement of Learning in Mathematics

a) Teacher ratings of relevancy of LAMMP Diagnostic Test items

Insofar as possible items comprising forms A and B of the LAMMP Diagnostic Test were selected to be representative of instructional goals of the LAMMP project. However, the major proportion of the items were selected during the summer of 1967 before it was possible to inspect LAMMP instructional materials, the bulk of which were developed later on during the 1967-68 school year. Decisions as to the content of the LAMMP Diagnostic Test were therefore guided by information gleaned during early planning meetings of the LAMMP teaching staff and from those instructional materials developed for the initial stages of the program. Since some assumptions had to be made about test content, there were grounds for concern about the test's relevancy as a major evaluation instrument for the LAMMP program.

This matter was investigated by having teachers of both experimental and comparison classes evaluate the relevance of each item of the LAMMP Diagnostic Test. Shortly before the end of the spring term the teachers were asked to judge the relevancy of

instruction in their classes to LAMMP Diagnostic Test items during the preceding year. Specifically, they were instructed to,

> "Make a judgment on the extent to which instruction in your mathematic classes this year would facilitate student's ability to answer each item correctly."

The following 5-point rating scale was provided.

Instruction in my classes this year.....

| Definitely should not facilitate ability to answer | Probably would not facilitate | Uncertain | Probably should facilitate | Definitely should facilitate ability to answer |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

Thirteen teachers of experimental and 13 teachers of comparison classes agreed to complete the judgments. Mean ratings for each of the 52 items administered in the pretest were computed separately for experimental and comparison teachers within each school. These and other related data were analyzed for the purpose of answering three questions.

(1) Did the teachers view the test as relevant to their instruction and were there differences in the ratings of experimental and comparison teachers?

Mean ratings averaged over the 52 items in each form of the LAMMP Diagnostic Pretest

are presented in Table 15. These ratings, identified according to school and experimental versus comparison teachers, show that the test as a whole was judged to be relevant to instructional goals as defined by the teachers themselves. Only one of the means (Pacoima, comparison subjects, form B) was slightly below the midpoint of 3.0 on the scale of judged relevancy. One might have anticipated that the experimental teachers would have seen the test as more relevant, since their expressed instructional goals provided the basis for item selection. This was certainly not the case. The ratings summarized in Table 13 reveal no clear pattern of differences between experimental and comparison ratings. Indeed, the ratings of comparison teachers at Belvedere are the highest in the table.

(2) To what extent did experimental and comparison teachers agree on the relevancy of individual items of the LAMMP Diagnostic Test? While overall ratings of item relevancy were similar for the experimental and comparison teachers, it may be that the two groups saw different items as relevant. If there were

Table 15

MEAN RATINGS OF RELEVANCY
OF LAMMP DIAGNOSTIC TEST ITEMS*

| School | Experimental | | Comparison | |
|---|---|---|---|---|
| | Fm A | Fm B | Fm A | Fm B |
| Belvedere | 3.71 (5) | 3.81 (5) | 4.28 (3) | 4.23 (3) |
| Edison | 4.0 (4) | 3.98 (4) | 3.94 (8) | 3.82 (8) |
| Pacoima | 3.61 (4) | 3.36 (4) | 3.59 (2) | 2.95 (2) |

*The numbers in parenthesis indicate the number of teachers contributing to each mean rating.

such differences on a number of items it may be possible to identify subsets of items on which the achievement of experimental and comparison students might be expected to differ. To answer this question mean ratings by experimental and comparison teachers of the 52 items were inter-correlated within each school for each test form. These correlations are reported in Table 16. Only for Edison is there a relatively high relationship between the ratings of experimental and comparison teachers. For the other two schools the relationships, while always positive, are quite low. One interpretation of this result is the report of members of the evaluation team that Edison appeared to be the only school where discussions about instructional goals and procedures occurred between experimental and comparison teachers. Although the LAMMP Diagnostic Test appears to be based on appropriate overall content for both experimental and comparison classes, evidently somewhat different subsets of items were seen as relevant by the two groups of teachers in two of the three schools.

(3) As reflected in the ratings of Diagnostic Test items, to what extent were curriculum

Table 16

CORRELATIONS BETWEEN EXPERIMENTAL AND COMPARISON TEACHER'S RATINGS OF LAMMP DIAGNOSTIC TEST ITEMS*

| | Belvedere | Edison | Pacoima |
|---|---|---|---|
| Form A | .35 | .62 | .20 |
| Form B | .12 | .64 | .18 |

*Correlations were computed over mean ratings on 52 test items.

Table 17

RELATIONSHIPS BETWEEN PROPORTION ANSWERING ITEMS OF LAMMP DIAGNOSTIC TEST CORRECTLY AT PRETEST AND TEACHER'S RATINGS OF RELEVANCY TO INSTRUCTION

| School | Experimental | | Comparison | |
|---|---|---|---|---|
| | Fm A | Fm B | Fm A | Fm B |
| Belvedere | .25 | .62 | .14 | .30 |
| Edison | .44 | .58 | .27 | .32 |
| Pacoima | .53 | .62 | .10 | .25 |

goals directed at weaknesses in mathematical skills exhibited by students at the beginning of the LAMMP program?

Educators would doubtless agree that instruction ought to be aimed at gaps in achievement rather than at what has already been mastered by most students. Having noted that teachers see the Diagnostic Test as generally reflecting their own instructional goals, it is legitimate and interesting to examine relationships between the proportion of students passing each item at pretest and the teacher ratings of item relevancy made at the end of the year. Such relationships should give an indication of the extent to which both experimental and comparison curricula were directed at needs of the students. It can be argued that ideally the correlations reported in Table 17 should be negative, indicating that teachers place greater emphasis on those skills in which students are initially weak. Correlations of approximately zero magnitude would indicate a lack of any relationship between student skills at entry and the instructional program, hardly a desirable situation. Positive correlations, as appear in Table 17, are even less desirable.

Teachers in experimental and comparison groups actually appear to have been directing instruction at those areas in which students were initially more capable! Moreover, since in every case the coefficient on the right side of Table 17 is higher than the corresponding value at the left, there appears to have been a greater tendency for experimental than for comparison teachers to emphasize those skills at which students were initially, more able. The results reported in Table 17 have significant implications not only for the instructional program under study but for educational practice in general, and great care must be taken in their interpretation. If it is really true that teachers in both experimental and comparison programs were relatively unaware of the particular pattern of entry skills characterizing their students and actually placed greater emphasis on topics with which students were already relatively more familiar, it would be unrealistic to anticipate significant improvement in performance. Most of the time would have been spent on topics which students already know. Why might teachers do this? One explanation

may be that the task of teaching students like those in the LAMP Program may be rendered easier by directing instruction at areas of achievement in which some competencies already existed.

The above interpretation, while consistent with the data of Table 17, is certainly difficult to accept in its unequivocal form. One might wonder, for example, how any learning at all can go on in the schools if the curricula were confined to what students already know? Of course, we are not dealing here with data taken from middle class, suburban schools. The present students are already very far behind in mathematics achievement. In the typical situation the deficit could be expected to widen in the future. Thus it may be that the frustrations encountered in teaching educationally handicapped students plus the need perceived by the teachers to provide such students with success experiences lead in the direction of making things easier by placing greater emphasis on those areas of content in which present capabilities of students are most developed.

There is at least one rather different explanation of the relationships in Table 17.

We are, after all, dealing with ratings, with what teachers said was emphasized in instruction. Moreover, these ratings were collected at the end of the year after the teachers had ample opportunity to become familiar with their students' strengths and weaknesses in mathematics. Perhaps the ratings of relevancy of items do not reflect what was done in the classroom at all, but simply teachers' perceptions of student capabilities. This interpretation does have the advantage of providing a possible basis for the correlations being higher for the experimental teachers. One might guess that at the end of the year the teachers, now quite familiar with their students' capacities, would unconsciously "take credit" by rating those items on which students would do relatively better as more relevant to their own instructional goals. The experimental teachers, acutely aware of the attention being paid to the achievement of their students, might be more unconsciously motivated to present such a picture, hence the higher correlations on the left side of Table 17.

It would be unwise on the basis of the data presently available to choose between the

two types of explanations. As a general issue of considerable significance for educational research, the relationship between instructional goals and student skills should be investigated further. For the purpose of the present report, the most probable conclusion is that both kinds of explanations are valid. Teachers from both experimental and comparison classes were evidently insufficiently aware of the specific patterns of skills their students possessed at entry to the 7th grade. Gross comparisons of total achievement test scores with national norms do not provide meaningful diagnostic information. Finally, the correlations are probably increased due to the very real motivation on the part of all teachers to effect improvements in their students' achievement.

Conclusions

1) The approach used in the development of the LAMMP Diagnostic Test appears to be a useful one in the sense that both comparison and experimental teachers see the content as in general relevant to their instructional goals. Moreover, there appear to be no overall differences in the relevancy of the

LAMMP Diagnostic Test items as perceived by comparison versus experimental teachers. In this sense the test is seen as a fair one by both groups.

2) In spite of the similarity of overall relevancy ratings by experimental and comparison teachers, there is evidence that some differentiation among curriculum goals of the two groups might be made for subsets of items. The nature of this differentiation will be explored in the later report and may provide guidelines as to the specific differences in the instructional outcomes of experimental and comparison classes, if such exist.

3) There is no evidence that specific strengths and weaknesses of the student populations were taken into account in the setting of instructional goals, both in the case of the LAMMP and comparison classes. According to at least one interpretation of the data, there is in fact evidence to the contrary. Teachers may have emphasized students' strengths rather than weaknesses. This tendency appears to be more characteristic of the experimental classes. While the finding needs to be replicated in other

research, it is of high relevance to the design of all kinds of instructional programs. In the future development of instructional materials for LAMMP, account should be taken of available diagnostic information as to the entry skills of the students.

b) Comparisons between total scores on the LAMMP Diagnostic Test and on the Iowa Test of Basic Skills

Comparisons between pre- and posttest scores on the LAMMP Diagnostic Test are presented for Items 1-40 in Table 11 and for Items 1-52 in Table 12. Tables 11 and 12 have also been presented graphically in Figures 2 - 5. Probably the only clear-cut result pertaining to achievement gains apparent from these data is that all groups made some gains over the year on overall test scores. There do not appear to be consistent differences in achievement gains between experimental and comparison groups. Because of the lack of comparability of experimental and comparison groups in two of the schools, such comparisons are not particularly meaningful in any case. Similar results are apparent in the case of the Arithmetic section of the ITBS Test, as reported in Table 10.

Figure 2

PRETEST - POSTTEST DIFFERENCES IN MEANS, BY SCHOOLS
(Form A, Items 1-40)

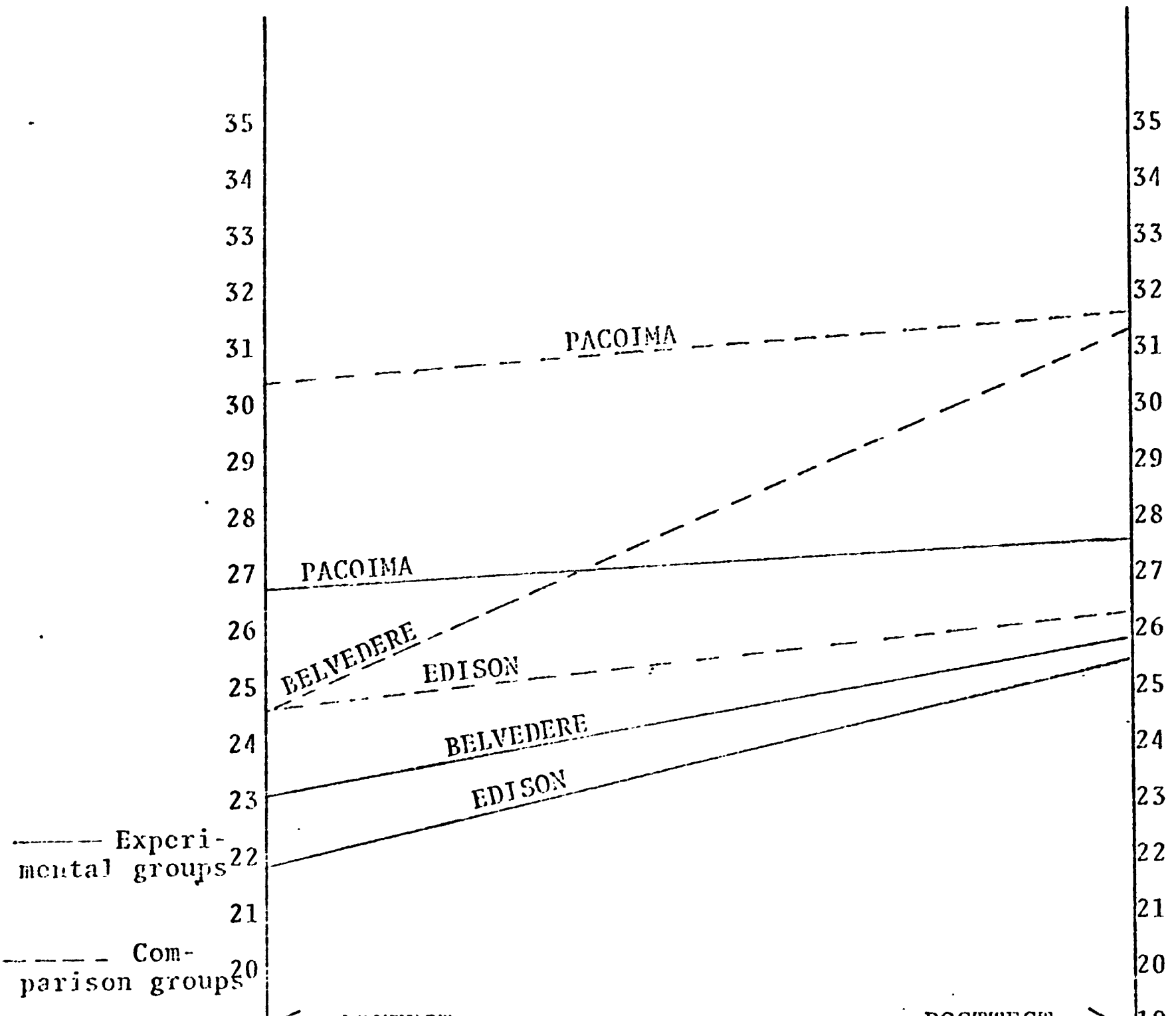


Figure 3

PRETEST - POSTTEST DIFFERENCES IN MEANS, BY SCHOOLS
(Form B, Items 1-40)

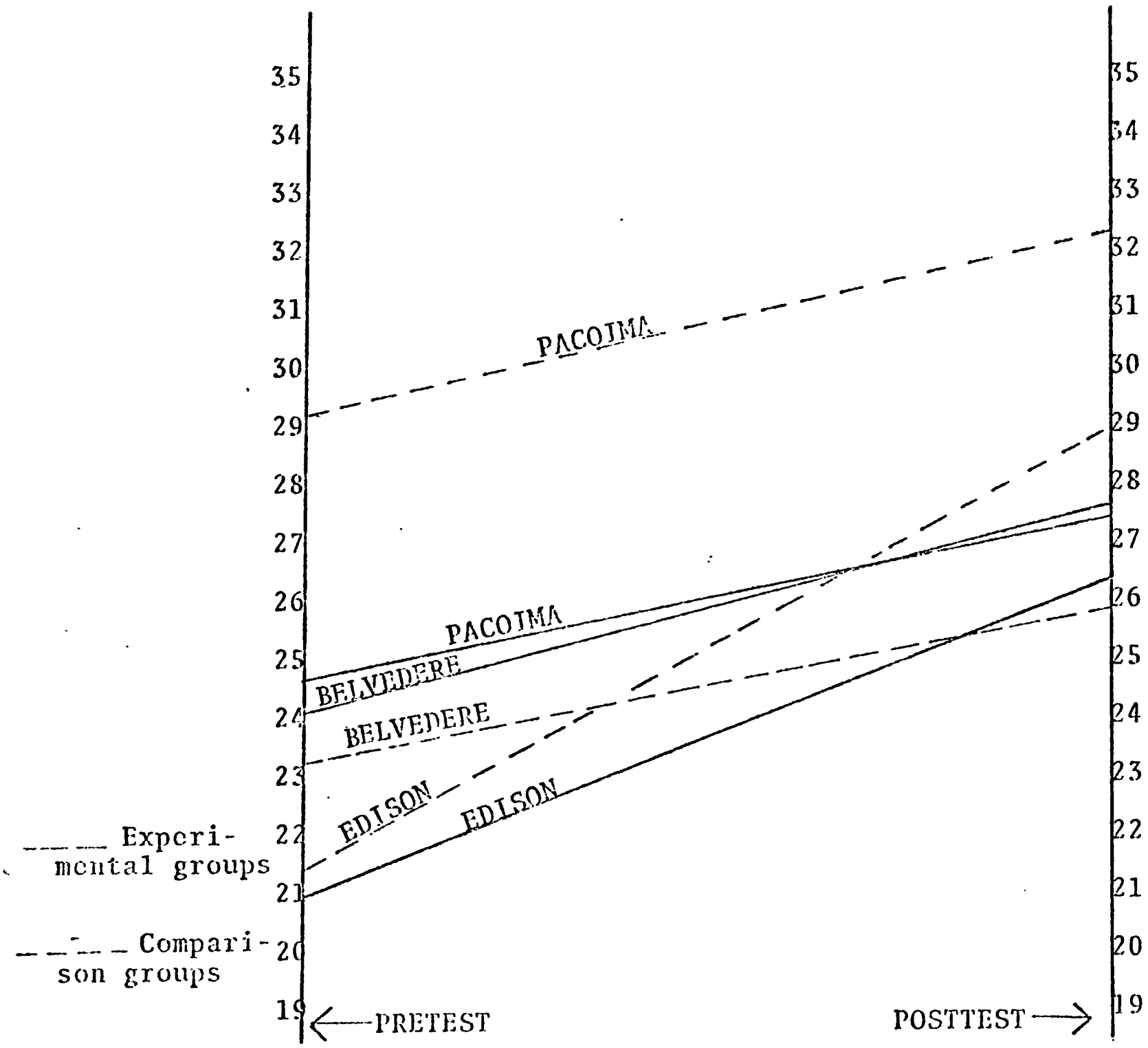


Figure 4

MATHEMATICS DIAGNOSTIC TEST RESULTS, FORM A

(Items 1-52)

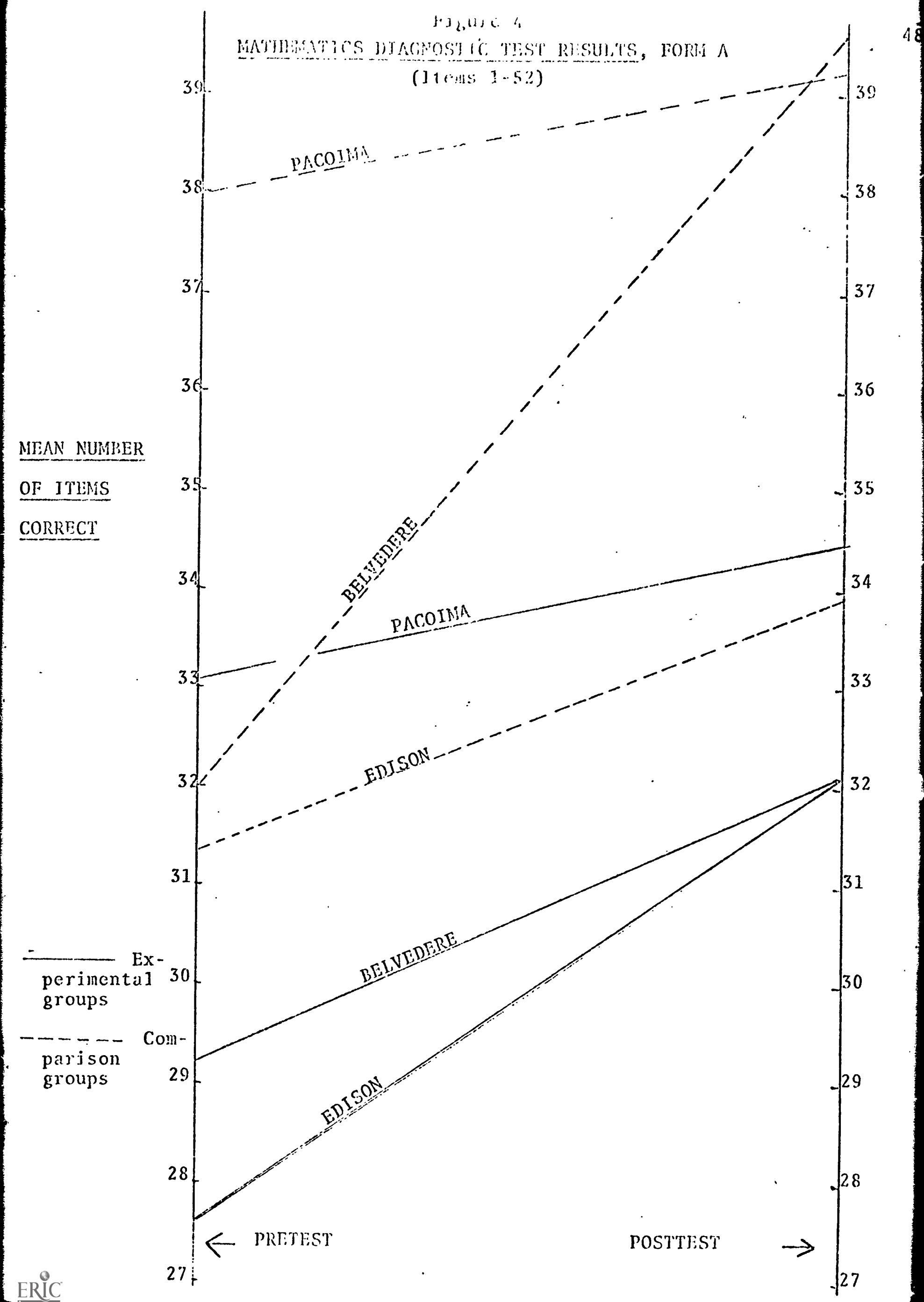
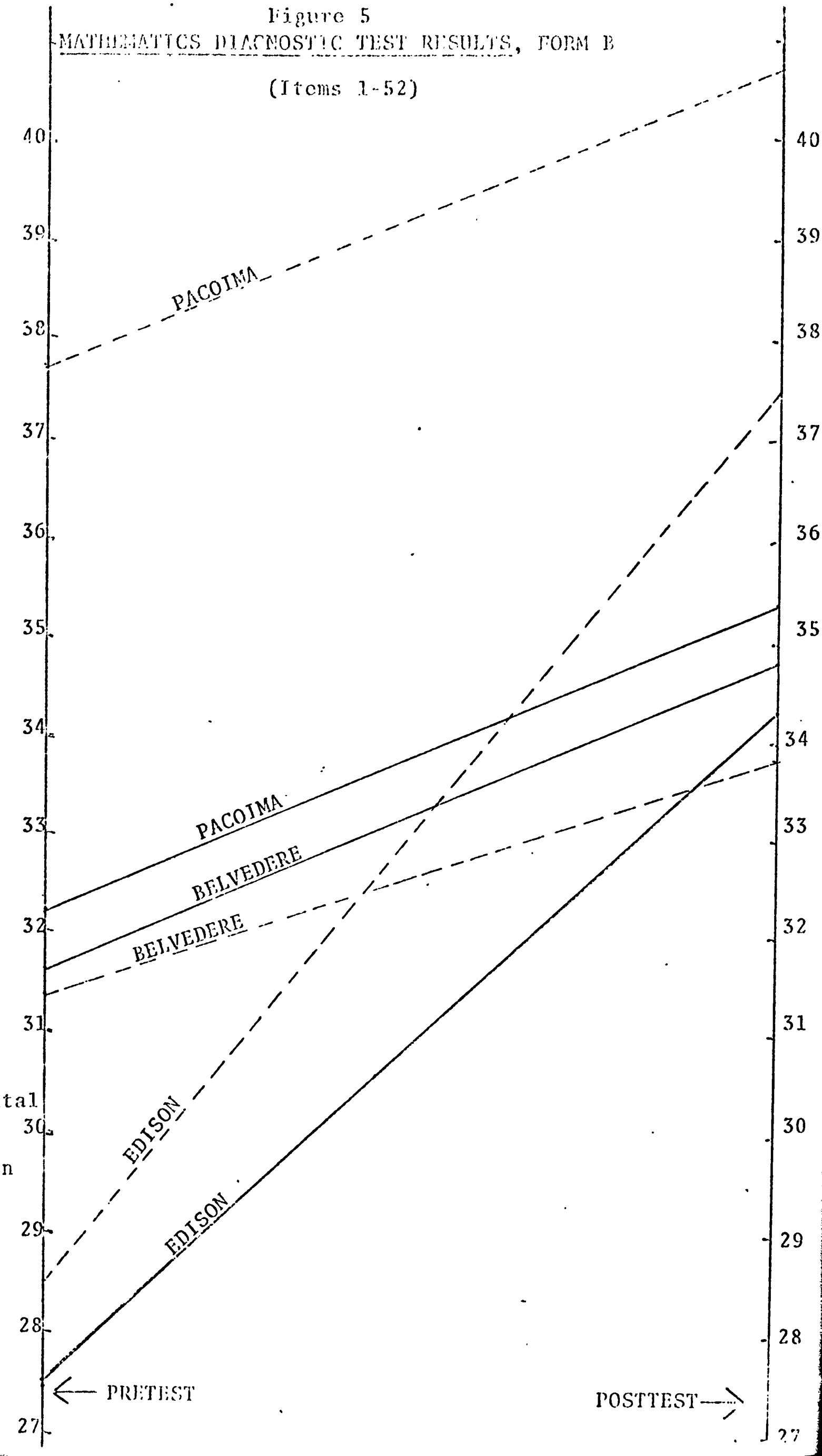


Figure 5

MATHEMATICS DIAGNOSTIC TEST RESULTS, FORM B

(Items 1-52)



Despite current practice, it is our view that comparisons based on total test scores are not especially useful, except in a very general summative sense. The content of the LAMMP Diagnostic Test, for example, is deliberately broad in scope so as to reflect as many as possible of the goals of the three LAMMP curriculum development teams. It is to be expected that gains in certain areas of concentration may be mainly obscured in a total test score by lack of gains in areas which received little or no emphasis.

Because total scores are not especially informative, we are in the process of combining LAMMP Diagnostic Test items from the same content and/or process dimension of the content-process matrix in an attempt to produce relatively homogeneous subscores, which will be more likely to reflect particular emphases in instruction. In this regard for example, we have some preliminary evidence of increased skills in algebraic concepts and operations for experimental classes in one of the schools and for increased achievement in concepts relating to set theory for comparison classes in another school. The original intention has been to use the teachers' ratings of item relevance to identify subsets of items on which gains might

be expected. The usefulness of this approach is somewhat in doubt since, as indicated earlier, teachers tended to see as relevant those items on which students were already performing at a relatively higher level.

3. Measurement of students' attitudes

A group of items on the Social Survey Instrument gives some information about students' view of the world and of themselves. Pre- and posttest scores on these items show certain trends in all students, both experimental and comparison. All groups at the end of the year express less liking of school, less satisfaction with themselves, are less in agreement with the statement that "children should obey all the rules their parents make for them", and are more in agreement with the statement, "The best way to settle some arguments is by a good fight". There may be some clues here about what tends to happen to these students in junior high school.

There do not appear to be any strong patterns of differences between pre-post test scores for experimental groups as contrasted with scores for comparison groups. In the case of Edison, however, there is some indication of a change in the outlook of experimental students. At the beginning of the year these students expressed less trust in people and more feeling of powerlessness than the comparison students. At the end of the year, agreement with the statement,

"In general people can be trusted", had decreased for both groups, but much less for the experimental group (47% to 37% for experimental; 52% to 26% for comparison). Pretest and posttest scores on the item, "There isn't much people can do about the way things are going to turn out in life", showed a decrease in agreement for the experimental group (78% to 67%) and an increase in agreement for the comparison group (57% to 66%). Finally, agreement with the statement, "There are times when I think that I am no good at all", dropped from 69% to 41% for the experimental students, while it reamined at the original level for comparison students (64% to 60%). Although slight, the evidence does suggest that the experimental group at Edison may be shifting toward a more favorable self-concept.

4. Combining and contrasting data from different schools

Analysis based upon combined data from different schools are probably acceptable where the purpose is to provide gross summary information. In contrast, where the purpose is to make inferences about the effects of instructional conditions (e.g., all LAMMP students versus all comparison students), such combinations are suspect from several points of view. For one thing, the nature of the ultimate reference population is undefined. Student bodies in the three schools differ sharply in ethnic character. While

all three schools are below average in verbal and mathematical achievement on the Iowa Test of Basic Skills, data reported in Tables 9 and 10 for Verbal and Arithmetic subtests do reveal differences among the three institutions. With respect to the ITBS Verbal score, for example, experimental students score at the 17th, 23rd and 31st percentiles for Edison, Belvedere, and Pacoima, respectively. In the case of the ITBS Arithmetic subtest the experimental students at the 3 schools fell at the 17th percentile for Belvedere and Pacoima, but at the 10th percentile for Edison. Differences were also apparent for number of items correct at pretest on the LAMMP Diagnostic Test (Tables 11, 12).

Similar conclusions would doubtless be reached from an examination of the cultural and demographic variables. For example, on the Social Survey Instrument of those students aware of their father's educational history, 48% at Belvedere, 26% at Edison, and 36% at Pacoima reported that their fathershad less than a high school education. (The relative ranking of the schools might change if all the students knew the number of years of education received by their fathers).

Since the student bodies of the three schools differ in so many ways, it is likewise not appropriate to

make distinctions between schools as to the relative effectiveness of LAMMP instructional programs. For example, measures of student achievement cannot provide a basis for concluding that the LAMMP program at Belvedere, emphasizing a "mathematics laboratory approach", is superior or inferior to the media-based program at Pacoima. If the differential effectiveness of the three LAMMP instructional programs had been of interest, a quite different and far more complex experimental design should have been adopted. In any case, comparative questions of this sort are usually of minor importance and rarely justify the trouble that must be taken in answering them. The most generalized objective of the LAMMP program presumably is to produce effective instructional materials in all three of the Centers. If it is at some point necessary to choose one approach over the others, comparative levels of student achievement will probably not turn out to be a decisive factor. Such a decision is more likely to be based on considerations like cost of materials and equipment, training requirements for staff, and similar matters.

## V. Recommendations

The following points should be taken into account, either with respect to the future operation of the LAMMP program or with reference to the evaluation procedures by which its outcomes are to be monitored.

A. In order to insure random assignment of students to experimental and comparison groups, it is recommended that assignment to classes for the 1968-69 school year be done under the supervision of appropriate technical personnel from the Los Angeles City Schools or from the LAMMP evaluation team.

B. It is of course hoped that the evaluation report for the first year will provide some useful suggestions for future development of the program, but in the main, the role of the evaluation team during this first year was perceived as that of providing evaluative data with respect to relatively long term outcomes rather than facilitating the day to day processes by which desirable outcomes are achieved.

This approach is described as "summative" in the literature of evaluation. In its more negative aspects this sort of evaluation can be likened to the role of St. Peter at the Golden Gates, judging the product but not participating in its shaping. A very different approach has been termed "formative" evaluation. In its extreme form, formative evaluation is concerned solely with

developing the best possible product, in the hope that the product will meet whatever summative standards are later imposed. In this sense the formative evaluator is more like the parish priest who does his best to guide the member of his congregation along the path of righteousness, but who, whatever the quality of the finished product, acts as the advocate at the critical moment.

Some of the experiences gained during the initial year of the LAMMP project have led members of the evaluation team to feel it would be desirable in the future to engage, at least to some extent, in a more formative role, vis a vis the LAMMP program. Although we would see the evaluation effort next year as having both summative and formative aspects, we do sense that the interests of LAMMP and the Center for the Study of Evaluation of Instructional Programs would be benefited by a closer connection between the evaluation and the process of developing the LAMMP instructional program. This need is especially apparent in the seeming lack of mesh between student skills at entry and specific instructional goals, as indicated by the teachers' ratings of LAMMP Diagnostic Test items. Had diagnostic information been communicated to the teachers in the initial phases of the program, there might have been an earlier re-assessment of instructional objectives.

We therefore propose that for the next year a procedure be worked out for providing teachers with feedback as to student accomplishment. One way of doing this would be to have the teachers themselves examine the revised content-process matrix, which was used to construct the LAMMP Diagnostic Test, and select those areas in which they hope to effect improvements in student achievement. This process should be guided by what we have learned this year about the entry skills of LAMMP students. The evaluation team would then develop appropriate tests for administration at two or three points during the year, in an effort to help LAMMP teachers monitor the progress of their students in the light of their own instructional objectives. This kind of activity on the part of the evaluation team need not interfere with the requirements of the more summative type of evaluation.

An even more direct approach to formative evaluation would involve working with LAMMP teachers in developing tests for assessing student progress with respect to units of instruction extending over relatively brief periods of time.

There are some strong arguments in favor of this approach. First, the more frequent testing would increase the teachers' opportunities to discover and provide for individual needs. The workshop sessions with Dr. Benjamin Bloom last summer represented an attempt to orient

teachers to this kind of procedure. Second, in order to assess the effectiveness of teacher-made materials at successive stages in their development, it will be necessary for teachers to gather systematic evaluative data as they try out and modify the materials. Since it is impractical for the evaluation staff to try to prepare all the tests which would be needed, we suggest that the evaluation staff serve in a consultant capacity to the teachers in a joint effort at formative evaluation.

C. Although this report has not presented systematically the data available from informal observations of the course of this project, two recommendations seem obvious from even a cursory consideration of the context within which the instructional program is proceeding.

1. It is recommended that the time of the resource personnel (e.g., counselor, mathematics consultant, illustrators) be distributed so that most of it is devoted to actual work with teachers in the three centers. It is also recommended that any vacancy which occurs among resource personnel be filled rapidly so that teachers will not be left for long periods without the assistance they need in order to accomplish the objectives of the program.
2. Conditions necessary for attaining the dual objectives of instruction and the creation of new materials by the project teachers should probably be

re-examined. The notion of teachers creating and trying out new instructional materials in close conjunction with teaching classes is an interesting idea to explore. To be successful, however, it may require a more flexible organization than that used this year. A systematic effort may have to be made to discover various balances between teaching and working on materials which would be favorable to the accomplishment of both objectives. For example, teachers in each center might rotate assignments to periods of concentration on developing materials and periods of testing materials in teaching; teaching assistants might be employed in a variety of ways; or different patterns of organization might be tried in different centers.

In any event, consideration should be given to the kinds of difficulties which were apparent this year in teachers' efforts to accomplish both objectives. First, as noted above, the special resources which have been built into the project to help in the development of materials need to be available to teachers on a more frequent and reliable basis. Second, the cost in teacher time which is required by publicity efforts, especially large numbers of visitors, and by frequent in-service training activities should be weighed

against the amount of teacher time required to accomplish both objectives of the program.

If the formative evaluation approach, which has been recommended, is adopted, it will also be important to provide teachers with enough time to work with the evaluation team and on their own in the preparation of testing materials. This additional demand on time could bring large benefits in the better monitoring of pupil progress and the adequate assessment of newly created instructional materials.