NIBBELINK, William Henry, 1939-
THE USE OF AN ANECDOTAL STYLE OF CONTENT
PRESENTATION AS A MOTIVATIONAL AND
INSTRUCTIONAL DEVICE FOR SEVENTH GRADE
UNDER-ACHIEVERS IN MATHEMATICS. [Pages
183-281, "The Narrative for the Experimental
Unit", and pages 283-352, "The Student
Workbook for the Experimental Unit", not
microfilmed at request of the author].

University Microfilms, A XEROX Company, Ann Arbor, Michigan

© 1971
WILLIAM HENRY NIBBELINK
ALL RIGHTS RESERVED

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED
THE USE OF AN ANECDOTAL STYLE OF CONTENT PRESENTATION AS A MOTIVATIONAL AND INSTRUCTIONAL DEVICE FOR SEVENTH GRADE UNDER-ACHIEVERS IN MATHEMATICS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of the Ohio State University

By

William H. Nibbelink, B.A., M.S.

The Ohio State University

1971

Approved by

[Signature]

Adviser

College of Education
PLEASE NOTE:

Some pages have indistinct print. Filmed as received.

UNIVERSITY MICROFILMS.
To the reader:

As of March 11, 1971; the copyright for appendixes "A" and "B", registration number "A186331", is assigned to the Charles E. Merrill Company of Columbus, Ohio.

Therefore, appendixes "A" and "B" are not included with copies of the dissertation distributed by University Microfilms.

Requests for the materials should be addressed to the Senior Editor, Mathematics; Charles E. Merrill Publishing Co.; 1300 Alum Creek Drive; Columbus, Ohio 43216.

Chapter II (pages 26 through 61) of this volume describe appendixes "A" and "B".
ACKNOWLEDGMENTS

A special thank you to Dr. Alan Osborne for encouraging me to follow my first interest during the period of choosing a dissertation project. Without his help in arranging for the project and his optimism, I believe I would have heeded voices that suggested little hope for introducing a sizable amount of "new" material into a school system's curriculum.

To members of the reading committee, Dr. Harold Trimble, Dr. Joe Crosswhite, and Dr. James Duncan, both gratitude and admiration are offered: gratitude for their helpful suggestions and comments; and admiration for their ability to deal kindly and patiently with those peculiar creatures called graduate students.

The cooperation and assistance offered by the Department of Evaluation and Research of the Columbus Public School System is much appreciated, as is that offered by the teachers who participated in the program.

Last, but certainly not least, a salute to my long-suffering wife and sometimes fatherless children. They must have wearied awfully of hearing each trial of the past few years described as the critical one.
VITA

April 7, 1939 ................ Born - Oskaloosa, Iowa

1961 ....................... B.A.; Calvin College; Grand Rapids, Michigan

1961-1963 ................. Teacher; S.W. Minn. Chr. High School; Edgerton, Minnesota

1965 ....................... M.S.; The University of Iowa; Iowa City, Iowa

1965-1968 ................. Assistant Professor, Mathematics; Central College; Pella, Iowa

1968-1971 ................. Graduate Teaching Assistant; The Ohio State University; Columbus, Ohio

FIELDS OF STUDY

Major Field: Mathematics Education

Studies in Mathematics....Professor James F. Jakobsen, The University of Iowa

Studies in Mathematics Education....Professor Harold C. Trimble, The Ohio State University
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>VITA</td>
<td>iii</td>
</tr>
<tr>
<td>Listing of Tables</td>
<td>vi</td>
</tr>
</tbody>
</table>

## Chapter

<table>
<thead>
<tr>
<th>I. Problem Statement, Background, and Intentions of the Study</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Definition</td>
<td>1</td>
</tr>
<tr>
<td>Materials intended for &quot;Under-achievers&quot; in Mathematics</td>
<td></td>
</tr>
<tr>
<td>A Brief Description of the Experimental Program</td>
<td></td>
</tr>
<tr>
<td>A Brief Description of the Tests Used</td>
<td></td>
</tr>
<tr>
<td>A Brief Description of the Sample Used</td>
<td></td>
</tr>
<tr>
<td>A Brief Sketch of the Research Design</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. The Experimental Unit</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>26</td>
</tr>
<tr>
<td>Primary Goals of the Narrative</td>
<td></td>
</tr>
<tr>
<td>Primary Goals of the Student Workbook</td>
<td></td>
</tr>
<tr>
<td>The Policy on Supplementary Exercises</td>
<td></td>
</tr>
<tr>
<td>Pilot Development of the Experimental Unit</td>
<td></td>
</tr>
<tr>
<td>A Comment on the Effect of &quot;New&quot; Materials</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Research Design and Methodology</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Pretest-Posttest Comparison</td>
<td>62</td>
</tr>
<tr>
<td>The Experimental-Control Study</td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1.</td>
<td>The Nature of the Experimental Mortality</td>
</tr>
<tr>
<td>2.</td>
<td>Results on Psychological Scales for the Pretest-Posttest Comparison</td>
</tr>
<tr>
<td></td>
<td>over outer-city and Inner-city Subjects</td>
</tr>
<tr>
<td>3.</td>
<td>Results on the Mathematics tests for the Pretest-Posttest Comparison</td>
</tr>
<tr>
<td></td>
<td>over individual Outer-city Classes</td>
</tr>
<tr>
<td>4.</td>
<td>Results on the Mathematics Tests for the Pretest-Posttest Comparison</td>
</tr>
<tr>
<td></td>
<td>over individual Inner-city Classes</td>
</tr>
<tr>
<td>5.</td>
<td>Results on the Mathematics Tests for the Pretest-Posttest Comparison</td>
</tr>
<tr>
<td></td>
<td>over Outer-city Subjects and Inner-city Subjects</td>
</tr>
<tr>
<td>6.</td>
<td>Results on the Psychological Scales for the Experimental-Control Study</td>
</tr>
<tr>
<td></td>
<td>over the Experimental and Control Classes</td>
</tr>
<tr>
<td>7.</td>
<td>Results on the Psychological Scales for the Experimental-Control Study</td>
</tr>
<tr>
<td></td>
<td>comparing the two Groups, Experimental and Control</td>
</tr>
<tr>
<td>8.</td>
<td>Results on the Mathematics Tests for the Experimental-Control Study</td>
</tr>
<tr>
<td></td>
<td>over the Experimental and Control Classes</td>
</tr>
<tr>
<td>9.</td>
<td>Results on the Mathematics Tests for the Experimental-Control Study</td>
</tr>
<tr>
<td></td>
<td>comparing the two Groups, Experimental and Control</td>
</tr>
<tr>
<td>10.</td>
<td>Display of Results from the Pretest-Posttest Comparison Study and &quot;NLSMA&quot;</td>
</tr>
<tr>
<td></td>
<td>Data on Test II (Open Sentences)</td>
</tr>
<tr>
<td>11.</td>
<td>Display of Results from the Pretest-Posttest Comparison Study and &quot;NLSMA&quot;</td>
</tr>
<tr>
<td></td>
<td>Data on Items &quot;1&quot; through &quot;6&quot; of test III (the Computations Test)</td>
</tr>
</tbody>
</table>
12. Display of Results from the Pretest-Posttest Comparison Study and "NLSMA" Data on Items "7" through "12" of Test III (the Computations Test) .................. 117

13. Display of Results from the Pretest-Posttest Comparison Study and "NLSMA" Data for Items "1" through "6" of Test IV (the Algorithms Test) .................. 122

14. Display of Results from the Pretest-Posttest Comparison Study and "NLSMA" Data for Items "7" through "12" of Test IV (the Algorithms Test) .................. 123

15. Rankings for Items on Test IV as Determined by Percentages of Correct Responses .................. 125

16. Correlation Matrix for Mathematics and Reading Test Results .................. 127

17. Differences between Correlations for Reading Scores and Mathematics Scores from Pretest to Posttest .................. 128
CHAPTER I
PROBLEM STATEMENT, BACKGROUND, AND
INTENTIONS OF THE STUDY

Problem Definition

The study described by this report concerns the use and effectiveness of an experimental program used by seventh grade under-achievers in the Columbus Public Schools during the first semester of the 1970-1971 school year.

The experimental program deals with counting and the four basic operations on non-negative integers. The program's distinguishing characteristic is its employment of an anecdotal style in content presentation. The anecdotal style is intended to serve both as a motivational device for pupils and as a means of facilitating the translation of certain learning and teaching principles into classroom practice.

Definition of the Target Population

The use of the experimental program is by pupils assigned to seventh grade "modified" classes. Seventh grade modified classes evolve from the first broad-scale use of "tracking" by the Columbus Public Schools. The primary consideration in assigning a pupil to a modified
class is his achievement level in reading. Generally, pupils two or more years behind chronological age in reading achievement level are assigned to such classes. Unless the pupil has demonstrated near-normal achievement in mathematics during the sixth grade, the assignment based on reading levels also places the pupil in a modified mathematics class. In addition, pupils with a history of low achievement in mathematics only may be placed in modified mathematics classes, while remaining in the regular sequence of studies for other content areas.

Some junior high schools in the Columbus System provide "special modified" classes for pupils with records of extremely low achievement. One such class is included in the sample of eight classes participating in the study. The other seven classes are described by the preceding paragraph.

It should be mentioned that the descriptions offered for "modified" and "special modified" classes were gained by talking with both administrators and teachers of the schools participating in the program. As is almost necessarily the case, given differences from school to school, no two of the classes participating in the program appear to have been defined by exactly the same set of criteria. (For example, if a given school's pre-registration plans provided for two modified sections, there were just two sections, with the result that the set of "under-achievers" in that school was
Rationale for Concern and Semantic Considerations

Regarding innate abilities and influences of the environment inherited by birth, it is evidently not the case that all men are equal. It is also the case that any instructional scheme devoted to making men equal by such connotations for the term is likely doomed to frustration,

or to a search for measurements which guarantee equality in exchange for a capacity to predict success in a given society. Briefly, with no aid from the test tubes of Brave New World, there appear to be alphas, betas, etc...relative to almost any criterion society is willing to honor. Claims for "equality", on the other hand, seem usually dependent on a narrow set of criteria, on criteria not honored by society's practices, or on an attempt to view man through the eyes of God.

However, the Brave New World analogy is short-lived. The phenomenon in that setting is that, almost without exception, alphas aspire to being alphas, betas to being betas, and so on. Beyond this, where any type of "human" was, there he was needed; and where he was needed, there he was. There were few "under-achievers", using the term either in relation to societal demands or in relation to the individual's potential,...and those few could be easily removed.

By either of the above criteria,...societal demands or the individual's potential,...contemporary American society
does have under-achievers, most noticeably in relation to societal demands of the individual.

Since "potential" is usually defined by one or several of many possible measures and is consequently in danger of putting yet another unattainable demand on the individual, "under-achievement" in this context will be used in the limited sense of a failure to meet societal demands. In other words, the "F-student" who makes "F" is an "under-achiever" and the "A-student" who makes "B" is likely not. A second semantic consideration concerns a choice between "under-achiever" and "low-achiever". In general, "under-achievement" connotes at least a hope that the achievement level demonstrated by the individual is below the possible, while "low-achievement" may or may not connote such. Since any serious attempt to raise achievement levels is premised on such hope, the terms "under-achiever" and "under-achievement" will be used in the following pages.

Agreeing that, to some extent anyway, the American position is that the societal structure must serve the individual, it becomes a moral concern whether or not the society may demand the impossible of an individual. If the answer is no, then the society has the obligation to respond to such occurrences in one of two ways. It can either adjust its demands to suit the individual, or it must change the individual to a point of his meeting the demands.

The above problem gains in complexity when widely-held
"values" are admitted as "demands". The naive position would be to divorce the two,... arguing that the high school dropout who finds daily food and shelter has no complaints. The more troublesome position would be to relate the two,... arguing that the would-be plumber not admitted to the union for want of a high school diploma has been dealt a foul blow, since somehow society allowed him to value becoming a plumber without taking care to see to the prerequisites.

The position taken in this context is that "values" do serve as or imply "demands". Further, the role of the public school is seen to include both establishing values (which may set demands) and equipping the individual to meet societal demands, both those demands peculiar to the school experience and those established by the broader community. Finally, where it is within the school's power to be an influence, both the setting of unattainable goals (demands) and the failure to equip pupils to meet existent demands are seen as wrongs.

By the above definitions and remarks, pupils in seventh grade modified classes may be thought of as "under-achievers". Except for a few pupils insensitive to the reward and report systems upheld by the schools, they are reminded of this fact by grade cards, achievement scores based on the "average" pupil's performance at a given age, differential treatment by teachers and peers, and by placement in special classes, each of which has its subtle (sometimes not so subtle) way
of establishing demands.

Beyond being an "under-achiever" relative to school demands, the broader community often includes as one of its requirements a certain level of performance in school, ....sometimes as an arbitrary screening device and sometimes with a view to the demands of a specific task. In either case, the "under-achiever" may be such on two counts.

Since the school's ability, assuming it had the right, to control societal demands appears questionable; and since the society now known shows few indications toward lowering its demands of the individual; it would appear to be the moral obligation of the school either to help each individual meet those demands or to know that the task is impossible. The position taken in this context is that efforts to raise achievement levels of under-achievers are required by that obligation.

Arguments like the above are unavoidably convincing or not convincing depending on the reader's views of the function of the school and of the societal structure. A second attempt to justify the efforts expended in this study may be made by appealing to federal government priorities over the 1960's. Cautiously assuming that the federal government's allocation of funds speaks the concerns of the society it serves, the federal funding of certain Title III programs of the 1960's may be cited as evidence of a general concern for the problem of under-achievement in mathematics.
These programs concentrated attention on the development of classroom materials and/or the development of in-service programs for teachers of under-achievers. Two such programs that focused attention on both these areas were the one operated from West Palm Beach, Florida serving counties in southern Florida and the one operated from Des Moines, Iowa serving an area in south-central Iowa.

Other federal allocations of funds which would indicate similar concerns are those in support of Upward Bound Programs, Head Start Programs, and Job Corps programs, which emphasized competencies in mathematics and reading as the two basic prerequisites to employability.

A final appeal for concern may be made by directing attention to the cost of welfare programs. While he was Secretary of Labor, W. Willard Wirtz stated that "each dropout will cost taxpayers $1000.00 per year for the rest of his life." (This statement sets an average figure, lowered by those dropouts fortunate enough to remain off welfare.)

There is evidence that pupils like those of seventh grade modified classes in the Columbus Schools are potential dropouts. A study directed by Dr. L. F. Cervantes of St. Louis University, drawing samples from six major U.S. cities, suggested that being "two years behind in reading or arithmetic at the seventh-grade level" was a common characteristic of the dropout.

Granting that the pupils identified for this study are potential dropouts, concentrating attention on the seventh
grade offers the advantage of dealing with such individuals while still at an age of required school attendance. If some few of these would-be dropouts can be influenced to remain in school by a given program, the program may stand justified by such mundane considerations as dollars and cents.

**Materials Intended for "Under-achievers" in Mathematics**

No attempt to list all materials alleged to be effective with under-achievers in mathematics will be made in this context... A humble attempt by the Des Moines CILAMP Project to collect such materials resulted in over six large boxfuls in less than three months.

A perusal of such materials would tend to indicate that they differ from the more traditional textbook in one or several of the following respects:

1. The content and/or vocabulary are "toned down."
2. A non-traditional means of presentation is employed.
3. There is a heavy emphasis on "variety."
4. The presentation is "small-frame programmed."
5. The packaging of materials is unique.

The SMSG "M"-series may serve as an example for materials in which "the content and/or vocabulary are 'toned down'." Common characteristics of such materials include slower pace, less difficult vocabulary, less insistence on abstract concepts, and less sophisticated symbolism. The effectiveness of such materials probably stems as much
from their tendency to slow down the teacher (thereby allowing fewer pupils to "get lost") as from any qualitative differences compared to a standard text.

The unit written by Eldert Groenendyk, Experimental Ninth Grade Mathematics, may serve as an example in which "a non-traditional means of presentation is employed."
The unit makes heavy use of flow-charting, both as a means of presenting mathematical content and as a technique for helping pupils clarify their own notions about mathematical procedures. (The unit is also one in which "there is a heavy emphasis on 'variety'.")

The "LAMP" booklet written by J. T. Zimmerman for use in the Des Moines Public Schools may serve as an example in which "there is a heavy emphasis on variety." Most such units are compilations of short learning tasks. They are frequently used in conjunction with a "math lab."

Although some teachers of under-achievers do suggest that "small-frame programmed" materials are effective with such pupils, most of these materials are not suitable to group use, but are, rather, useful to some pupils, working individually, and enjoying reading achievement levels higher than would typify the seventh grade under-achiever.

The booklets produced through the Title III Program operated from West Palm Beach may serve as an example in which "the packaging of materials is unique." Topics which would ordinarily be presented under a single cover are presented by short booklets. In a sense, such efforts might
be classified as emphasizing "variety".

(It should be mentioned that naming a particular work as an example for a single characteristic is not intended to imply that the characteristic named does justice to describing the work. More often, the characteristic named is one of several that might be mentioned.)

In addition to seeing the development of printed materials, the 1960's saw an emphasis on the notion of a "math lab." A study of several settings defined as "math labs" would indicate that, as a common feature, the "math lab" is a room used for mathematics class and containing at least some hardware, calculators, manipulative devices, games related to mathematical concepts, and so on. The effectiveness of the "math lab" appears heavily dependent on teacher skills. It is perhaps this recognition that moves some mathematics educators to refer to the "math lab" as an approach to teaching, along with a conspicuous hesitance to define the setting in physical terms.

Two characteristics of the flurry of attacks on the problems of under-achievement seem to prevail, neither contributing to the research respectability of such ventures. First, most evaluation of materials and methodologies is in the form of informal teacher reactions to their use. Often, those who speak loudest are heeded. Second, many such materials or methodologies which gain the endorsement of an "expert" are taken seriously and not
questioned as to their effectiveness.

It may be that such neglects of careful evaluative techniques are a near-necessary condition to the early development of materials, since it is certainly the case that there must be something to evaluate before evaluation can take place. A second possible explanation is that "the stakes are very low" in the eyes of many teachers and administrators when it comes to trying new programs with under-achievers....If present programs are not effective, there is really very little to lose by experimenting a bit with something new; anything at all is likely to show up as well.

A final concern to the use of new (different) materials which must be mentioned is the possibility of the "newness" being the prime (perhaps the only) factor contributing to apparent success of the materials. In such cases the materials do not necessarily stand condemned. It is rather the case that the expedience of their use becomes dependent on the nature of pupils' prior experiences in school.

The greatest danger in failing to recognize the possible effects of noticeably differential treatment is that the "new" materials (or methods) are likely to be prescriptive to earlier and/or later curricular and instructional decisions, thereby forcing the loss of their being "new", and consequently the loss of effectiveness. It may be that tendencies toward "fadism" are rooted in just such failures to
seek the conditions for and limits of a given technique's effectiveness. Relating the above to what might be an instance of the Hawthorne effect, it would be the error of painting every room in the factory a permanent orange upon observing that production from one room increased when its green walls were painted orange.

A Brief Description of the Experimental Program Used For Purposes of This Study

The experimental package includes three items: a narrative, a student workbook, and a teacher's handbook. Both the narrative and the workbook are divided into five sections, a section on counting and a section on each of the four basic operations with natural numbers.

The narrative imbeds the mathematical content in a fictitious historical account of the discovery of non-negative integers with the four basic operations. The detailed, step-by-step presentation aims neither to assume nor demand a student's facility for mathematical discovery. The pacing and sequencing of mathematical content are similar to what might typify a small-frame programmed presentation.

Each section of the student workbook includes a set of elementary exercises followed by optional exercises, some dealing with different number bases. Where the student's lack of competence with basic skills and understanding seems acute, the optional exercises may be ignored in favor of
more drill. Where the student is better prepared, most of his time may be spent with the optional exercises.

The teacher's handbook includes suggestions on classroom methodology, answers to exercises in the workbook, and notes on discussion topics which are suggested by the narrative. The handbook also includes explanations for including what might appear to be "excess baggage" in the narrative and workbook.

The unit claims a relatively large number of consciously included variables. There are no provisions in the unit's design for an easy isolation of these variables; neither do any of the research hypotheses attempt to identify specific variables as playing key roles in any change in performance which might be observed.

The following list presents the primary goals of both the narrative and the student workbook. Implied by the listing are some of the "variables" referred to in the above paragraph.

1. Controlling "closure".
2. Controlling "pacing".
3. Avoiding anxiety-producing stimuli.
4. Encouraging emotional involvement.
5. Entertaining.
6. Dealing with the need for and danger in repetition.
7. Controlling ego-involvement.
8. Managing "variety" in presentation and in task.
9. Providing for individual differences.

Both the narrative and the student workbook are appended. The teacher's handbook, however, is not. The teacher's handbook is identical to the student workbook except that answers to exercises are recorded and pages concerning classroom methodology and suggestions are inserted. Contents of these inserts are treated implicitly and expanded in chapter II. The reader interested in a more complete description of the materials is invited to refer to both chapter II and to the appended materials.

A Brief Description of the Tests Used for Purposes of the Study

Two forty minute tests were administered to pupils participating in the program, a pretest and a posttest, one given the day before beginning the program and the other given the day after completing the program. Some sections of the two tests were identical, and others were designed to be equivalent.

Each test consisted of two main parts, a set of psychological scales and a set of mathematics sub-tests. Items used as psychological measures were selected from the NLSMA batteries, including some from each of the following NLSMA classifications: "easy vs. hard", "fun vs. dull", "actual self-concept", and "ideal self-concept".

Most items used for the mathematics sub-tests are also NLSMA
items. The mathematics part of the test is made up of three sub-tests: an "open sentences" test, a "computations" test, and an "algorithms" test. The psychological scales and the computation test appear in identical form on the pretest and posttest. The open sentences test and the algorithms test appear in equivalent forms on the two tests. (A more complete description of the tests, along with statements of considerations in their design, may be found on pages 65 through 69 of chapter III.)

The reasons for including the psychological measures are not ambitious. Their primary function, in the view of the investigator, is to serve a "watch-dog role", hopefully toward identifying any peculiarity that might be suggestive of further investigation.

Of the mathematics tests, the "open sentences" test is included more as a check for transfer than as a direct check for the effectiveness of the program, while the "computation" test and the "algorithm" test are intended to deal directly with the concerns of the program. (For further clarification, the reader is referred to pages 65 through 69 of chapter III.)

A Brief Description of the Sample Used for the Study

Eight classes, seven "modified" and one "special modified", used the experimental program. (Definitions for "modified" and "special modified" appear on pages 1 and 2 of this chapter.) The initial selection of schools available
for the study was made by the Department of Evaluation and Research of the Columbus Public School System. Of the eight classes, four were inner-city and four were outer-city.

A ninth inner-city class was tested as a control class. This class was taught by one of the teachers who also taught a class using the program, was judged by that teacher to be a near-equivalent to the class using the program, and dealt with the same topics as covered by the program over an identical time period.

(For details concerning the selection of the sample, the reader is referred to pages 62, 63, and 73 of chapter III.)

Choosing to work with classes of seventh grade "modified" classes does impose some rather severe limitations with respect to sampling. The following are among such limitations:

1. School to school variance in placement of pupils. Already hinted on page 63 of this chapter, the achievement levels of pupils in "modified" classes in any given school will depend in part on the number of such classes scheduled and in part on whether "special modified" classes are scheduled. (Scheduling "special modified" classes will tend to result in higher mean achievement levels for "modified" classes; scheduling too few "modified" classes will tend to result in lower mean achievement
levels for "modified" classes.) In addition to scheduling differences, there is school to school variance in the degree to which pupils are placed in modified mathematics classes on the basis of reading scores.

2. School to school variance in mobility policies. In some schools, pupils who show promise while in a modified class are advanced to a regular class as soon as the teacher's opinion suggests they are capable of the more difficult work, .... and pupils showing poor performance are placed in special modified classes. In other schools, the placement in modified classes is viewed as somewhat permanent. Policies also vary from school to school on the movement of pupils from regular classes to modified classes. At the one extreme, pupils are moved from one class to another almost as soon as there is reason to believe that their performance suggests a move. At the other extreme, pupils are given time to firmly establish that a move is almost demanded.

3. The impossibility of random assignment. Since each school has relatively few modified classes, the chances are high that any given pupil is in a specific modified class because his schedule of course work demands it. For that reason, even
where several modified classes are scheduled, it is still impossible to assign pupils to classes in a way that would insure that any one class is typical of modified classes in that school.

In summary, the inhibitors to ideal sampling procedures are as follows. First, school to school variance in placement of pupils and the impossibility of random assignment make questionable assuming that pupils from different classes represent the same population. Second, school to school variance in mobility policies is likely to cause variance in experimental mortality patterns from class to class and to affect the meaning of achievement gains from class to class.

An added problem to working with under-achievers is that such pupils often have records of high absenteeism. Also, such pupils are frequently members of highly mobile families. These factors further aggravate the problem of experimental mortality.

A final sampling difficulty is that the distinction between "inner-city" and "outer-city" is perhaps less valid for modified classes from schools so classified than it is for the schools' total populations. If achievement levels in mathematics do vary from "inner-city" to "outer-city", and if schools are not purely "inner-city" or "outer-city"; then "modified" classes in either setting are likely to draw mainly from the same group, "inner-city" pupils or "outer-
city" pupils, whichever group tends to show lower achievement levels.

In summary, the choice to deal with seventh grade underachievers is an invitation to impurities in sampling. The alternative position, however, which would be to allow the quest for elegant samples to largely dictate the available areas for concern, is unacceptable to the investigator. School problems, after all, often seem to give little heed to lending themselves to investigation by laboratory techniques.

A Brief Sketch of the Research Design

Four uses will be made of the data in an effort to determine the extent of and nature of gains that may be observable with the use of the experimental program. (A more complete description of these uses of the data may be found in chapter III.) The four uses may be named as follows: "The pretest-posttest comparison" study, "the experimental-control" study, "the comparison with NLSMA data", and the correlations with reading scores".

Formal hypotheses (those to be tested by statistical techniques) will be stated in just one of two forms, null or directional. Where the investigator anticipates no differences or where a substantial degree of uncertainty is present, the null hypothesis will be stated, identified by "Ho(--)". On the other hand, where differences are anticipated by the investigator, the directional hypothesis will
be stated, identified by "Hi(--)".

The Pretest-Posttest Comparison

This treatment of the data will attempt to determine whether or not changes from pretest to posttest are statistically significant. The following hypotheses will be tested for the psychological scales:

Ho(j): There will be no change over items "1" through "10" from pretest to posttest over class "j".
\[ j = 1, 2, 3, 4, 5, 6, 7, 8. \]

Ho(S,T): Pupils in "S" will show no change over "T" from pretest to posttest.
\[ S = \text{Inner-city, Outer-city}. \]
\[ T = \text{Items "1" and "2" ("easy vs. hard" items); Items "3" through "6" ("fun vs. dull"); Items "7" through "10" ("actual self-concept" items); Items "11" and "12" ("ideal self-concept" items); and Items "1" through "10".} \]

The following hypotheses will be tested for the mathematics tests:

Hi(j,T): Posttest scores will be higher than pretest scores for class "j" over test "T".
\[ j = 1, 2, 3, 4, 5, 6, 7, 8. \]
\[ T = \text{II, III, IV, Composite of (II, III, IV).} \]

Note: "II" is the "open sentences" test, "III" is the "computations" test, and "IV" is the "algorithms" test.
Hi(S,T): Posttest scores will be higher than pretest scores for pupils in "S" over test "T".
S = Inner-city, Outer-city.
T = II, III, IV, Composite of (II, III, IV).

Hi(T): Class means and medians will be higher on the posttest than on the pretest for test "T".
T = II, III, IV, Composite of (II, III, IV).

All hypotheses stated for "the pretest-posttest comparison" will be tested by the Wilcoxon's Matched-Pairs Signed-Ranks Test (Wilcoxon's T). A discussion of the choice of statistics relative to the limitations imposed by sampling is included in chapter III.

The purpose of "the pretest-posttest comparison" is two-fold. First, it is intended to identify areas in which there appear to be gains. Second, it will dictate the nature of comparisons allowable for "the comparison with NLSMA data", a comparison which will be limited to the mathematics tests. No statistically significant change from pretest to posttest will limit comparisons to describing achievement levels of pupils in the modified classes. Statistically significant changes from pretest to posttest will invite an attempt to quantify the observed gains (or loses).

The Experimental-Control Study

Two inner-city, modified seventh grade classes were identified for purposes of this study, both taught by the same teacher, both covering the same mathematical topics,
both in the same time period.

The "experimental" class was one of the eight classes used for the pretest-posttest comparison. The "control" class was a class considered to be equivalent to the experimental class by the teacher. The control class used a traditional textbook. (For more complete descriptions, see pages 73 through 77 of chapter III.)

The following hypotheses will be tested for the psychological scales for the experimental-control study:

Ho(j,T): There will be no change over "T" from pretest to posttest for class "j".
\[ j = \text{Experimental, Control.} \]
\[ T = \text{"easy vs. hard" items, "Fun vs. dull" items, "actual self-concept" items, "Ideal self-concept" items, and items } 1 \text{ through } 10. \]

Ho(P,T): There will be no difference between experimental and control groups relative to "T" on "P".
\[ P = \text{Pretest, Posttest.} \]
\[ T \text{ ranges as above.} \]

The following hypotheses will be tested for the mathematics tests for the experimental-control study:

Hi(j,T): Posttest scores will be higher than pretest scores for class "j" over test "T".
\[ j = \text{Experimental, Control.} \]
\[ T = \text{II, III, IV, COMposite(II, III, IV).} \]
There will be no difference between experimental and control groups relative to "T", on test "P".

\[ P = \text{Pretest, Posttest.} \]

\[ T = II, III, IV, \text{COMposite of (II, III, IV).} \]

Hypotheses of the types \( \text{"Ho}(j,T)" \) and \( \text{"Hi}(j,T)" \) will be tested by a Wilcoxon's T. Those of the type \( \text{"Ho}(P,T)" \) will be tested by a Mann-Whitney U.

Support of all the hypotheses by the data will indicate no distinguishable differences between the two treatments relative to the measures employed.

The Comparisons with NLSMA data.

As already mentioned, statistical significance of gains by classes on the mathematics tests from "the pretest-post-test comparison" will be considered license to attempt an interpretation of such gains by an appeal to NLSMA data. A lack of statistical significance will restrict such comparisons to being descriptive of seventh grade modified classes as found in the Columbus Public Schools.

No formal research hypotheses belong to the comparisons with NLSMA data, since the comparison is ex-post-facto in nature. (For further discussion of such comparisons, the reader is referred to pages 77 through 79 of chapter III.)

Correlations with Reading Scores

The final use of the data will also be ex-post-facto in nature,....an attempt to see possible relationships
between pupils' reading abilities and performance during the ten week period of the program. The two reading tests used are the "vocabulary" test and the "comprehension" test of the "California Comprehensive Reading Test", Level 3.

The following hypotheses will be tested for the correlations with reading scores:

\[ H_0(r_{pre-A,B} : r_{post-A,B}) : \] There will be no difference between the correlation of pretest "A" with "B" and the correlation of posttest "A" with "B".

\[ A = III, IV, COMPose of (II, III, IV). \]
\[ B = Vocabulary, Comprehension. \]

The reading tests will be given about mid-way through the ten week experimental program. (For further information, the reader is referred to pages 80 through 81 of chapter III.)

Summary

The four uses of the data may be summarized as follows.

"The pretest-posttest comparison" has as its primary function checking for the statistical significance of changes that may occur over the ten week period of the program's usage.

Without statistically significant changes, "the comparison to NLSMA data" serves the function of describing the sample. With statistically significant changes, "the comparison to NLSMA data" may serve both to describe the sample and to suggest substantive significance for changes.

"The experimental-control study" has the function of
providing an additional frame of reference. It might serve as a more useful frame of reference than the NLSMA data in case achievement levels of the modified classes are not near levels identified by the NLSMA data, or in case item-by-item responses by modified classes suggest their being from a population too unlike the NLSMA population to allow for meaningful comparisons.

"Correlations with reading scores" stands somewhat by itself relative to the other three uses of the data. Its main function will be to watch for any peculiarities that might be suggestive of further investigations.
CHAPTER II
THE EXPERIMENTAL UNIT

The experimental package includes three items: a narrative, a student workbook, and a teacher handbook. Both the narrative and the workbook are divided into five sections, a section on counting and a section on each of the four basic operations with natural numbers.

The narrative imbeds the mathematical content in a fictitious historical account of the discovery of non-negative integers with the four basic operations. The detailed, step-by-step presentation aims neither to assume nor to demand a student’s facility for mathematical discovery. The pacing and sequencing of mathematical content are similar to what would tipify a small-frame programmed presentation.

Each section of the student workbook includes a set of elementary exercises followed by optional exercises, some dealing with different number bases. Where the student’s lack of competence with basic skills and understanding seems acute, the optional exercises may be ignored in favor of more drill. Where the student is better prepared, most of his time may be spent with the optional
exercises.

The teacher's handbook includes suggestions on classroom methodology, answers to exercises in the workbook, and notes on discussion topics which are suggested by the narrative. The handbook also provides explanations for including what might appear to be "excess baggage" in the narrative. For example, what the hero in the story refers to as his "fix-the-answer-later" method for addition is included for the following reasons:

1. To prepare the student for the "carry number" method for addition.
2. To aid the student in understanding "borrowing" when he encounters subtraction.
3. To provide a reference for later work with measurement systems where changing readings to lowest denomination is required.

(See pages 203, 204, 205, and 225 of the appendix.)

Both the narrative and the student workbook are appended. The teacher's handbook, however, is not. The teacher's handbook is identical to the student workbook except that answers to exercises are recorded and pages concerning classroom methodology and suggestions are inserted. Contents of these inserts are treated implicitly and expanded by the remaining pages of this chapter.

The unit claims a relatively large number of consciously included variables. There are no provisions in
the unit's design for an easy isolation of these variables; neither do any of the research hypotheses attempt to identify specific variables as playing key roles in any change in performance which might be observed. The choice of variables to be included was made by looking at variables independently, without a research-based regard for interactions between variables. The decision to create a unit in such a way rests on the following considerations.

First, and perhaps foremost, the position taken in this context is that the limitations of time, financial resources, and experimental control advise against attempting a complex analysis of interactions between the variables incorporated by the unit. A shoddy treatment of such a challenge runs the risk of leaving behind no more than a few questionable conclusions, inviting a further study to gain opposing results, and then demanding a third study to serve as "tie-breaker."

Another consideration is the fact that where treatment does make a difference relative to a given test, the certainty with which statistical techniques can claim a difference is a function of sample size. An immediate consequence of this fact is that "no statistically significant difference" is perhaps more often a matter of insufficient sensitivity in design and instrumentation than a matter of no actual difference. Regarding the above, the experimenter's choice in this context is to "over-arm" for the
task at hand; that is, to include in one package many factors, each intended to increase achievement gains from pretest to posttest. The quest is for traditionally accepted levels of statistical significance, the license to speak further.

An alternative would be to limit the number of variables from the start. In its most extreme form, this would be tantamount to considering each variable independently, the complete ignoring of interactions between the variables. Any such procedure runs the risk of suggesting that incompatible "parts", each deemed "good" in isolation from the others, go into the making of the "whole". (This argument is valid, of course, only where the "whole" is pre-defined.) In addition, given the restrictions of sample size and the difficulties in testing "extreme" groups, limiting the number of variables runs the risk of falling short of statistical significance where there may actually be a difference.

In summary, the scheme behind the experimental treatment in this context is akin to that of going after a fly with a shotgun. The fly must be taken, if not by one or several of the projectiles intended for that purpose, then by powder burn. Discussions of gauge, magnum, and shot size may follow the autopsy, which may in turn be followed by more refined experimentation.

The following pages will outline the primary goals of
both the narrative and the student workbook. Implied by the listing are some of the "variables" referred to by preceding paragraphs. Each of the following will be discussed:

1. Controlling "closure".
2. Controlling "pacing".
3. Avoiding anxiety-producing stimuli.
4. Encouraging emotional involvement.
5. Entertaining.
6. Dealing with the need for and danger in repetition.
7. Controlling ego-involvement.
8. Managing variety in presentation and in task.
9. Providing for individual differences.

Some of the above appeal to principles from educational psychology and/or learning theory. The critical reader is almost sure to wonder if psychological considerations preceded the unit's design or if they were consulted after the fact. The answer:...a touch of both.

While some may decry the fact that psychological justifications are frequently "tacked on" after the fact, the position taken by the investigator is that it is no less noble to employ psychology's vocabulary in describing sound treatment than it is to employ its theory in prescribing treatment. To the keen perceiver, after all, a large part
of psychology would do no more than to efficiently verbalize the obvious... As evidence that psychologically sound practices may be displayed before (or without) regard for the formalized discipline, consider the six year old who displays an astute sense for Skinnerian "shaping" in teaching a younger cousin to tie shoes.

On the other hand, the formalized discipline, beyond being descriptive, seeks to be predictive. Regarding curricular decisions, the predictive capacity is perhaps as often useful in deciding what not to do as in deciding what to do. The difficulty is that the desired and result (the conclusion) is often stated first with the task of identifying an appropriate curricular scheme (premiss set) coming second.

Given an "impossible" end result, the deductive process could do no more than reject or fail to confirm candidate curricular schemes. Given a "possible" end result, the deductive process would most likely reject or fail to confirm some curricular schemes and confirm others. In either case, the process of defining a candidate curricular scheme is an inductive one while the process of checking its appropriateness is a deductive one....assuming, of course, the existence of an adequate theoretical base.

The main point of the above is that the deductive process is perhaps properly "after the fact": that is, after
the definition of a curricular scheme intended to attain a desired and prestated end. The concern over whether or not psychological principles are consulted "before the fact" has to do with the inductive process. It is not a concern to providing "justification" for a curricular scheme.

In the following pages, where a concern over psychological principles is implicit, the discussions deal mainly with defining desired end results and defining candidate curricular schemes. Since the concern is over the appropriateness of such schemes, the question of whether or not psychological considerations preceded the design is not seen as important.

Primary Goals of the Narrative

Controlling "Closure"

In-service teachers of under-achievers who participated in workshops sponsored by the Des Moines GILAMP Project made the following observation:...If it takes four weeks to teach such children to deal with feet and inches, it will likely take another four weeks to establish such facility for pounds and ounces, and another four weeks for pints and quarts. If there were any transfer at all taking place, one would expect that the time periods needed for similar learning tasks presented in sequence would display a decreasing order,...not remain constant.

The problem evidenced by the learner may be seen as that of establishing "closure" prematurely. It is not a problem
peculiar to under-achievers, but a problem chronic in under-achievers. After all, science teachers have marveled over how a bright pupil can leave all his knowledge of algebra behind in attempting to balance a chemistry equation. But with the bright pupil, a little prodding will likely end with the treasured gestalt "aha!" The under-achiever, on the other hand, seems far more persistent in dealing with learning as a host of isolated and unrelated bits.

As its quantity increases, such "bits and pieces" learning is more difficult to establish initially, more subject to forgetting, and more in need of repeated rehearsal. It may also be more difficult to establish because it lacks a frame of reference. Borrowing Ausubel's terminology, it is likely to be "meaningless" learning, or at best only slightly "meaningful" learning. (This fact is often especially difficult for mathematics teachers to deal with since they view meaning as a function of the subject matter. Taking Ausubel seriously insists that "meaning" is a function of the learner's encounter with the material.) Viewing it another way, the learning is difficult to establish because it lacks an ideational frame of reference. Former learning which could serve to help organize or subsume the new is either no longer present or is available to the new task. New learning, in that case, is initially encountered as rote learning. Even when it gains a degree of "meaning", the meaning is dependent on too narrow a set of internal
and external stimuli. Viewing the above in light of Gagne's hierarchy, the end result is learning just at or still short of "concept" learning.

It is reasonable to assert that rote learning is more subject to forgetting than "meaningful" learning. First, if new learning is "meaningful", it may be expected that it is gained more efficiently relative to time. This is particularly important to the school setting where group instruction tends to fix the learning time available as a constant. Given a fixed time interval, "meaningful" learning is likely superior learning. Second, rote learning may be less often reinforced by additional experiences because of its lower level of abstraction. In short, low-abstraction learning allows fewer experiences to serve as reinforcers.

It is not surprising, in light of the above, that teachers of under-achievers often find it necessary to "start all over" at the beginning of the school year. Nor is it surprising that such pupils "seem to forget everything over the summer vacation", "do perfectly well on the unit test and fail the semester final", or "get all mixed up if any new approach is used". Only learning that is very similar to the initial learning is likely to serve as reinforcement. Any deviation in presentation or task requirement is likely to yield "interference".

One intention in presenting mathematical content via a story is to gain added control over the learner's tendency
to establish "closure" too soon and too frequently. Many under-achievers appear to regard the last problem in a set of exercises, the sounding of the bell, or "all right, let's turn the page" as a mandate to reach "closure". A story, by maintaining a single plot until a closure point more compatible with the nature of the learning task is reached, may serve to abort tendencies toward premature closure.

The most naive use of the story format is to use chapter endings as suggestions for closure points. A slightly more subtle use of the story format to control closure is that of having the hero "search for a better method" without resting until something satisfactory is found. (The chapter on division most consciously uses this scheme.) Another scheme is to use introductory and transitional paragraphs as occasions for relating past learning to previews of new learning tasks, the very essence of establishing "meaning".

Now and then a teacher encounters a pupil who never seems to reach "closure", ....the type who utters mystical, excited profundities relating number, song, and the whirling of planets, ....and then soundly fails the Wednesday quiz on subtraction. No deliberate attempt was made in writing the unit to accommodate such behavior. If, by some chance, the strategy aimed at controlling for premature closure also relieves this opposite malady, only a feeling of good fortune will be claimed.
Controlling "Pacing"

Although research seems to indicate that the expedience of external pacing depends on individual student characteristics, a teacher faced with a class of under-achievers usually sets a single pace for the group as a unit. In view of the difficulty of managerial problems frequently presented by such classes, this approach is perhaps the most fruitful, given the one room, one hour, one teacher, and twenty-five pupils. (It is assumed that the narrative will be treated with the class as a group.)

One understandable tendency teachers of under-achievers often display is that of moving on to the next topic after the class has given superficial evidence that learning has taken place. It is understandable for at least two reasons. First, where review of formerly encountered material is involved, the class is likely to gain rapidly in a short period of time; not in understanding, but in performance of what was probably rote learning, in the first place, merely in need of rehearsal. Second, teachers of under-achievers often feel a pressure to get pupils to "catch up" with the performance level suggested by chronological age. Even teachers willing to acknowledge the rate at which under-achievers usually show gains are likely to feel impatience or even guilt over spending a third week on counting and addition.

Beyond the above is the problem of deciding whether or
not apparent gains in performance are superficial. An under-achiever may score seventy percent on a test over subtraction at the end of the sixth grade, score thirty percent over the same test at the beginning of the seventh grade, and again score seventy percent on the test one week later after a review of subtraction. He may be succeeding only on subtraction problems involving no borrowing or on problems where the tens place is reduced to provide more units, as was the case in the sixth, fifth, fourth, and third grades. The point is that the review in such cases may not place the pupil in a better position than he enjoyed at the end of the sixth grade, even though the teacher is likely to take the jump from thirty percent to seventy percent as indicative of a very substantial gain, which would likely be the case if the learning were new learning.

Relative to the above, the narrative spends as much time as it dares to spend on the early, more basic concepts. The main purpose of chapter one is to force the pupil to look at place value, a critical notion to "meaning" in subsequent tasks. The chapters on addition and subtraction move almost as slowly, at a pace which would almost surely insult pupils if there were no story to serve as a diversionary measure. Less effort is put toward slowing the pace in the chapters on multiplication and division, since the probability of rapid, superficial gains
is much lower for these topics.

Another problem, hopefully dealt with by pacing considerations, is that of extreme response patterns commonly observed in under-achievers. Low achievers frequently behave as one or the other of two extreme types,....those who refuse to offer responses and those who offer responses almost randomly, often at a rapid rate. (Some under-achievers oscillate between the two patterns.)

A poor reader who suddenly becomes frustrated and begins to utter a wide variety of words which are not even remotely suggested by the spelling is demonstrating the second response pattern. In mathematics, such a student is one who "completes" every algebra assignment, often by filling the assignment sheet with almost arbitrary permutations of the algebraic symbols at his disposal. This second response pattern may be described as "run-away trial and error"....with a marked emphasis on the latter. A common symptom of "run-away trial and error" is that the pupil is likely to pay more attention to the teacher's reactions to responses than to the stimuli intended to elicit a particular response. To make matters still more difficult, many teachers inadvertently reward such response patterns either by unknowingly providing evidence that the guessing is getting better or by simply stopping the pupil with an announcement of the right answer. As John Holt points out for cases like these, the session is over;.....
...the right answer is now known. The subsequent explanation is merely a formality to be endured with nods, smiles, and a wistfully intelligent look.

The other extreme, a refusal to offer responses, may also be rewarded inadvertently. It is very often met by a request from the teacher that the pupil at least try to accomplish the task. There is very little ego-threat in being asked to try, not nearly so much as in being corrected after making a wrong response.

Provisions for pupils who refuse to offer responses are more consciously imbedded in the student workbook than in the narrative, primarily by the inclusion of exercises which seek to minimize the level of ego-involvement. To the extent that the narrative can awaken an interest in such a pupil, it may serve to encourage his offering responses.

The technique intended to deal with "run-away trial and error" is the small-frame programmed manner in which the mathematical content is covered by the story. Especially in the early chapters of the narrative, the very moderate pace is an attempt to reduce frustration and consequently to reduce the tendency of the pupil to succumb to a round of uncritical multiple response. Assuming an effectiveness for the provisions aimed at preventing premature closure, the accompanying moderate pace, at best, might make it unnecessary for a pupil to respond without confidence.
As mentioned earlier, many low achievers oscillate between the two extreme response patterns treated by these paragraphs. This fact presents problems. It would be ruinous to push for responses if the pupil is about to discharge a volley of undisciplined guesses; and it would be futile to exercise patience while the pupil is enjoying a period of mental hibernation. Having to deal with several such pupils at one time further intensifies this problem. It is for this reason that the experimental unit chooses to pick a path somewhere between "push" and "patience". It is probable that a keen tutor, working with a single pupil, could speed the rate of learning considerably by pushing and by exercising patience at appropriate times. But working with a group of pupils, it seems more reasonable to focus on variables which allow for control of one extreme response type without inviting the other. For this reason the unit chooses to focus on ego-involvement and "meaning" to gain control.

Avoiding Anxiety-Producing Stimuli

It may be assumed,...or at least conjectured,...that many stimuli which are supportive or neutral to higher achievers are debilitating to under-achievers. Examples might include the bench outside the principal's office, a blackboard, any printed matter in ditto blue, any book over a hundred pages long, a grade card,...and so on. It is beyond the reach of any classroom program to establish the
bucolic atmosphere that may be required to completely avoid such stimuli, but some small efforts in the direction of such avoidance may be attempted.

The claim made for the narrative regarding such stimuli is that the appearance of a light-hearted story in a mathematics class is not likely to elicit the same feelings as would the appearance of yet another, perhaps thicker, standard textbook. In addition to the novelty in the format, the narrative intentionally avoids mathematical vocabulary which might demand relatively abstract understanding to insure proper usage. In general, the narrative avoids mathematical vocabulary wherever the writer believes the learning can be established as easily without it. One of the most extreme examples of such avoidance occurs in the chapter on division, where the conversation centers around the "little number" and the "big number under the bent line". Such descriptions are almost sure to cause mathematical purists to blush,.... but not many junior high under-achievers are such purists.

(Provisions for avoiding anxiety-producing stimuli are also made by the student workbook. These will be discussed in later paragraphs of this chapter.)

**Encouraging Emotional Involvement**

The narrative does not pretend to be a historically accurate account of the development of natural numbers. The choice for using a fictitious setting was made primarily because it allows for greater control of closure...
points and pacing. It also allows for some play with emotions, an ingredient which would have to be managed delicately with any attempt to preserve historicity.

One of the secondary goals of the narrative is to invite pupils to empathize with characters in the story, the hope being that more attention to content will be given by a pupil who can identify with one or several of these figures. For example, chapter one of the narrative sees the hero a defensive under-achiever on page 186, excited on page 187, abused on pages 187 and 188, up on page 189, down on page 190, and completely depressed by the middle of page 191. The idea behind presenting such a roller coaster ride through emotions is not based on the premise that all under-achievers behave as the hero does. Some do, but the fact is that many exhibit little more than apathy. The idea is rather that the under-achiever may see himself a time or two, along with an invitation to run the gamut of emotions if his taking mathematics seriously may demand his doing so.

The hero is sometimes given to extreme displays of emotion. Page 263 of the narrative sees him at his very worst, just after despairing over a division problem treated by repeated subtraction. Pages 264 and 265 find him at his best, having found "a better method" for dealing with such problems.

The hero is also a man who sometimes shows a reluctance to begin working when a new task presents itself. He
balks on pages 233 and 234 when the problem of multiplication first appears; again on pages 257 and 258 when he is asked to deal with a question that begins the metamorphosis toward "long division" techniques. Many under-achievers should feel one with the hero through these mopings and procrastinations. If such feelings for the hero are strong enough, they may carry the pupil along when the push for a better method replaces sullenness in the story.

For pupils who find it difficult to associate with the hero, there is the village musician,....a gentle, flower-child type, not given to violent emotions, nor to the ravings of ambition. On page 224 he helps the hero develop an attack for subtraction where regrouping is required. His help is accidental. The hero must explain "borrowing" to the musician after the musician himself made the comment that motivated the discovery. The same phenomenon occurs when the hero reaches his conclusion of commutativity for multiplication on pages 253 through 255. The musician's comments serve as a catalyst for the hero's reaching a discovery, and again the hero must explain the discovery to the musician. But this time the musician not only fails to gain an insight,....he completely loses the intent of the scientist's argument by concentrating on details in the example. Such behavior is known to at least some under-achievers.
The difference between the village musician and many under-achievers, hopefully a contagious difference, is that the musician maintains an interest in mathematics throughout the narrative, in spite of the several shortcomings he shares with under-achievers.

It should be emphasized that including provisions for the pupils' emotional involvement with the story is not a research-based decision. It is based on a hunch, the hunch that emotional involvement may raise interest and in turn change behavior. Such phenomenon seem to occur outside the mathematics classroom. For example, consider the young girl who suddenly begins blinking and walking differently because of idolizing a TV star, or the admirer of an auto racer who soon develops the traits of a Jehu, or, for the skeptic, consider the loyal consumer of TV serials.

Entertaining

Watching pupils' faces during classroom hours sometimes seems to suggest that school is a matter of investing the present in anguish in order to draw future joy. Whether or not school ought to be fun is a philosophical concern, beyond the scope of this discussion. However, if an appeal to precedent may be made, it would appear that American values are not adverse to a touch of hedonism. Fruit flavors for children's medicines and music for grocery shoppers and dental patients may serve as examples.

In general, where entertainment is not in opposition
to the task at hand, it will probably be approved. Beyond the argument that the most palatable packaging of the medicine may be deemed the best, one may hope that a more pleasurable experience will achieve a positive attitude shift. To whatever extent more positive attitudes enhance achievement, an additional "good" may be claimed.

**Primary Goals of the Student Workbook**

**Dealing with the Need for and Danger in Repetition**

A case could be made for the claim that under-achievers are often in need of much more repetition than are "normal" pupils. The claim could be made by appealing to the fact that under-achievers tend to treat a larger share of learning as rote learning; or it could be made by appealing to the fact that fewer past experiences are available to lend "meaning" to the new learning task. In any case, looking to classroom practices for dealing with under-achievers would suggest a strong commitment to repetition (drill).

On the other hand, the gestalt position would indicate a danger of narrowing, a "blinding" influence, as a result of too much drill. In a sense, to the extent that a given response becomes available to a rather specific set of stimuli through repetitive drill, it becomes less available to variations in stimuli presentations.

Recognizing both the above as valid, the resolution of the dilemma appears to call for an adequate number of moderately varied exercises. To be avoided are such practices
as assigning an additional thirty simple addition problems to a pupil who is already capable of working such problems very well. Some evidence that such considerations went into the making of the student workbook may be gained by a perusal of the following sets of pages: pages 299, 302, 303, 305, 308, and 309 for addition; pages 311 through 315 for subtraction, and so on. Pages black with drill exercises are excluded from the workbook, both because of the above consideration and because of the likelihood that they would serve as anxiety-producing stimuli. (The policy on drill exercises to supplement the workbook will be discussed in later pages of this chapter.)

**Controlling Ego-involvement**

John Holt's discussion of "strategies for failure" presents a powerful case toward suggesting that many of the extreme, debilitating behaviors in under-achievers result from perceived threats to ego. The case is powerful in spite of its lack of data displays and significance levels.

For the reader given to a heavier dependence on reported research results, "anxiety" will be defined as fear which is rooted in perceiving a threat to self-esteem. Among the respectably established symptoms of anxiety are: rapid trial and error behavior within a narrow set of responses; escape, often in the form of apathy; and the refusal to invest time in original thinking. These behaviors almost define a trademark for under-achievers.
Research is also reported which would indicate that anxiety will facilitate or debilitate learning, depending on the nature of the learning task. For example, positive correlations are reported between anxiety level and achievement for learning that is rote in nature; and positive correlations are reported between anxiety level and achievement where learning is considered "easy". On the other hand, negative correlations are reported between anxiety level and achievement where learning is considered to be difficult or where the learning task requires problem-solving behaviors.

In light of the above, it is not surprising that many teachers of under-achievers become convinced that the most effective teaching strategy is to emphasize drill, often by the ream. So long as anxiety levels remain high, this is, indeed, perhaps the only workable strategy. The sorrow, often unseen by common testing procedures, is that the learning is likely of a type highly subject to forgetting. It may also be learning which suffers the "blinding" effect mentioned earlier.

Also, assuming the validity of the above, it would appear ill-advised to simply change the nature of the learning tasks without taking anxiety levels into account. Embracing this assumption, the exercises in the student workbook aim at lowering anxiety levels, not by lowering standards for pupil performance, but by attempting to lower the degree of
ego-involvement.

One form of this strategy is an appeal to "safety in numbers" by including a substantial number of class discussion exercises. Such exercises appear with higher frequency when the learning is considered to be difficult or at a level of abstraction not readily lending itself to rote memorization. Examples of this strategy include the discussion of counting on page 288 of the student workbook, the discussion of place value and addition on pages 295 through 297, more discussion aimed at understanding place value on pages 308 and 309, and the discussion of division algorithms on page 337. (This listing is not intended to be complete.)

A second form of this strategy is that of placing the pupil in the role of the judge rather than the judged. Whether rightly or not, pupils seem more willing to accept and defend divergent response patterns when the task at hand is evaluation,....as do adults. Pages 288 and 289 employ this strategy by inviting pupils to discuss and observe "counting" as done by smaller children.

The problem in getting an under-achiever to deal with a topic like counting is two-fold. First, most of the pupils will know they can count as far as they please without difficulty. To talk about counting directly runs a high risk of insulting the pupil,....much as a group of high school grammar teachers might be insulted by a lecture on using capital letters for proper nouns. Insult, a vehicle
for ego-threat, is not likely to enhance learning. The second aspect of the problem is that many under-achievers are in dire need of a close look at counting and the notion of place value.... A possible solution, suggested by item eleven on page 289, is to invite pupils to theorize about the mental processes of others, a pastime enjoyed by all ages.

A milder form of camouflage is shown by exercises like those on pages 295 and 310 by which the pupils are asked to discuss certain methods of presentation as to potential effectiveness with younger children. Again, the primary intent of the exercise's is nothing more than to get the under-achiever to look at mathematical content. An even less subtle invitation to role-playing is extended by exercises like those of page 304, where the pupils are asked to correct and evaluate work done by contemporaries of the hero in the narrative. With such assignments there are "right answers" to be found. But it may be less painful to have failed to detect another's error than to have made the error. It may also be hoped that the pupil who often makes such error(s) as appear in these exercises will be able to ponder it painlessly because it is not his error.

A possible criticism of such "correct the error" tasks is that seeing the wrong approach may increase the probability that the wrong approach will be used upon encountering similar stimuli. For example, spelling tests which ask
the examinee to find incorrect spellings have been criticised on these grounds. While the criticism may appear valid for near-rote or rote learning tasks, the opinion taken in this context is that such is not the case where the appropriate response to a stimuli set is dependent upon the learner's possession of an abstract concept or principle. The opposite may be the case. For example, a mechanic who thinks in terms of oxygen-fuel ratios may actually learn (further clarify his notions) from encountering a carburetor peculiarly out of adjustment. Implicit to the above, of course, is that the pupil will be operating from some level of abstraction, however humble, by the time such assignments are reached in the workbook.

The strategy of inviting pupils to play roles is gradually abandoned as the mathematical content approaches multiplication. At this point the probability that a direct encounter with the content will be insulting becomes much lower. The strategy of providing "safety in numbers", however, is evident throughout the workbook.

Managing "Variety" in Presentation and in Task

The cry for "variety" has been sounded often and loudly, particularly where under-achievers are concerned. Characterizations of under-achievers have included: "they can't concentrate", "they have a very short attention span", and "they become bored much sooner than 'normal' pupils". There may or may not be truth in such pronouncements....One may wonder
upon seeing an under-achiever spend each period of the school day with the same low-yield novelette tucked behind the textbook for the hour; or upon seeing an under-achiever faithfully practice free-throws for extended periods each day. To make matters still more puzzling, there are college professors who exhibit all of the above alleged marks of under-achievement during the usual faculty meeting.....

The kinder conclusion is that a number of the alleged marks of under-achievement are largely a function of the individual's values and tolerances relative to a given situation.

To the radical right of variety-advocates are those who smile over classrooms heavy with sundry unrelated games and puzzles, with little regard for what constitutes appropriate content. The phenomenon is like that of a judge who holds court without considering the case at hand, or who sometimes holds court without a case at all. Perhaps a main reason for such unqualified commitments to variety is that much of the commotion which would otherwise be seen as managerial difficulty may be justified.....If a child is given to slamming his book closed every five minutes, simply change activities at such intervals, thereby making his actions normal and proper.

The position taken in this context is that variety must be judged in light of prestated, or at least prepondered, goals,.....where the mere display of activity is not an acceptable goal. A willingness to call "good" whatever learning
may result from a near-random flurry of activities is also withheld. Variety in task, then, is justified or condemned by the intentions for the course. Interjecting variety in task to control managerial problems is seen as symptomatic of defeat, even though it may appear to be the only available escape in some cases.

Variety in presentation, on the other hand, is viewed as sound policy whenever the goal for the course asks that the pupil reach some higher level of abstract thinking or where transfer is a critical concern. In effect, varying presentation is no more than providing for a wider range of stimuli sets to elicit the desired response, the prescribed strategy for teaching concepts. It might also be hypothesized that variety in presentation, if seen as such, will lower tendencies toward "premature closure:...while excessive variety in task will strengthen such tendencies.

The workbook treatment of subtraction may serve as an example of providing variety in presentation, an example which also intends to combat tendencies toward premature closure. The variety may be viewed by comparing pages 310, 311, 312, 313, 314, and 319.

In summary, the workbook displays rather extensive variety for two reasons. First, relative to a fixed task, providing variety is sound teaching practice. Second, it may be that descriptions of under-achievers offered by variety-advocates are accurate....With a touch of good
fortune, variety in presentation may erase a need to manage or "justify" classroom behavior by varying task.

**Providing for Individual Differences**

A knotty problem concerning individual cognitive styles is that recognizing the phenomenon is not adequate toward knowing how to accommodate it. First, there is the question of whether or not it is possible for one style (the teacher's) to know how to effectively deal with another's; and, if possible, whether the course content should be adjusted to the given cognitive style or whether the given style should be viewed as an entry to establishing a style more amicable to the teacher's perception of the course content. Second, there is the question of whether the pupil with an unusual style should be fed his preferred diet of stimuli or whether he should be harshly thrust into settings more likely to typify his remaining days in school and his post-school experiences.

The design of the workbook does not pretend to deal with the above questions. If the content choices which honor variety in presentation happen to be sound relative to the above, it falls under serendipity,...without being realized.

Less subtle connotations for "individual differences" include differences in achievement levels and in learning rates. The workbook, dependent on the teacher's skill, intends to accommodate such differences by two schemes.
First, where classes are significantly heterogeneous, the so-called "optional" exercises may be assigned differentially. For example, a more advanced group of pupils may struggle to add base twenty-two numbers with the remnants of a torn addition table while others work with base eight addition or continue work with base ten. (The base eight and base twenty-two exercises appear on pages 305 and 307 of the workbook.)

Second, a large number of class discussion exercises may serve to divide pupils as tutors and tutored without having the relationships made explicit. Again, the belief here is that classroom materials can only facilitate, not guarantee an effective dealing with individual differences. (As a possible example, item two on page 310 could split the class on an agree-disagree basis where the split may really be determined by levels of understanding.)

Providing for Pupil Interaction

Some (perhaps most) teachers have known the following experience:.....A pupil has been showing a remarkable inability to "catch on" to an idea or process. Every known method considered sound has been exhausted in attempts to make the point. The teacher, becoming desperate, begins behaving like an under-achiever, trying one poorly conceived strategy after another;....and then, after an extremely awkward or questionable exposition, the pupil suddenly professes complete understanding. Sure that
understanding couldn't result from the inferior exposition, the teacher asks the student to demonstrate his understanding,.....which the student does handsomely.

Parents of smaller children are often privileged to see similar improbabilities. For example, a mother may devote many fruitless sessions to getting her young son to secure a shoe lace;....only to see his crude peer, in less than a minute, bring fruition by instructing the young man to "make that string ram your other finger out of the hole by the bent strong."

The tendency of both the teacher and the mother might be to conclude that teaching is a despairingly mystical business. This tendency would be justified if it really were the last learning experience that accomplished the entire task. It is more likely that the rare last session merely served to evoke an insight or performance that was already very near.

The above statement does not, however, belittle the importance of the peculiar last session that saw the task completed....The mother's "sound" instructions might have been given over another two months before a single shoe was tied. The point is that the peculiar session, however inadequate toward the entire learning task, did cause the learner to focus on a critical aspect of the task at hand. Such "peculiar sessions" may be particularly valuable where the task is concept learning and where tradi-
tional teaching techniques have failed to see the task completed, often leaving the appearance that absolutely nothing was learned.

The belief which motivates the inclusion of several of the class discussion exercises is that "peculiar sessions" abound when under-achievers discourse things mathematical. An additional belief held in this context is that where a pupil still fails to understand place value by the seventh grade, the more traditional teaching techniques may be suspected of needing assistance. Since both the teacher and the textbook author are prone to teach as they were taught, part of the burden of providing "peculiar sessions" is placed on the pupils.

Where under-achievers are involved, peer interaction may provide several other advantages over teacher talk. First, and already treated by earlier paragraphs, class discussions offer "safety in numbers." Second, and also discussed in earlier paragraphs, teacher talk may be an anxiety-producing stimulus for some under-achievers.

It should be emphasized that the above discussion does not intend to suggest unconditionally that encouraging peer interaction is the most efficient scheme for establishing mathematical learning. Where new (unfamiliar to the learner) learning tasks are involved, peer interaction is too likely to be no more than noise. Beyond that, where learning has been well established by traditional techniques,
any subsequent "peculiar session" is likely to be just that,...peculiar,...as well as uninstructive.

Notes on Goals Stated for the Narrative

The suggestion (by format of presentation) that some of the stated goals for the unit belong to the narrative and others to the workbook is not intended to imply an absence of concerns common to both. The design of the workbook means to support goals stated for the narrative.

Efforts to control "closure" are quite conscious in the design of the workbook. The section on division provides an example of such efforts: On page 335 the "repeated subtraction" method is related to the "tall division" method by a class discussion exercise; on page 337 the "tall division" method is related to the "long division" method by another discussion; and on page 340 the three methods are involved simultaneously by questions concerning the number of subtractions required by each method. Beyond such exercises, both subtle and overt reminders to "seek a better method" are included in pages 334 through 344.

Efforts to avoid anxiety-producing stimuli are also conscious in the design of the workbook. Complete avoidance of potentially anxiety-producing stimuli is, of course, impossible. Just the same, even drill exercises may be adorned or camouflaged. One such example is exercise six on page 312 in which the instruction to carry out one of the hero's assignments to his pupils has no loftier goal
beyond providing a bit of drill. Another effort to avoid such stimuli is the intentional lack of large sets of drill exercises.

The two goals which belong most exclusively to the narrative are the control of pacing and the encouragement of emotional involvement. This does not mean that such considerations are foreign to the design of the workbook. The number and placement of class discussion exercises will certainly have an influence on pacing. Also, a number of the exercises in the workbook intend to capitalize on any affection that may have developed for characters in the story. One such example is the class discussion exercise on page 330 which speaks of the village musician, the hero, and his secretary.

Entertainment, if novelty may serve as such, is also intended by the workbook. An example of such intentions is offered by the discussions involving the "peg-wheel adding machine" described on page 309. Looking only at the mathematical content, the odometer of an auto might as well have been used to make whatever point may be made.

The Policy on Supplementary Exercises

For purposes of the study associated with the unit's use, the nature of supplementary exercises was restricted to drill involving the four basic operations. In almost all cases, such drill was administered by dittoed, teacher-prepared work sheets. On occasion, teachers had the pupils
take drill problems directly from a standard text.

The amount of drill, however, was not specified. Each teacher was encouraged to use his own judgment in deciding on frequency and quantity. Also, decisions whether or not to use workbook exercises labeled "optional" rested entirely with the teachers.

Pilot Development of the Experimental Unit

A narrative, sharing the goals stated for an anecdotal style by this chapter and developing the same two main characters as in the appended narrative, was tested in a rough first-draft, base eight form by several teachers of under-achievers in the Des Moines, Iowa area in 1966-1968. Both seventh and eighth graders used those materials. Though feedback was not formal, but rather by unstructured interviews with the teachers, the following tentative conclusions were drawn:

1. Pupils demonstrated a much longer attention span than teachers expected in cases where the story was read aloud in class.

2. The story stimulated livelier discussion (regarding mathematics) than is typical for classes of junior high under-achievers.

3. Leaving the task of reading the story to the pupils outside of classroom hours was nearly equivalent to ignoring the story completely.

4. Both teachers and pupils showed a remarkable,
almost passionate, resistance to investing many hours of time in the study of base eight.

5. Pupils did not, as was dreaded as a possibility, turn against the main characters in the story, but rather found them intriguing.

6. The level of humor, while admittedly banal in large part, was deemed appropriate to the task at hand.

A second trial run with the pilot narrative (in base eight) was made in 1968-1969 using two classes of ninth grade under-achievers in the Columbus Public Schools. The reason for a second trial, using older subjects, was to further check the possibilities that a class might turn on the story's main characters, or reject the level of humor as belonging to the romper set,...or both. Feedback from teachers supported each of the six statements listed above.

The primary conclusion from the pilot studies toward building the unit used in the present study was that the story's characters and the chosen level of humor would survive in most seventh grade classes for under-achievers. A second conclusion, that base eight should be abandoned in favor of base ten, appeared obvious,...even though working in base eight would seem to give the appearance that the content was new and would thereby require less adornment and camouflage in presentation.
A Comment on the Effect of "New" Materials

If the effect of "new" materials does influence achievement, then it probably will be a factor in the study involving the experimental unit. There is almost sure to be a pupil awareness that something about the mathematics course is unusual. For that reason, no opposing argument will be offered to any suggestions that research results are not clearly attributable to the thoughtfulness of the experimental package.

Nor, on the other hand, will apologies be made. The main purpose of the experimental unit is, after all, to see achievement levels raised,...and not to attribute any such rise to one or several of the variables that were consciously included in the design of the package. If it raises achievement levels by successfully employing the effect of "new" materials, it does so just the same.

Returning to the fly and the shotgun of page , the autopsy might conclude death due to pre-blast shock,... which is still preferred to seeing the insect once again depart the target area of a poised swatter.
CHAPTER III
RESEARCH DESIGN AND METHODOLOGY

This chapter is divided into four sections, one for each of four different uses of the data. Chapter IV is divided in like manner. Each description for a use of the data is followed by page references to chapter IV for analysis and conclusion.

The Pretest-Posttest Comparison Study

The Description of the Sample

Eight classes of seventh grade modified pupils were used in the study. Initial selection of schools available to the experimenter was made by the Department of Evaluation and Research of the Columbus Public School System. The request by the experimenter to use an equal number of "inner-city" and "outer-city" schools for the study was honored. No principal, nor any teacher who was approached with the materials, refused to participate in the program. While this fact does not guarantee a random sample of classes, it does offer research advantages over a sample of eight agreeable out of twenty approached.

Of the eight classes considered for the pretest-posttest comparison, seven were "modified" classes and one was
a "special modified" class. (Definitions for the terms "modified" and "special modified" may be found on pages "1" through "3" of Chapter I.) In general, such classes are made up of pupils two or more years behind in reading and/or mathematics achievement levels.

In addition to a probable lack of random sampling with regard to choosing classes for the study, there was also a lack of random assignment within schools. For schools with several modified classes, the decision on which one would participate in the program was left to the principal. In no school was there an attempt to assign randomly pupils to the several such classes,...not because principals were disagreeable to the notion, but because it was clearly a scheduling impossibility. (Pupils were in the sixth period section because of band,...in the first period section because of shop,...and so on.)

In summary, from a research purist's point of view, sampling was less than ideal,...a fact justified only by a lack of options.

In addition to the sampling problems already listed, it should also be mentioned that the criteria for placing pupils in modified classes appeared to vary from school to school.

A final problem contributing to less than ideal sampling is that policies on moving pupils to and from such classes varied from school to school.
The Choice of a Starting Time for the Program

Each teacher was encouraged to delay beginning the program until he was confident that the transition from summer vacation to school had been completed. In addition, each was encouraged not to begin until after a quick review of the four basic operations on non-negative integers. The reason for requesting the delay was to lower the probability that pretest results would be unrealistically low. In keeping with the requested delay, teachers began the program in the third, fourth, or fifth week of the Fall semester.

The Description of the Treatment (The Experimental Unit)

For an outline of the experimental unit, the reader is referred to pages 12 through 14 of chapter I. A more complete description may be gained from chapter II and a perusal of the appended unit.

The following ground rules were established for the use of the experimental unit:

1. The narrative must be read aloud in class, either by the teacher or by pupils. No section may be omitted.
2. Exercises not marked "optional" must be assigned, except in cases where the teacher is sure that pupils are already familiar with the material.
3. Additional work must be restricted to drill.
4. Classroom testing policies are left to the teacher.
No exact time period was set for completing the program. Each teacher was encouraged to use his own judgment in pacing the class. (As it turned out, all teachers completed the program in nine to eleven weeks. Since this variation in time was not considered significant by the investigator, all references to the time period for the program are given as "the ten week period" of the program.)

The Tests Used for Purposes of the Study

Two forty minute tests were administered to the seventh graders working with the program. The first test was administered on the day before work with the unit began and the second test on the day after the ten week program was completed.

Items were chosen from the NLSMA batteries on the basis of their being relevant to the concerns of the program's intentions and, regarding the mathematical items, on the basis of their being of a difficulty level believed to be sensitive to achievement levels of pupils participating in the program.

Both tests (pretest and posttest) consisted of two main sections, a section made up of psychological scales and a mathematics section consisting of three subtests. In the following pages, the set of psychological measurements will be referred to as "test I"; the "open sentences" test of the mathematics section as "test II"; the "computations" test involving the four basic operations on non-negative
Integers as "test III"; and the "algorithms" test as "test IV".

Test I (the psychological scales) appears in identical form in both pretest and posttest. The twelve items were chosen with an awareness of the reading levels typical of pupils in seventh grade modified classes. Using the classificatory scheme developed by NLSMA for the scales:

1. Items "1" and "2" fall under "easy vs. hard".
2. Items "3", "4", "5", and "6" fall under "fun vs. dull".
3. Items "7", "8", "9", and "10" fall under "actual self-concept".
4. Items "11" and "12" fall under "ideal self-concept".

The reasons for including the psychological measures are not ambitious. The primary function of test I, in the view of the investigator, is to serve a "watch-dog" role,....hopefully identifying any peculiarity that might be suggestive of further investigation. The main reason for a low commitment to test I is that pupils will be encountering their first experience with tracking, which might easily be the prime factor in any response patterns.

Test II (the "open sentences" test) appears in equivalent form in pretest and posttest. The reason for not giving identical items is that pupils are not likely to see many "open sentences" during the program, thereby increasing the probability of pretest-posttest contamina-
Items "1" through "6" are NLSMA items or equivalent to NLSMA items. Items "7" and "8" were added as a curiosity to the investigator, and were subsequently discarded when it was observed that they were correlated negatively with items "2" through "6". The reason for discarding items "7" and "8" is that almost all pupils who misread directions to the test gave correct responses to these items, while pupils who did well on "2" through "6" apparently refused to believe that the operation sign appearing with the open sentence might also be that of the correct response. (See pages 353 through 372 of the appendix for the tests used.)

The primary purpose for including test II is to provide some indication of transfer of learning. No gains or low gains on test II accompanied by strong gains on tests III and IV would suggest that pupils rely heavily on the format employed in presenting a problem, since the items of test II might be routine to the pupil who understands "checking" answers, one of the topics covered by the experimental unit. Test II, then, is not viewed as a direct check upon the effectiveness of the unit.

Test III (computation) appears in identical form in both the pretest and the posttest. Items "1" through "8", "11", and "12" are NLSMA items. The test includes three computations for each of the four basic operations. Items on test IV are of direct concern to the intentions of the experimental program.
Relative to test III, several teachers of under-achievers and several "experts" were asked to state an opinion as to whether more sensitivity would be lost by an attempt to construct equivalent items or by allowing for pretest-posttest contamination that might result from using the same items twice. The decision was unanimous that it would be ill-advised to attempt a construction of equivalent items, for the following two reasons: First, it is difficult to know which problems might be "equivalent" to a pupil barely in possession of so-called "basic facts". Second, the pupils will see enough problems like those of test III during the program to adequately "interfere" with any learning peculiar to the pretest.

Test IV ("algorithms") appears in equivalent form in pretest and posttest. Equivalent form tests are chosen for two reasons. First, some of the items are different from anything confronted during the program. Second, the amount of time spent on each item of text IV is possibly great enough to create some suspicion that appreciable recall of particulars may be carried from pretest to posttest.

Test IV is the only one of the tests used for the program that is taken intact from the set of NLSMA tests. To some extent it may be viewed as a check for transfer in that it contains items not treated directly by the program. On the other hand, its intended function is to check for pupils' comprehension (or understanding) of algorithms, a primary
concern to the experimental program.

It should be mentioned that the test does depend rather heavily on the pupils' ability to read directions. For that reason it is possible that a gain in performance on test IV would result from improvements in reading ability over the period of the program. No provisions are made for determining the extent of gain attributable to reading.

The two tests are appended (See pages 353 through 372). For "equivalent" tests, pages are marked as belonging to test (a) or (b). As a check against any failure to meet the condition of equivalence, half of the pupils were given form (a) as pretest and (b) as posttest, while the others were given (b) as pretest and (a) as posttest.

To help rid the tests of a dependence on pupils' reading abilities, the teachers were instructed to read aloud the introductory page to each sub-test.

The Administration of the Tests

Pretests were administered the day before pupils began the program. The teachers were instructed not to inform pupils until after the pretest that new materials were to be introduced into the class.

The posttest was administered on the day after the program was completed.

All tests were scored by the investigator. Results were available to the teachers after the program was completed, along with comments on their individual classes.
The Research Hypotheses

The following hypotheses will be tested for the psychological scales:

$H_0(j)$: There will be no change over items "1" through "10" from pretest to posttest over class "j".
\[ j = 1, 2, 3, 4, 5, 6, 7, 8. \]

$H_0(S,T)$: Pupils in "S" will show no change over "T" from pretest to posttest.
\[ S = \text{Inner-city, Outer-city}. \]
\[ T = \text{"Easy vs. hard" items, "Fun vs. dull" items, "Actual self-concept" items, "Ideal self-concept" items, and items "1" through "10".} \]

The following hypotheses will be tested for the mathematics tests:

$H_1(j,T)$: Posttest scores will be higher than pretest scores for class "j" over test "T".
\[ j = 1, 2, 3, 4, 5, 6, 7, 8. \]
\[ T = \text{II, III, IV, Composite of (II, III, IV)}. \]

$H_1(S,T)$: Posttest scores will be higher than pretest scores for pupils in "S" over test "T".
\[ S = \text{Inner-city, Outer-city}. \]
\[ T = \text{II, III, IV, Composite of (II, III, IV)}. \]

$H_1(T)$: Class means and medians will be higher on the posttest than on the pretest for test "T".
\[ T = \text{II, III, IV, Composite of (II, III, IV)}. \]
Methods for Analyzing the Data

All of the above hypotheses will be tested for statistical significance by the Wilcoxon's Matched-Pairs Signed-Ranks Test (Wilcoxon's $T$). The choice of a distribution-free test seems, in the opinion of the investigator, to be in order for such "extreme" groups as "modified" classes are presumed to be.

Neither means nor medians will be computed for the results of the psychological scales, for the following two reasons: First, the reasons for including the scales do not demand such information. Second, the effect of lifting such items from the NLSMA text context and the effect of pupils' low reading abilities are likely to render any comparison of results from this study to NLSMA figures meaningless.

Both means and medians will be displayed for the mathematics tests.

Interpretations of the Data

The pretest-posttest comparison study is capable of providing little substantive significance, even with statistical significance, for want of comparative data. At best, it can suggest that there were or were not differences between pretest and posttest results.

However, the primary purpose of the "pretest-posttest comparison" is to seek license to interpret gain scores by comparing them to data provided by NLSMA. If gains for the
mathematics tests are statistically significant, such license will be assumed. If not, the comparison of test results to NLSMA data can merely provide an indication of achievement levels of seventh grade modified classes in the Columbus Public Schools.

**Weaknesses in the Design**

The most glaring weakness in the pretest-posttest comparison is the lack of a firm claim to generalizability. Any claim that there is no abuse of statistics relative to hypotheses of types "\(H_0(j)\)" and "\(H_1(j,T)\)" is perhaps tantamount to insisting on eight small studies, each concerned with a unique population. Hypotheses of types "\(H_0(S,T)\)" , "\(H_1(S,T)\)" , and "\(H_1(T)\)" suggest that generalizability is a concern by the samples used. A claim of no abuse relative to these hypotheses is perhaps indefensible.

By itself, the pretest-posttest comparison also lacks comparative data, thereby rendering any gains uninterpretable. However, granted gains, this weakness may be answered by the comparison to NLSMA data.

The design is also vulnerable to condemnations relative to most of the common design deficiencies, with the possible exceptions of "no point of reference" and "unstable instrumentation". The determination of the degree to which such deficiencies may contaminate results is beyond the grasp of the study. Each reader is invited to reach his own decision. The investigator holds the opinion that the experimental
program will be the main factor in the case of statistically significant gains. (Data displays for "the pretest-posttest comparison" appear on pages 86 through 98 of chapter IV.)

The Experimental Control Study

The Description of the Sample

Two inner-city, modified seventh grade classes were used for purposes of this study, both taught by the same teacher.

The "experimental" group was one of the eight classes used for the pretest-posttest comparison. The "control" group was a class judged by the head of the mathematics department, who also taught both classes, to be a near-equivalent to the experimental group. (Talk of making random assignments was dropped almost at once, since variations in pupils' programs appeared to be a main factor in assigning them to one section of mathematics or the other.)

The Choice of a Starting Time

Starting time considerations were the same as those for the pretest-posttest comparison. (See page 64.) Both the experimental and control classes took the pretest on the same day and took the posttest on the same day.

The Description of the Two Treatments

The treatment given the "experimental" group is described by preceding pages of this paper. The "control" group covered the same mathematical topics, but by using a standard
textbook, one of the textbooks widely recognized as being appropriate to the needs of pupils like those of the two classes.

The Tests Used for Purposes of the Study

The pretest and posttest given both to the experimental and to the control group are described in earlier pages of this paper. The testing is identical to that of the pretest-posttest comparison. Both the "experimental" and "control" groups took the pretest on the same day and took the posttest on the same day.

The Administration of the Two Tests

The administration of the two tests is described by that of the pretest-posttest comparison.

The Research Hypotheses

The following hypotheses will be tested for the psychological scales for the experimental-control study:

Ho(j,T): There will be no change over "T" from pretest to posttest for class "j".

j = Experimental, Control.
T = "Easy vs. hard" items, "Fun vs. dull" items, "Actual self-concept" items, "Ideal self-concept" items, and items "1" through "10".
Ho(P,T): There will be no difference between experimental and control groups relative to "T" on "P".
P = Pretest, Posttest.
T = "Easy vs. hard" items, "Fun vs. dull" items, "Actual self-concept" items, "Ideal self-concept" items, and items "1" through "10".

The following hypotheses will be tested for the mathematics tests for the experimental-control study:

Hi(j,T): Posttest scores will be higher than pretest scores for class "j" over test "T".
J = Experimental, Control.
T = II, III, IV, Composite of (II, III, IV).

Ho(P,T): There will be no difference between experimental and control groups relative to "T" on "P".
P = Pretest, Posttest.
T = II, III, IV, Composite of (II, III, IV).

Methods for Analyzing the Data

Hypotheses of types "Ho(j,T)" and "Hi(j,T)" will be tested by a Wilcoxon's T. Hypotheses of type "Ho(P,T)" will be tested by a Mann-Whitney U. Neither means nor medians will be computed for the psychological scales. Both means and medians will be displayed for the mathematics tests.
Interpretation of the Data

If the data should support all of the above hypotheses, the suggestion would be that there is no difference between the "experimental" and "control" treatments relative to the questions asked of the data.

Weaknesses in the Design

The primary weakness in the design is the lack of random assignment of pupils to the two classes. As was the case with "the pretest-posttest comparison" study, an insistence that no statistical assumptions have been violated might deny both generalizability and the grounds for comparisons between control and experimental groups....Generalizability might be denied in that each group must be considered as a total population; comparisons might be denied in that no check is available for possible interactions between populations and treatments.

Again, each reader is invited to come to his own conclusions as to the seriousness of the weaknesses. The investigator, as before, holds the opinion that classes used in this study are adequately like each other and like other classes of under-achievers to allow for inferences both toward generalizability and toward comparisons of the experimental and control treatments.

The teacher who taught both classes was considered to be strong by the investigator and by one "expert" who was acquainted with his work. The investigator reached his
conclusions about the teacher by several sessions of eavesdropping, ... with his presence in the hallway not known to the teacher.

**The Comparisons to "NLSMA" Data**

**The Conditions Which will justify making Comparisons**

Statistical significance of gains by classes on the mathematics tests from "the pretest-posttest comparison" will be considered license to compare those results to data supplied by the NLSMA studies. In the event of statistical significance over inner-city and outer-city classes but not over individual classes, the investigator will see only a delicate invitation to make such comparisons, ... but they will be made just the same. In case no claim of achievement gains can be made by an appeal to the data from "the Pretest-posttest Comparison", no inference will be sought beyond an attempt to describe the status of pupils in seventh grade "modified" classes in the Columbus Schools.

**The Nature of the Comparisons to NLSMA Data**

Assuming that gains can be claimed for the pretest-posttest comparisons, the data available from the NLSMA study may be used toward providing a frame of reference, particularly in determining the substantive significance of such gains. For that reason, most items chosen for the two tests were those for which information was available from NLSMA usage with several grade levels.
In any case, with or without statistical significance for the pretest-posttest comparison, the NLSMA data may serve to identify specific weaknesses of pupils in seventh grade modified classes.

It should be mentioned that the choice of items with an eye to being able to make such comparisons involved a degree of guess-work. In general, items were chosen on the assumption that pupils in seventh grade modified classes are about two years behind in reading and in mathematics. (As mentioned earlier, this determination resulted from interviews with both administrators and teachers.)

The ideal situation would have been to have scores from standardized tests available in advance of making item selections. Such was not the case for the following reasons:

1. Seeking approval from the Columbus Public Schools to execute the program in the early Fall required its completion before the school year began.

2. School-to-school variations in assigning pupils to "modified" classes made it difficult to fix an achievement level as characteristic of such pupils without knowing which schools would be available.

3. Plans by a number of schools to schedule "special modified" classes for the first time seemed to advise against looking to the previous year's classes as descriptive of "modified" classes.

4. Such information, even if obtained, would have been
been derived relative to tests other than those used by NLSMA.

(The reader who suspects that the writer is preparing him for a jolt is correct. As it turned out, several of the "special modified" classes were dissolved early in the Fall with pupils being assigned to "modified" classes; even where there were "special modified" classes, pupils who belonged to such classes but who had scheduling problems were sometimes placed in "modified" classes; and, finally, pupils in "modified" classes appeared to be more than two years behind both in reading and in mathematics.)

Results from the psychological scales will not be compared to NLSMA data. The decision not to make such comparisons is for the following reasons:

1. It is likely that pupils encountering a first experience with "tracking" will respond in large part to that experience.

2. The eclectic nature of the item bank used is likely to influence responses.

3. The psychological scales are intended to serve no more than a "watch-dog" function.

4. Reading difficulties of pupils are likely to be a factor in response patterns.
Correlations with Reading Scores

The final use of the data will rather boldly abandon respect for checking against the assumptions of statistical techniques in an attempt to find possible relationships between pupils' reading abilities and performance over the ten week period of the experimental program.

The following hypotheses will be tested for the correlations with reading scores:

\[ H_0(r_{pre-A,B}, r_{post-A,B}) \]: There will be no difference between the correlation of pretest "A" with "B" and the correlation of posttest "A" with "B".

\[ A = III, IV, COMPosite of (II, III, IV) \]
\[ B = Vocabulary, Comprehension \]

The two reading tests are the "vocabulary" test and the "comprehension" test of the "California Comprehensive Reading Test", Level 3. The reading tests will be given about midway through the ten week period of the experimental program. Correlations will be computed by a Pearson "r". The statistical significance of each Pearson "r" will be checked by an F-test. The statistical significance of differences between "r's" will be tested by the Hotelling Formula for Correlated Correlations.

If hypotheses of type "\[ H_0(r_{pre-A,B}, r_{post-A,B}) \]" are supported by the data, it will simply indicate that evidence is non-existent or not strong enough to suggest that the pupils' reading abilities are a factor in responses to the
experimental program. If one (or several) of the hypotheses is rejected by the data, the experimenter will feel free to at least suggest further study regarding interactions between reading abilities and the effectiveness of the experimental program.

In any event, due to the irreverence for the assumptions that underlie the statistical techniques, results for "correlations with reading scores" will be suggestive at best. In addition, the following weaknesses prevail for the correlations with reading scores:

1. The narrative is read aloud in class with the experimental program. It is somewhat presumptuous to believe that "reading" and "listening" are the same.

2. There is some reading ability demanded by the mathematics tests. With only one reading test given midway through the program, it is impossible to determine if gains are in reading, in mathematics, or in both.

In preceding pages of this paper, the correlations with reading scores have been referred to as an ex-post-facto treatment of the data. There are two reasons for viewing them as such. First, there are serious design weaknesses. Second, the investigator lacks the grounds for predicting results.
CHAPTER IV
ANALYSIS OF DATA AND INTERPRETATIONS

Before a discussion of each of the four uses for the data, the reader should be acquainted with a rather outstanding feature of this study,...the degree of experimental mortality. Of the eight classes participating in "the pretest-posttest comparison" study, 54 subjects took the pretest and did not take the posttest, 43 subjects took the posttest and did not take the pretest, and 129 subjects took both the pretest and the posttest.

The table on page 83 of this chapter displays the nature of the experimental mortality. Classes (1), (2), (3), and (4) are outer-city classes. Classes (5), (6), (7), and (8) are inner-city classes.

In the table, classes are given by columns. The first pair of rows gives the number of subjects who took only the pretest and provides the mean score for those subjects on the pretest. The second pair of rows provides information for those subjects who took only the posttest. The third set of (three) rows gives the number of subjects who took both tests and provides pretest and posttest means for those subjects.
<table>
<thead>
<tr>
<th>Row Identification</th>
<th>Outer-city Classes</th>
<th>Inner-city Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>(1) (2) (3) (4)</td>
<td>(5) (6) (7) (8)</td>
</tr>
<tr>
<td>&quot;N&quot; for Pretest Only</td>
<td>5 2 5 8</td>
<td>7 11 6 10</td>
</tr>
<tr>
<td>Mean, Pretest Only</td>
<td>9.8 11.5 14.6 10.5</td>
<td>11.3 9.5 11.7 9.9</td>
</tr>
<tr>
<td>&quot;N&quot; for Posttest Only</td>
<td>5 2 7 4</td>
<td>5 8 4 8</td>
</tr>
<tr>
<td>Mean, Posttest Only</td>
<td>12.6 10.0 10.9 9.3</td>
<td>10.8 7.3 14.5 11.8</td>
</tr>
<tr>
<td>&quot;N&quot; for Both Tests</td>
<td>19 21 16 14</td>
<td>17 11 18 13</td>
</tr>
<tr>
<td>Mean for Pretest</td>
<td>11.5 14.7 9.9 10.5</td>
<td>11.6 9.8 13.8 12.1</td>
</tr>
<tr>
<td>Mean for Posttest</td>
<td>16.4 17.8 14.6 15.1</td>
<td>14.1 12.6 16.9 17.2</td>
</tr>
</tbody>
</table>

Note: See the preceding page for comments on the above.
Several factors contributed to the mortality patterns, and those factors varied from class to class. Classes (1) and (8) were influenced by changes in the daily schedule of classes that took place during the program. In addition, pupils were moved between classes (1) and (8) and "special modified" classes, and pupils were moved between regular classes and classes (1) and (8) during the program.

The main factor in the pattern displayed for class (3) was a movement of pupils from class (3) to regular classes accompanied by a movement of pupils from regular classes to class (3). Such movement of pupils to and from regular classes during the program was also a factor in patterns displayed for classes (4), (5), and (6).

A second factor which played a significant role in the mortality patterns was absenteeism, a trait more typical of poorer than of better students.

Pooling data for all subjects who took one or both of the two tests yields the following results:

1. The 54 who took only the pretest had a pretest mean of 10.76.
2. The 43 who took only the posttest had a posttest mean of 10.70.
3. The 129 who took both tests had a mean of 12.0 on the pretest and a mean of 15.8 on the posttest.

Little further mention will be made of the experimental
mortality in remaining sections of this chapter. It is, however, the opinion of the investigator that the nature of the mortality tends to contribute to less pronounced gains in achievement than might be expected with more stable class enrollments. The reasons for holding this opinion are as follows. First, it is likely that those pupils who would have shown the most remarkable gains over the ten weeks of the program were no longer available for the posttest. (The assumption is, of course, that pupils identified by teachers as belonging in regular classes would have done nicely on the posttest.) Second, classroom efficiency is rarely enhanced by having to take on new pupils mid-way through a rather "new" program that is sequential in nature.

The remaining pages of this chapter follow an outline identical to that of chapter III. Pages 86 through 98 provide data, statistical results, and interpretations for "the pretest-posttest comparison" study; pages 98 through 107 do the same for "the experimental-control" study; the "comparison to NLSMA data" appears on pages 107 through 126; and results for "correlations with reading scores" appear on pages 126 through 129.

Should the reader wish for an informal description for any of the classes, he may refer to chapter V, which compiles informal feedback by classes treated individually.
The Pretest-Posttest Comparison

The Psychological Scales

The following hypotheses were stated for the psychological scales for the pretest-posttest comparison:

Ho(j): There will be no change over items "1" through "10" from pretest to posttest over class "j".

j = 1, 2, 3, 4, 5, 6, 7, 8.

Ho(S,T): Pupils in "S" will show no change over "T" from pretest to posttest.

S = Inner-city, Outer-city.

T = Items "1" and "2" ("Easy vs. Hard" items); Items "3" through "6" ("Fun vs. Dull"); Items "7" through "10" ("actual self-concept" items); Items "11" and "12" ("Ideal self-concept" items); and Items "1" through "10".

Relative to the set of "Ho(j)" hypotheses, the data did not deny Ho(j) for classes (2), (5), (6), and (8). Contrariwise, classes (1), (3), (4), and (7) appeared to show higher "attitudes" on posttest than on pretest on items "1" through "10", at the following levels of statistical significance yielded by a Wilcoxon's T: 0.02 for class (1), 0.10 for class (3), 0.05 for class (4), and 0.02 for class (7).

Of the type "Ho(S,T)" hypotheses, the two most clearly
suggested by the data are for \( T = \) "ideal self-concept" and \( S = \) outer-city, inner-city. As is shown by the table on the following page, the data suggests rejecting most of the hypotheses of type \( \text{Ho}(S,T) \).

"\text{Ho}(\text{outer-city, } T)" is rejected for each of the following: \( T = \) "easy vs. hard", "fun vs. dull", "actual self-concept", and items "1" through "10". Each of these is rejected in favor of suggesting gains over the ten weeks of the experimental program.

"\text{Ho}(\text{inner-city, items } "1" \text{ through } "10")" is also rejected in favor of suggesting gains over the ten weeks of the program.

"\text{Ho}(\text{inner-city, } T)" is cautiously rejected for each of the following: \( T = \) "easy vs. hard", "fun vs. dull", and "actual self-concept". Again, each is cautiously rejected in favor of suggesting gains over the ten week period of the program.

Two probable factors toward lower levels of statistical significance for inner-city classes are as follows. First, there were fewer inner-city subjects than outer-city subjects. Second, as suggested by the informal reports by teachers, it is likely that abilities to read both the instructions and the test items varied between inner-city and outer-city subjects. (There is also a possible difference regarding attitudes toward attitude tests.)
TABLE 2
RESULTS ON PSYCHOLOGICAL SCALES FOR "THE PRETEST-POSTTEST COMPARISON" OVER OUTER-CITY AND INNER-CITY SUBJECTS

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Items</th>
<th>Higher of Pre-test, Posttest</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer-city</td>
<td>&quot;Easy vs. Hard&quot;</td>
<td>Posttest</td>
<td>.01</td>
</tr>
<tr>
<td>Inner-city</td>
<td>&quot;Easy vs. Hard&quot;</td>
<td>Posttest</td>
<td>.11</td>
</tr>
<tr>
<td>Outer-city</td>
<td>&quot;Fun vs. Dull&quot;</td>
<td>Posttest</td>
<td>.001</td>
</tr>
<tr>
<td>Inner-city</td>
<td>&quot;Fun vs. Dull&quot;</td>
<td>Posttest</td>
<td>.15</td>
</tr>
<tr>
<td>Outer-city</td>
<td>&quot;Actual Self-concept&quot;</td>
<td>Posttest</td>
<td>.05</td>
</tr>
<tr>
<td>Inner-city</td>
<td>&quot;Actual Self-concept&quot;</td>
<td>Posttest</td>
<td>.17</td>
</tr>
<tr>
<td>Outer-city</td>
<td>&quot;Ideal Self-concept&quot;</td>
<td>Posttest</td>
<td>(.38)</td>
</tr>
<tr>
<td>Inner-city</td>
<td>&quot;Ideal Self-concept&quot;</td>
<td>Posttest</td>
<td>(.82)</td>
</tr>
<tr>
<td>Outer-city</td>
<td>items &quot;1&quot; through &quot;10&quot;</td>
<td>Posttest</td>
<td>.001</td>
</tr>
<tr>
<td>Inner-city</td>
<td>items &quot;1&quot; through &quot;10&quot;</td>
<td>Posttest</td>
<td>.05</td>
</tr>
</tbody>
</table>

Notes: "Higher" in the above table is defined to be in agreement with the computational form for the Wilcoxon's T. As it happened, in all cases "higher" also identified the test with the higher mean.

The above table provides data for judging the hypotheses of type "H0(S,T)".
The Mathematics Tests

The following hypotheses were stated for the mathematics tests:

\( H_1(j, T) \): Posttest scores will be higher than pretest scores for class "j" over test "T".
\( j = 1, 2, 3, 4, 5, 6, 7, 8. \)
\( T = II, III, IV, COMposite of (II, III, IV). \)

\( H_1(S, T) \): Posttest scores will be higher than pretest scores for pupils in "S" over test "T".
\( S = \text{Inner-city classes, Outer-city classes.} \)
\( T = II, III, IV, COMposite of (II, III, IV). \)

\( H_1(T) \): Class means and medians will be higher on the posttest than on the pretest for test "T".
\( T = II, III, IV, COMposite of (II, III, IV). \)

Note: "II" is the "open sentences" test, "III" is the "computations" test, and "IV" is the "algorithms" test.

Table 3 on page 90 provides data toward accepting or rejecting hypotheses of type "\( H_1(j, T) \)" for outer-city classes, classes (1), (2), (3), and (4).

Table 4 on page 91 provides data toward accepting or rejecting hypotheses of type "\( H_1(j, T) \)" for inner-city classes, classes (5), (6), (7), and (8).

Table 5 on page 92 provides data toward accepting or rejecting hypotheses of type "\( H_1(S, T) \)" for both outer-city and inner-city subjects.

Levels of statistical significance given by the three
### TABLE 3

RESULTS ON THE MATHEMATICS TESTS FOR "THE PRETEST-POSTTEST COMPARISON" OVER INDIVIDUAL OUTER-CITY CLASSES

<table>
<thead>
<tr>
<th>Class-Test</th>
<th>Pretest Mean</th>
<th>Posttest Mean</th>
<th>Pretest Median</th>
<th>Posttest Median</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)-II</td>
<td>2.9</td>
<td>3.9</td>
<td>2.4</td>
<td>4.1</td>
<td>.01</td>
</tr>
<tr>
<td>(1)-III</td>
<td>7.0</td>
<td>9.0</td>
<td>7.3</td>
<td>9.8</td>
<td>.01</td>
</tr>
<tr>
<td>(1)-IV</td>
<td>1.6</td>
<td>3.4</td>
<td>1.4</td>
<td>3.6</td>
<td>.01</td>
</tr>
<tr>
<td>(1)-COMP</td>
<td>11.5</td>
<td>16.4</td>
<td>11.8</td>
<td>18.0</td>
<td>.01</td>
</tr>
<tr>
<td>(2)-II</td>
<td>4.2</td>
<td>4.7</td>
<td>4.7</td>
<td>5.4</td>
<td>.05</td>
</tr>
<tr>
<td>(2)-III</td>
<td>8.2</td>
<td>9.1</td>
<td>8.3</td>
<td>9.7</td>
<td>(.15)</td>
</tr>
<tr>
<td>(2)-IV</td>
<td>2.3</td>
<td>4.0</td>
<td>2.2</td>
<td>4.1</td>
<td>.01</td>
</tr>
<tr>
<td>(2)-COMP</td>
<td>14.7</td>
<td>17.8</td>
<td>14.9</td>
<td>19.0</td>
<td>.01</td>
</tr>
<tr>
<td>(3)-II</td>
<td>2.4</td>
<td>3.4</td>
<td>2.3</td>
<td>3.1</td>
<td>.05</td>
</tr>
<tr>
<td>(3)-III</td>
<td>6.3</td>
<td>8.5</td>
<td>6.0</td>
<td>9.8</td>
<td>.01</td>
</tr>
<tr>
<td>(3)-IV</td>
<td>1.2</td>
<td>2.8</td>
<td>.5</td>
<td>2.5</td>
<td>.01</td>
</tr>
<tr>
<td>(3)-COMP</td>
<td>9.9</td>
<td>14.6</td>
<td>9.5</td>
<td>14.5</td>
<td>.01</td>
</tr>
<tr>
<td>(4)-II</td>
<td>2.5</td>
<td>3.2</td>
<td>2.3</td>
<td>3.5</td>
<td>(.15)</td>
</tr>
<tr>
<td>(4)-III</td>
<td>6.1</td>
<td>8.6</td>
<td>5.5</td>
<td>8.8</td>
<td>.01</td>
</tr>
<tr>
<td>(4)-IV</td>
<td>1.8</td>
<td>3.2</td>
<td>1.8</td>
<td>3.0</td>
<td>.01</td>
</tr>
<tr>
<td>(4)-COMP</td>
<td>10.5</td>
<td>15.1</td>
<td>9.5</td>
<td>14.5</td>
<td>.01</td>
</tr>
</tbody>
</table>
TABLE 4

RESULTS ON THE MATHEMATICS TESTS FOR "THE PRETEST-POSTTEST COMPARISON OVER INDIVIDUAL INNER-CITY CLASSES"

<table>
<thead>
<tr>
<th>Class-Test</th>
<th>Pretest Mean</th>
<th>Posttest Mean</th>
<th>Pretest Median</th>
<th>Posttest Median</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5)-II</td>
<td>2.5</td>
<td>2.5</td>
<td>2.0</td>
<td>2.0</td>
<td>--</td>
</tr>
<tr>
<td>(5)-III</td>
<td>7.2</td>
<td>8.4</td>
<td>7.3</td>
<td>9.0</td>
<td>.01</td>
</tr>
<tr>
<td>(5)-IV</td>
<td>2.0</td>
<td>3.2</td>
<td>1.9</td>
<td>3.0</td>
<td>.05</td>
</tr>
<tr>
<td>(5)-COMP</td>
<td>11.6</td>
<td>14.1</td>
<td>12.0</td>
<td>13.7</td>
<td>.01</td>
</tr>
<tr>
<td>(6)-II</td>
<td>1.2</td>
<td>1.3</td>
<td>1.1</td>
<td>1.1</td>
<td>--</td>
</tr>
<tr>
<td>(6)-III</td>
<td>6.9</td>
<td>7.9</td>
<td>6.8</td>
<td>8.7</td>
<td>(.25)</td>
</tr>
<tr>
<td>(6)-IV</td>
<td>1.7</td>
<td>3.4</td>
<td>1.7</td>
<td>3.2</td>
<td>.01</td>
</tr>
<tr>
<td>(6)-COMP</td>
<td>9.8</td>
<td>12.6</td>
<td>9.8</td>
<td>13.6</td>
<td>.02</td>
</tr>
<tr>
<td>(7)-II</td>
<td>3.1</td>
<td>3.6</td>
<td>3.0</td>
<td>3.5</td>
<td>--</td>
</tr>
<tr>
<td>(7)-III</td>
<td>8.0</td>
<td>9.4</td>
<td>8.0</td>
<td>9.8</td>
<td>.02</td>
</tr>
<tr>
<td>(7)-IV</td>
<td>2.7</td>
<td>3.9</td>
<td>2.8</td>
<td>3.9</td>
<td>.05</td>
</tr>
<tr>
<td>(7)-COMP</td>
<td>13.8</td>
<td>16.9</td>
<td>13.7</td>
<td>17.0</td>
<td>.01</td>
</tr>
<tr>
<td>(8)-II</td>
<td>2.4</td>
<td>3.7</td>
<td>2.3</td>
<td>4.0</td>
<td>.05</td>
</tr>
<tr>
<td>(8)-III</td>
<td>6.8</td>
<td>9.1</td>
<td>6.8</td>
<td>8.3</td>
<td>.02</td>
</tr>
<tr>
<td>(8)-IV</td>
<td>2.8</td>
<td>4.4</td>
<td>2.8</td>
<td>3.3</td>
<td>(.15)</td>
</tr>
<tr>
<td>(8)-COMP</td>
<td>12.1</td>
<td>17.2</td>
<td>11.0</td>
<td>16.0</td>
<td>.01</td>
</tr>
</tbody>
</table>
TABLE 5
RESULTS ON THE MATHEMATICS TESTS FOR "THE PRETEST-POSTTEST COMPARISON" OVER OUTER-CITY SUBJECTS AND INNER-CITY SUBJECTS

<table>
<thead>
<tr>
<th>Group-Test</th>
<th>Pretest Mean</th>
<th>Posttest Mean</th>
<th>Pretest Median</th>
<th>Posttest Median</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(O)-II</td>
<td>3.1</td>
<td>3.9</td>
<td>2.9</td>
<td>4.0</td>
<td>.001</td>
</tr>
<tr>
<td>(O)-III</td>
<td>7.0</td>
<td>8.8</td>
<td>7.1</td>
<td>9.5</td>
<td>.001</td>
</tr>
<tr>
<td>(O)-IV</td>
<td>1.8</td>
<td>3.4</td>
<td>1.6</td>
<td>3.4</td>
<td>.001</td>
</tr>
<tr>
<td>(O)-COMP</td>
<td>11.9</td>
<td>16.1</td>
<td>11.9</td>
<td>16.3</td>
<td>.001</td>
</tr>
<tr>
<td>(I)-II</td>
<td>2.4</td>
<td>2.9</td>
<td>1.8</td>
<td>2.8</td>
<td>.03</td>
</tr>
<tr>
<td>(I)-III</td>
<td>7.3</td>
<td>8.7</td>
<td>7.3</td>
<td>9.1</td>
<td>.001</td>
</tr>
<tr>
<td>(I)-IV</td>
<td>2.3</td>
<td>3.7</td>
<td>2.3</td>
<td>3.5</td>
<td>.001</td>
</tr>
<tr>
<td>(I)-COMP</td>
<td>12.1</td>
<td>15.4</td>
<td>11.9</td>
<td>14.5</td>
<td>.001</td>
</tr>
</tbody>
</table>

tables are determined by a two-tail Wilcoxon's T. As mentioned in chapter III, the assumption of random sampling, a pre-requisite to use of a Wilcoxon's T has likely been violated. Such violation would be more a concern to the data of table 5 than that of tables 3 and 4. As mentioned earlier, the reader may decide for himself if such violations are severe enough to discredit the given levels of statistical significance as being suggestive relative to
the stated hypotheses. The investigator suggests that they are not.

"Hi(j,II)" is accepted for \( j = (1), (2), (3), \) and (8); is cautiously suggested for \( j = (4) \); and is rejected for \( j = (5), (6), \) and (7). In other words, classes (1), (2), (3), (8), and possibly (4) appear to have gained over the ten week period of the program relative to the "open sentences" test, while classes (5), (6), and (7) do not appear to have gained.

"Hi(j,III)" is accepted for \( j = (1), (3), (4), (5), (7), \) and (8); and is cautiously suggested for \( j = (2) \) and (6). In other words, all of the classes appear to have gained over the ten week period of the program relative to the "computations" test, although gains for classes (2) and (6) were not strongly convincing.

"Hi(j,IV)" is accepted for \( j = (1), (2), (3), (4), (5), (6), (7), \) and (8). That is, all classes appear to have gained over the ten week period of the program relative to the "algorithms" test.

"Hi(j,COMP)" is also accepted for each of the eight classes. That is, all classes appear to have gained over the ten week period of the program relative to the composite of tests II, III, and IV.

Each of the hypotheses of type "Hi(S,T)" is accepted. In other words, both outer-city subjects and inner-city subjects appear to have gained over the ten week period
relative to each of the mathematics tests, and relative to
the composite of those tests.

Hypotheses of type "H1(T)" demand that classes be used
as the units of sampling. By the rather humble "Sign Test", "H1(II)" is supported at the .05 level; and "H1(III)", "H1(IV)", and "H1(COMP)" are supported at the .01 level.
That is, means and medians are higher on posttest than on
pretest for each of the mathematics tests and for the com­
posite of the mathematics tests.

Conclusions and Suggestions

Relative to the psychological scales, included in the
testing to serve a "watch-dog" function, the investigator
concludes that the experimental program did not lower at­
titudes toward mathematics. The reasons for the very limited
use of the data include the following:

1. Pupils were encountering their first experience
   with "tracking", which might be a primary factor
   in any pattern displayed by the data.

2. If the effect of "new" materials can be a factor in
   response patterns to the psychological scales, it is
   likely to have been such in this study.

3. Reading levels of the subjects were such that they
   were likely a factor in response patterns.

4. The eclectic nature of the item set comprising the
   psychological scales used for this study may
invalidate any item-by-item comparisons to data provided by the NLSMA results.

The only peculiarity seen by the investigator is that response patterns to "ideal self-concept" items seemed independent of response patterns to the other three parts of the test. Although not supported at respectable statistical significance levels, the investigator suggests that "ideal self-concept" may be largely a function of teacher characteristics.

Relative to the mathematics tests, the investigator concludes that achievement levels rose over the ten week period of the program relative to each of the three sub-tests.

As pointed out earlier in this context, "the pretest-posttest comparison" is capable of little toward defining the substantive significance of any gains in achievement. However, statistical significance of such gains, claimed by the analysis of the data, serves as license to seek substantive significance for those gains by comparisons to NLSMA data. (Such comparisons are made by the third use of the data in this chapter.)

Among less solid conclusions suggested by the data from "the pretest-posttest comparison" are the following:

1. It would appear that gains on test II are less pronounced for inner-city subjects than for outer-
city subjects. It is likely that differences in reading skills account for differences in performance on test II. On the other hand, it is possible that there are qualitative differences in cognitive styles common to inner-city subjects as compared to those of outer-city subjects, and that these differences influenced response patterns to test II. (Test II was included in part as a check for "transfer" of learning.)

2. It would appear that the definition of a "modified" seventh grade class varies from school to school (Such suspicions were expressed by the early pages of chapter I.)

3. Distracters for test IV (the "algorithms" test) were extremely effective for the under-achievers participating in the program, almost astoundingly effective in the opinion of the investigator. Test IV included thirteen items, each asking for the choice of one-of-five responses. Had pupils responded by using well-oiled game spinners, the expected mean score would have been 2.6. As it happened, the mean over all outer-city subjects on the pretest was 1.8; the mean over all inner-city subjects on the pretest was 2.3. (Posttest means were 3.4 for outer-city and 3.7 for inner-city.)

Two examples may serve to make the point.
Of the 129 subjects, 9 responded correctly on pretest item "2" with "e", 43 responded to item "2" with "a"; 20 responded correctly on posttest item "2" with "e", 33 responded to item "2" with "a".

Of the 129 subjects, 8 responded correctly on pretest item "6" with "a", 34 responded to item "6" with "c"; 25 responded correctly on posttest item "6" with "a", 35 responded to item "6" with "c".

In view of the distracter effectiveness for test IV, it may be suggested that a graph similar to the one below would result if it were possible to plot understanding (x-coordinate) against performance as the number of correct responses on test IV (y-coordinate).

![Graph](image)

The most dreadful argument that could be leveled against the experimental program is that the gain in mean scores from pretest to posttest on test IV is actually indicative
of a loss in understanding. That is, a move in the positive direction relative to the y-coordinate could correspond to a move in the negative direction relative to the x-coordinate.

The investigator rejects that argument for two reasons. First, it is unthinkable. Second, posttest means are considered, in the opinion of the investigator, to be enough above the chance level to discard the argument.

The "Experimental-Control Study"

The Psychological Scales

The following hypotheses were stated for the psychological scales for the experimental-control study:

Ho(j,T): There will be no change over "T" from pretest to posttest for class "j".

j = Experimental, Control.

T = "Easy vs. hard" items, "Fun vs. dull" items, "Actual self-concept" items, "Ideal self-concept" items, and items "1" through "10".

Ho(P,T): There will be no difference between experimental and control groups relative to "T" on "P".

P = Pretest, Posttest.

T ranges as above.

Table 6 on the following page provides data for judging
TABLE 6
RESULTS ON THE PSYCHOLOGICAL SCALES FOR "THE EXPERIMENTAL-CONTROL STUDY" OVER THE EXPERIMENTAL AND CONTROL CLASSES

<table>
<thead>
<tr>
<th>Class</th>
<th>Items</th>
<th>Higher of Pre-test, Posttest</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>&quot;Easy vs. Hard&quot;</td>
<td>Posttest</td>
<td>.05</td>
</tr>
<tr>
<td>Control</td>
<td>&quot;Easy vs. Hard&quot;</td>
<td>Posttest</td>
<td>.10</td>
</tr>
<tr>
<td>Experimental</td>
<td>&quot;Fun vs. Dull&quot;</td>
<td>Posttest</td>
<td>.02</td>
</tr>
<tr>
<td>Control</td>
<td>&quot;Fun vs. Dull&quot;</td>
<td>Posttest</td>
<td>.15</td>
</tr>
<tr>
<td>Experimental</td>
<td>&quot;Actual Self-concept&quot;</td>
<td>Posttest</td>
<td>--</td>
</tr>
<tr>
<td>Control</td>
<td>&quot;Actual Self-concept&quot;</td>
<td>Posttest</td>
<td>.01</td>
</tr>
<tr>
<td>Experimental</td>
<td>&quot;Ideal Self-concept&quot;</td>
<td>Posttest</td>
<td>--</td>
</tr>
<tr>
<td>Control</td>
<td>&quot;Ideal Self-concept&quot;</td>
<td>Pretest</td>
<td>--</td>
</tr>
<tr>
<td>Experimental</td>
<td>Items &quot;1&quot; through &quot;10&quot;</td>
<td>Posttest</td>
<td>.02</td>
</tr>
<tr>
<td>Control</td>
<td>Items &quot;1&quot; through &quot;10&quot;</td>
<td>Posttest</td>
<td>.10</td>
</tr>
</tbody>
</table>

the hypotheses of type "Ho(j,T)". Table 7 provides data for judging the hypotheses of type "Ho(P,T)". Significance levels for table 6 are given by a two-tail Wilcoxon's T. Significance levels for table 7 are given by a two-tail Mann-Whitney U. The symbol "--" is sometimes used to indicate a low level of statistical significance.
TABLE 7

RESULTS ON THE PSYCHOLOGICAL SCALES FOR "THE EXPERIMENTAL-CONTROL STUDY" COMPARING THE TWO GROUPS, EXPERIMENTAL AND CONTROL

<table>
<thead>
<tr>
<th>Items</th>
<th>Test</th>
<th>Higher of the Two Classes</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Easy vs. Hard&quot;</td>
<td>Pre</td>
<td>Control</td>
<td>(.28)</td>
</tr>
<tr>
<td>&quot;Easy vs. Hard&quot;</td>
<td>Post</td>
<td>Control</td>
<td>(.20)</td>
</tr>
<tr>
<td>&quot;Fun vs. Dull&quot;</td>
<td>Pre</td>
<td>Experimental</td>
<td>(.55)</td>
</tr>
<tr>
<td>&quot;Fun vs. Dull&quot;</td>
<td>Post</td>
<td>Experimental</td>
<td>(.29)</td>
</tr>
<tr>
<td>&quot;Actual Self-concept&quot;</td>
<td>Pre</td>
<td>Experimental</td>
<td>(.58)</td>
</tr>
<tr>
<td>&quot;Actual Self-concept&quot;</td>
<td>Post</td>
<td>Control</td>
<td>(.29)</td>
</tr>
<tr>
<td>&quot;Ideal Self-concept&quot;</td>
<td>Pre</td>
<td>Experimental</td>
<td>(.44)</td>
</tr>
<tr>
<td>&quot;Ideal Self-concept&quot;</td>
<td>Post</td>
<td>Experimental</td>
<td>(.04)</td>
</tr>
<tr>
<td>Items &quot;1&quot; through &quot;10&quot;</td>
<td>Pre</td>
<td>Experimental</td>
<td>(.75)</td>
</tr>
<tr>
<td>Items &quot;1&quot; through &quot;10&quot;</td>
<td>Post</td>
<td>Experimental</td>
<td>(.82)</td>
</tr>
</tbody>
</table>

The number of subjects in the "experimental" group is 18. The number of subjects in the "control" group is 21.

As with the data provided for "the pretest-posttest comparison" study, and for some of the reasons given in that context, neither means nor medians are displayed for the responses to the psychological scales.
"Ho(experimental, T)" is rejected in favor of suggesting gains from pretest to posttest for T = "easy vs. hard", "fun vs. dull", and items "1" through "10".

"Ho(experimental, T)" appears to be suggested by the data for T = "actual" and "ideal self-concept".

"Ho(control, T)" is rejected in favor of suggesting gain from pretest to posttest for T = "actual self-concept".

"Ho(control, T)" is cautiously rejected in favor of suggesting gain from pretest to posttest for T = "fun vs. dull", "easy vs. hard", and items "1" through "10".

"Ho(control, T)" appears to be suggested by the data for T = "ideal self-concept".

Of hypotheses of type "Ho(P,T)" only one is rejected in favor of suggesting statistically significant differences between experimental and control groups, namely P = posttest and T = "ideal self-concept". In view of the fact that the experimental group scored higher than did the control group on both pretest and posttest, a claim of any substantive significance is questionable.

The Mathematics Tests

Regarding the "experimental-control study", the reader is reminded that the primary weakness in the design is the lack of random assignment of pupils to classes. (The discussion of weaknesses in design occurs in chapter III). In a sense, then, a claim of research purity implies that each class is being viewed as an entire population.
The following hypotheses were stated for the mathematics tests for the experimental-control study:

\( H_1(j,T) \): Posttest scores will be higher than pretest scores for class "j" over test "T".

\( j = \text{Experimental, Control} \)

\( T = \text{II, III, IV, COMComposite of (II, III, IV).} \)

\( H_0(P,T) \): There will be no difference between experimental and control groups relative to "T" on "P".

\( P = \text{Pretest, Posttest.} \)

\( T = \text{II, III, IV, COMComposite of (II, III, IV).} \)

Note: "II" is the "open sentences" test, "III" is the "computations" test, and "IV" is the "algorithms" test.

Table 8 on the following page provides data for judging the hypotheses of type "\( H_1(j,T) \)". Statistical significance levels are given by a two-tail Wilcoxon's T. The symbol "(E)" is used for "experimental" and the symbol "(C)" is used for "control".

Table 9 provides data for judging the hypotheses of type "\( H_0(P,T) \)". Statistical significance levels are given by a two-tail Mann-Whitney U. "Pr" is used to mean "pretest" and "Po" is used to mean "posttest". The word "higher" is defined to be consistent with the computational form for the Mann-Whitney U.

As with "the pretest-posttest comparison" study, both means and medians are displayed.
### TABLE 8

RESULTS ON THE MATHEMATICS TESTS FOR "THE EXPERIMENTAL-CONTROL STUDY" OVER THE EXPERIMENTAL AND CONTROL CLASSES

<table>
<thead>
<tr>
<th>Class-Test</th>
<th>Pretest Mean</th>
<th>Posttest Mean</th>
<th>Pretest Median</th>
<th>Posttest Median</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E)-II</td>
<td>3.1</td>
<td>3.6</td>
<td>3.0</td>
<td>3.5</td>
<td>--</td>
</tr>
<tr>
<td>(C)-II</td>
<td>2.4</td>
<td>3.0</td>
<td>1.7</td>
<td>2.7</td>
<td>--</td>
</tr>
<tr>
<td>(E)-III</td>
<td>8.0</td>
<td>9.4</td>
<td>8.0</td>
<td>9.8</td>
<td>.02</td>
</tr>
<tr>
<td>(C)-III</td>
<td>8.7</td>
<td>9.1</td>
<td>9.0</td>
<td>9.9</td>
<td>.05</td>
</tr>
<tr>
<td>(E)-IV</td>
<td>2.7</td>
<td>3.9</td>
<td>2.8</td>
<td>3.9</td>
<td>.05</td>
</tr>
<tr>
<td>(C)-IV</td>
<td>3.0</td>
<td>2.8</td>
<td>3.0</td>
<td>2.6</td>
<td>--</td>
</tr>
<tr>
<td>(E)-COMP</td>
<td>13.8</td>
<td>16.9</td>
<td>13.7</td>
<td>17.0</td>
<td>.01</td>
</tr>
<tr>
<td>(C)-COMP</td>
<td>14.1</td>
<td>14.8</td>
<td>13.8</td>
<td>15.6</td>
<td>--</td>
</tr>
</tbody>
</table>

"Hi(experimental, T)" is supported by the data for T = III, IV, and COMposite of the mathematics tests. It is questionable for T = II.

"Hi(control, T)" is supported by the data for T = III, and is questionable for T = II and for T = COMposite of the mathematics tests. It is not supported for T = IV, but is rejected in favor of suggesting no gain from pretest to posttest relative to test IV.
TABLE 9

RESULTS ON THE MATHEMATICS TESTS FOR "THE EXPERIMENTAL-CONTROL STUDY" COMPARING THE TWO GROUPS, EXPERIMENTAL AND CONTROL

<table>
<thead>
<tr>
<th>Test</th>
<th>Exp. Mean</th>
<th>Con. Mean</th>
<th>Exp. Median</th>
<th>Con. Median</th>
<th>Higher of Two Classes</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr-II</td>
<td>3.1</td>
<td>2.4</td>
<td>3.0</td>
<td>1.7</td>
<td>Exp.</td>
<td>(.37)</td>
</tr>
<tr>
<td>Po-II</td>
<td>3.6</td>
<td>3.0</td>
<td>3.5</td>
<td>2.7</td>
<td>Exp.</td>
<td>(.29)</td>
</tr>
<tr>
<td>Pr-III</td>
<td>8.0</td>
<td>8.7</td>
<td>8.0</td>
<td>9.0</td>
<td>Con.</td>
<td>(.20)</td>
</tr>
<tr>
<td>Po-III</td>
<td>9.4</td>
<td>9.1</td>
<td>9.8</td>
<td>9.9</td>
<td>Exp.</td>
<td>(.90)</td>
</tr>
<tr>
<td>Pr-IV</td>
<td>2.7</td>
<td>3.0</td>
<td>2.8</td>
<td>3.0</td>
<td>Con.</td>
<td>(.55)</td>
</tr>
<tr>
<td>Po-IV</td>
<td>3.9</td>
<td>2.8</td>
<td>3.9</td>
<td>2.6</td>
<td>Exp.</td>
<td>.05</td>
</tr>
<tr>
<td>Pr-COMP</td>
<td>13.8</td>
<td>14.1</td>
<td>13.7</td>
<td>13.8</td>
<td>Con.</td>
<td>(.72)</td>
</tr>
<tr>
<td>Po-COMP</td>
<td>16.9</td>
<td>14.8</td>
<td>17.0</td>
<td>15.6</td>
<td>Exp.</td>
<td>.10</td>
</tr>
</tbody>
</table>

"Ho(posttest, IV)" is rejected in favor of suggesting that the experimental group scored higher than the control group on posttest IV. (This rejection gains substantive significance from the indication that the control group scored higher than did the experimental group on posttest IV.)

"Ho(posttest, COMPosite)" is cautiously rejected in favor of suggesting that the experimental group scored higher
than did the control group on posttest COMPosite of the mathematics tests. (Again, this rejection may have substantive significance by the indication that the control group scored higher than did the experimental group on the pretest of the COMPosite of the mathematics tests.)

With varying degrees of confidence, "Ho(P,T)" appears to be suggested by the data for remaining values for "P" and "T".

Conclusions and Suggestions

Relative to the psychological scales, the investigator is of the opinion that no strong claims can be made for different response patterns to experimental and control treatments. The following, however, may be conjectured as hypotheses for some further study, blessed with larger "N's", more comprehensive tests, and more ideal sampling:

1. The experimental treatment is both more "fun" and considered to be "easier than more traditional treatment. (This suspicion is suggested by an inspection of the first four rows of table 6 and the first four rows of table 7.)

2. The experimental treatment tends to raise "ideal self-concept" more than does the control treatment; and the experimental treatment tends to raise "actual self-concept" less than does the control treatment. (This suspicion is suggested by an
inspection of rows five through eight of table 6 and rows five through eight of table 7.)

Relating the second conjecture to the results on the mathematics tests, a possible explanation for the differential effects of the two treatments on "self-concept" is that the pupils in the experimental group became more aware of what there was to know regarding "simple" mathematics, and were thereby convinced that the "actual" state of their knowledge was something less than "ideal";...while the pupils in the control group did not become aware that there was more to know than was already known, thereby allowing themselves the comfort of seeing as "ideal" the "actual" state of their knowledge. (In the opinion of the investigator, the above explanation seems a possibility from the suggestions in table 8 that the control group showed no gains on the "algorithms" test while the experimental group did show gains.)

Relative to the mathematics tests, the following appear to be suggested by the data:

1. The experimental treatment is more effective than the control treatment relative to test IV, the "algorithms" test.

2. The experimental treatment is not less effective than the control treatment relative to any of the three mathematics tests.

3. The experimental treatment is more effective than
the control treatment relative to the COMPosite of the three mathematics tests, ... mainly because of its effectiveness relative to test IV.

A possible threat to the claims made for the experimental treatment relative to test IV is that the drop in performance shown by the control group over the ten week period on test IV is actually indicative of a gain in understanding. (The reader is referred to the discussion of distracter-effectiveness on page 97 of this chapter to lend a touch of credibility to such a threat.)

However, the investigator dismisses the threat for two reasons. First, means by both groups on both pretest and posttest were at or above the chance level. An insistence that a drop in means may indicate a gain in understanding is therefore tantamount to claiming that the curve on page should show several minima points. The second reason for dismissing the threat is that such a claim is perhaps far fetched, ... and certainly not a concern to most reported research in which multiple-choice items are employed.

The Comparison with "NLSMA" Data

As mentioned earlier, statistical significance of gains by classes on the mathematics tests from "the pretest-post-test comparison" is considered license to attempt providing substantive significance to such gains by comparing them to data supplied by NLSMA. On the basis of results displayed
on pages 89 through 94, that license is claimed.

The following should be kept in mind relative to comparisons to NLSMA data:

1. The treatment period was ten weeks. It was begun after pupils had opportunity for a brief review of the four basic operations. (The intentions for this delay in beginning and the period for review were to avoid pretest results that might be heavily biased by easily recovered learning forgotten over the summer months.)

2. Testing periods reported by NLSMA are at least one year apart, as compared with ten weeks between pretest and posttest for the experimental program.

3. On the other hand, the posttest given in conjunction with the experimental program was given soon after the material treated by the test was covered in class, which may bias those results relative to the NLSMA data which were not derived with a concern for testing soon after a topic was covered by the subjects.

For reasons implied by the above statements, attempts to describe gains precisely in terms of grade level of performance will be generally avoided, and done cautiously when it might seem instructive to qualify gains by such comparisons.

It should also be kept in mind that the sample selected
for purposes of this study is an "extreme" group, while the sample(s) used by NLSMA are not. Relative to the "extreme" group, the following considerations are in order for any attempt to compare data from that group to data provided by NLSMA:

1. The high (and peculiar) mortality endured by this study is likely to affect gains from pretest to posttest. (A full discussion of the experimental mortality appears on pages 82 through 85 of this chapter.)

2. Since the sample used for this study is an "extreme" group, attempts to relate the two by looking at measures of central tendency is likely not in order relative to any of the three mathematics tests. Even if means for the two groups were the same for a given test, item-by-item response patterns would probably be quite different.

3. Finally, already previewed by page 79 of chapter III, the suggestion that pupils in seventh grade "modified" classes are about two years behind in reading and in mathematics appears to be in error if determined by grade level achievement scores reported with NLSMA data. As will become apparent by the following pages, pupils in the seventh grade modified classes appear to be more than two years behind relative to the NLSMA measures.
In summary, the preceding considerations tend to call for seer's powers regarding any attempt to make comparisons between results derived from this study and those derived by NLSMA. On the other hand, the position taken relative to "societal demands" upon the individual (outlined in chapter I) asks that the effort be made.....The more respectable attempt to grant substantive significance to gains, from a research point of view, is the "experimental-control study" already treated; but comparisons of "under-achievers" to other "under-achievers" fall short both of describing the degree of under-achievement and of pin-pointing the nature of the under-achievement.

Comparisons for Test II ("Open Sentences" Test)

As mentioned earlier, a primary reason for including test II was as a check for "transfer". "Open sentences" were not treated directly by the experimental program. However, pupils who have an understanding of "checking" work done in computations involving the four basic operations could see "open sentences" as equivalent to "checking", except that the candidate "right answer" is not supplied.

Since "the pretest-posttest comparison" least convincingly provided statistical significance for gains on test II (as compared to tests III, IV, and COMposite), the comparison of results from test II to NLSMA data will be largely limited to describing the status of the sample, with no
claims made for the treatments effectiveness.

Table 10 on the following page provides pretest and posttest percentages for correct responses over the items of test II for each of the following groups:

1. All 129 subjects participating in the program.
2. The 78 subjects participating in the program who did not show evidence of "mis-reading" directions on either pretest or posttest, and who showed scores of "2" to "6" on the pretest. ("Mis-reading directions" is defined to mean that the pupil responded to all items by naming the operation belonging to the operation sign appearing in the "open sentence". A more complete discussion of "mis-reading" may be found on page 153 of chapter V.)
3. The 57 subjects participating in the program who did not show evidence of "mis-reading" directions on either pretest or posttest, and who showed pretest scores of "3" to "6".
4. The NLSMA x-population, 4th grade, spring.
5. The NLSMA x-population, 6th grade, spring.
6. The NLSMA x-population, 8th grade, fall.

In table 10, the symbol "--" is used to indicate that no comparative figure is available. Entries give the percentages of pupils who responded correctly by naming the operation which would be used in completing the "sentence": 
TABLE 10
DISPLAY OF RESULTS FROM "THE PRETEST-POSTTEST COMPARISON STUDY" AND "NLSMA" DATA ON TEST II (OPEN SENTENCES)

<table>
<thead>
<tr>
<th>Test</th>
<th>Subjects</th>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
<th>Item 5</th>
<th>Item 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>All 129</td>
<td>89</td>
<td>53</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>Posttest</td>
<td>All 129</td>
<td>93</td>
<td>60</td>
<td>46</td>
<td>43</td>
<td>37</td>
<td>63</td>
</tr>
<tr>
<td>Pretest</td>
<td>The 78</td>
<td>90</td>
<td>79</td>
<td>47</td>
<td>45</td>
<td>46</td>
<td>67</td>
</tr>
<tr>
<td>Posttest</td>
<td>The 78</td>
<td>95</td>
<td>78</td>
<td>65</td>
<td>63</td>
<td>55</td>
<td>83</td>
</tr>
<tr>
<td>Pretest</td>
<td>The 57</td>
<td>91</td>
<td>88</td>
<td>63</td>
<td>53</td>
<td>63</td>
<td>81</td>
</tr>
<tr>
<td>Posttest</td>
<td>The 57</td>
<td>98</td>
<td>88</td>
<td>75</td>
<td>67</td>
<td>65</td>
<td>86</td>
</tr>
<tr>
<td>Gr. 4, Spr.</td>
<td>NLSMA</td>
<td>--</td>
<td>67</td>
<td>51</td>
<td>51</td>
<td>49</td>
<td>54</td>
</tr>
<tr>
<td>Gr. 6, Spr.</td>
<td>NLSMA</td>
<td>--</td>
<td>84</td>
<td>75</td>
<td>68</td>
<td>69</td>
<td>80</td>
</tr>
<tr>
<td>Gr. 8, Fall</td>
<td>NLSMA</td>
<td>--</td>
<td>87</td>
<td>81</td>
<td>71</td>
<td>74</td>
<td>82</td>
</tr>
</tbody>
</table>

The six items on test II are as follows:

Item 1: \(71 + 24 = \text{___}\) (or) \(61 + 34 = \text{___}\).

Item 2: \(17 + \text{___} = 64\) (or) \(15 + \text{___} = 73\).

Item 3: \(13 \times \text{___} = 104\) (or) \(14 \times \text{___} = 126\).

Item 4: \(\text{___} - 23 = 46\) (or) \(\text{___} - 32 = 65\).

Item 5: \(\text{___} \div 12 = 7\) (or) \(\text{___} \div 13 = 6\).

Item 6: \(\text{___} + 23 = 655\) (or) \(\text{___} + 31 = 566\).
In general, the investigator found the display of table 10 fairly un-instructive. The gain for "the 78" on item "6" does give the appearance of being a leap of two grade levels, but the probability of several such leaps by chance alone, given eighteen occasions for such, would tend to discredit claims based on isolating one leap.

The following suggestions may be made by the data display of table 10:

1. Relative to test II, seventh grade "modified" classes appear to be made up of pupils more than two years behind in mathematics. (This suggestion is derived from comparing rows one, two, three, and seven of table 10.)

2. Relative to test II, pupils in seventh grade "modified" classes do not appear to be qualitatively different from the "normal" pupils of the NLSMA samples. (That is, percentages given for correct responses by the two groups tend to rank items similarly as to difficulty.)

3. Relative to test II, pupils in seventh grade "modified" classes appear to have gained in achievement by more than one half of a grade level. (This suggestion is derived from looking at those pairs of entries from rows one through six which allow for comparisons with rows seven through nine of the table on the preceding page.)
Comparisons for Test III ("Computations" Test)

Test III is very directly concerned with the intentions of the experimental ten week program. While the program is more weighted toward establishing understanding than toward teaching by repetitive drill, the intentions certainly include raising achievement levels relative to computational skills. Test III consists of the following items:

1. 103 + 7 2. 96 + 85 3. 378 + 63 + 504

4. 72 - 65 5. 834 - 49 6. 600 - 123

7. 56 x 3 8. 32 x 12 9. 235 x 305

10. 7/238 11. 40/280 12. 32/9792

Of particular interest are items "6", "9", "11", and "12" since each of these items includes "snags" for pupils whose understanding of algorithms might be very weak. Item "6" demands borrowing in order to borrow; item "9" invites pupils to err by mis-placing the partial product for the hundreds; item "11" invites an answer of seventy by placing seven above the eight and then throwing in a zero for good measure; and item "12" invites pupils to ignore placing a zero in the tens place for the quotient. (All of the potential "snags" proved to be such, on both tests,....but more so on the pretest.)
Table 11 and table 12 on the following pages display the data for comparing results from "the pretest-posttest comparison" to results derived by NLSMA. The different entries listed under "subjects" are defined as follows for those tables:

1. "(LOW)" identifies the 41 subjects who scored "0" to "5" on the pretest on computations.
2. "(MIDDLE)" identifies the 45 subjects who scored "6" to "8" on the pretest on computations.
3. "(HIGH)" identifies the 43 subjects who scored "9" to "12" on the pretest on computations.
4. "(ALL)" identifies all of the 129 subjects participating in the experimental program.
5. "(BELOW)" identifies the 66 subjects with a sum of scores for pretest and posttest less than or equal to 16.
6. "(ABOVE)" identifies the 63 subjects with a sum of scores for pretest and posttest greater than or equal to 17.
7. Remaining entries (rows) give data provided for different grade levels of the NLSMA x-population.

The breakdown of the subjects into "(BELOW)" and "(ABOVE)" is admittedly questionable. Its intention is to distribute subjects evidencing a possible "Galton Effect" so that high posttests are favored toward inclusion with the high achievement group and low posttests are favored.
TABLE 11

DISPLAY OF RESULTS FROM "THE PRETEST-POSTTEST COMPARISON STUDY" AND "NLSMA" DATA FOR ITEMS "1" THROUGH "6" OF TEST III (COMPUTATIONS TEST)

<table>
<thead>
<tr>
<th>Test</th>
<th>Subjects</th>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
<th>Item 5</th>
<th>Item 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>(LOW)</td>
<td>68</td>
<td>73</td>
<td>76</td>
<td>39</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>Posttest</td>
<td>(LOW)</td>
<td>63</td>
<td>80</td>
<td>88</td>
<td>73</td>
<td>61</td>
<td>32</td>
</tr>
<tr>
<td>Pretest</td>
<td>(MIDDLE)</td>
<td>78</td>
<td>87</td>
<td>91</td>
<td>82</td>
<td>64</td>
<td>47</td>
</tr>
<tr>
<td>Posttest</td>
<td>(MIDDLE)</td>
<td>91</td>
<td>96</td>
<td>96</td>
<td>91</td>
<td>78</td>
<td>64</td>
</tr>
<tr>
<td>Pretest</td>
<td>(HIGH)</td>
<td>91</td>
<td>98</td>
<td>93</td>
<td>100</td>
<td>93</td>
<td>91</td>
</tr>
<tr>
<td>Posttest</td>
<td>(HIGH)</td>
<td>91</td>
<td>93</td>
<td>100</td>
<td>100</td>
<td>93</td>
<td>88</td>
</tr>
<tr>
<td>Gr. 4, Fall</td>
<td>NLSMA</td>
<td>82</td>
<td>84</td>
<td>76</td>
<td>61</td>
<td>44</td>
<td>32</td>
</tr>
<tr>
<td>Gr. 7, Spr.</td>
<td>NLSMA</td>
<td>83</td>
<td>93</td>
<td>95</td>
<td>96</td>
<td>91</td>
<td>87</td>
</tr>
<tr>
<td>Pretest</td>
<td>(ALL)</td>
<td>79</td>
<td>86</td>
<td>87</td>
<td>74</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>Posttest</td>
<td>(ALL)</td>
<td>82</td>
<td>90</td>
<td>95</td>
<td>88</td>
<td>78</td>
<td>62</td>
</tr>
<tr>
<td>Pretest</td>
<td>(BELOW)</td>
<td>71</td>
<td>79</td>
<td>83</td>
<td>56</td>
<td>38</td>
<td>27</td>
</tr>
<tr>
<td>Posttest</td>
<td>(BELOW)</td>
<td>71</td>
<td>85</td>
<td>89</td>
<td>77</td>
<td>61</td>
<td>39</td>
</tr>
<tr>
<td>Pretest</td>
<td>(ABOVE)</td>
<td>87</td>
<td>94</td>
<td>90</td>
<td>94</td>
<td>83</td>
<td>73</td>
</tr>
<tr>
<td>Posttest</td>
<td>(ABOVE)</td>
<td>94</td>
<td>95</td>
<td>100</td>
<td>100</td>
<td>95</td>
<td>86</td>
</tr>
<tr>
<td>Test</td>
<td>Subjects</td>
<td>Item</td>
<td>Item</td>
<td>Item</td>
<td>Item</td>
<td>Item</td>
<td>Item</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Pretest</td>
<td>(LOW)</td>
<td>54</td>
<td>22</td>
<td>7</td>
<td>10</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Posttest</td>
<td>(LOW)</td>
<td>78</td>
<td>49</td>
<td>24</td>
<td>34</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td>Pretest</td>
<td>(MIDDLE)</td>
<td>96</td>
<td>84</td>
<td>24</td>
<td>36</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>Posttest</td>
<td>(MIDDLE)</td>
<td>87</td>
<td>87</td>
<td>47</td>
<td>58</td>
<td>49</td>
<td>31</td>
</tr>
<tr>
<td>Pretest</td>
<td>(HIGH)</td>
<td>100</td>
<td>98</td>
<td>84</td>
<td>88</td>
<td>65</td>
<td>44</td>
</tr>
<tr>
<td>Posttest</td>
<td>(HIGH)</td>
<td>100</td>
<td>98</td>
<td>91</td>
<td>88</td>
<td>91</td>
<td>63</td>
</tr>
<tr>
<td>Gr. 4, Fall</td>
<td>NLSMA</td>
<td>40</td>
<td>6</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Gr. 5, Spr.</td>
<td>NLSMA</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>76</td>
<td>44</td>
</tr>
<tr>
<td>Gr. 7, Spr.</td>
<td>NLSMA</td>
<td>97</td>
<td>92</td>
<td>--</td>
<td>--</td>
<td>89</td>
<td>67</td>
</tr>
<tr>
<td>Gr. 8, Spr.</td>
<td>NLSMA</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>92</td>
<td>69</td>
</tr>
<tr>
<td>Pretest</td>
<td>(ALL)</td>
<td>84</td>
<td>69</td>
<td>39</td>
<td>45</td>
<td>29</td>
<td>16</td>
</tr>
<tr>
<td>Posttest</td>
<td>(ALL)</td>
<td>88</td>
<td>78</td>
<td>54</td>
<td>60</td>
<td>59</td>
<td>37</td>
</tr>
<tr>
<td>Pretest</td>
<td>(BELOW)</td>
<td>68</td>
<td>45</td>
<td>9</td>
<td>17</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Posttest</td>
<td>(BELOW)</td>
<td>79</td>
<td>59</td>
<td>23</td>
<td>35</td>
<td>35</td>
<td>14</td>
</tr>
<tr>
<td>Pretest</td>
<td>(ABOVE)</td>
<td>100</td>
<td>94</td>
<td>70</td>
<td>75</td>
<td>52</td>
<td>32</td>
</tr>
<tr>
<td>Posttest</td>
<td>(ABOVE)</td>
<td>98</td>
<td>98</td>
<td>87</td>
<td>87</td>
<td>84</td>
<td>62</td>
</tr>
</tbody>
</table>
toward inclusion with the low achievement group. The break­
down would likely be indefensible except for the nature of
the experimental mortality endured by this study.

The data seems to suggest that under-achievers demon­
strate their being such more as the complexity of the task
increases. Comparing the rows for "(ALL)" with those for
"NLSMA" on table 11 would suggest that the under-achievers
score at grade level at posttest time on addition; score at
slightly less than grade level at posttest time on sub­
traction, with the most pronounced evidence of under-achiev­
ment yielded by problem "6", the first of the "snag" items.
Comparing the rows for "(ALL)" with those for "NLSMA" on
table 12 further confirms this suggestion in that the dif­
ferences between row nine and row twelve appear to be more
pronounced for division (items "11" and "12") than for mul­
tiplication (items "7" and "8").

This suggestion seems still further supported by ob­
serving that posttest results for "(HIGH)" and "(ABOVE)"
show only slight tendencies toward lower grade level achieve­
ment as task complexity increases (by a comparison to "gr. 7,
spr.") while posttest results for "(MIDDLE)" show a stronger
tendency toward lower grade level achievement as task com­
plexity increases (by a comparison to "gr. 7, spr.").

Comparing pretest results to NLSMA data makes an even
stronger case for the suggestion. For example, pretest
results for "(HIGH)" shows this group to be at or above
grade level through multiplication, but somewhere in the fifth grade on division.

In general, "(LOW)", "(MIDDLE)", AND "(BELOW)" groups display pretest and posttest percentages that fail to lend themselves to identifying the nature of gains by a comparison of gains on "snag" items to gains on other items, or by a comparison of gains on simpler tasks to gains on more complex tasks.

On the other hand, comparing pretest percentages to posttest percentages on items "11" and "12" for either the "(HIGH)" or "(ABOVE)" group would indicate achievement gains of nearly two years relative to those items, with far less convincing gains (or losses) on other items.

In summary, the investigator gleans the following two suggestions from the "comparison to NLSMA" relative to test III. First, under-achievement is evidenced more clearly as task complexity increases. While this suggestion is certainly not shocking, it might appear to be somewhat in opposition to claims made by some teachers that the primary difficulty in dealing with under-achievers is that they have never learned the basic "facts", ...that is, have never memorized the addition and multiplication tables. Rather, as suggested by the data displays, it may be the case that the primary difficulty in dealing with under-achievers is their tendency to "over-compartmentalize" learning, which could explain how some such pupils do well on addition, subtrac-
tion, and multiplication, the prerequisite skills to doing division, and then do poorly on division, viewing it as a completely new sort of a task.

Second, and very cautiously made, is the suggestion that the experimental program is effective toward yielding more "normal" relationships between the probability of success and task difficulty. That is, the tendency for underachievement to be more clearly visible as task difficulty increases is less evidenced by posttest results than by pretest results.

A possible third suggestion, denied a convincing verification both by the data displays and by the restrictions on comparisons listed on page 109, is that achievement levels relative to test III gained by more than half a year over the ten week period of the program.

Comparisons for Test IV ("Algorithms" Test)

While the inclusion of test IV appeared to have been the wisest decision relative to determining differences between the experimental and control treatments of the last section of this chapter, it can best be described as a modest disaster relative to any attempt toward determining the substantive significance of gains by comparisons to data provided by NLSMA.

Response patterns yielded by pupils participating in the experimental program were both quantitatively and perhaps
qualitatively of a nature denying much enlightenment from a comparison with NLSMA data.

Table 13 and table 14 on the following two pages will serve to add substance to the above remarks. The different entries listed under "subjects" are defined as follows:

1. "(ALL)" identifies all of the 129 subjects participating in the ten week experimental program.
2. "(LOW)" identifies the 82 subjects who scored "0" to "2" on the pretest.
3. "(HIGH)" identifies the 47 subjects who scored "3" or above on the pretest.
4. Remaining entries (rows) give data provided for different grade levels of the NLSMA x-population.

Numerical entries give the percentages of pupils in the group indicated who responded correctly to the item. (Item "13" of test IV is omitted from table 14 since there is no comparative data available from the NLSMA studies.

The most obvious suggestion made by the data displays is that pupils of seventh grade modified classes do very poorly on test IV,....so poorly, in fact, that it is impossible to indicate a grade level of performance for either pretest or posttest by a comparison to NLSMA data.

It is likely that a primary factor in the results for test IV is the fact that many of the pupils participating in the experimental study suffered severe reading handicaps.

In the opinion of the investigator, the data strongly
TABLE 13

DISPLAY OF RESULTS FROM "THE PRETEST-POSTTEST COMPARISON STUDY" AND "NLSMA" DATA FOR ITEMS "1" THROUGH "6" OF TEST IV (ALGORITHMS TEST)

<table>
<thead>
<tr>
<th>Test</th>
<th>Subjects</th>
<th>Item &quot;1&quot;</th>
<th>Item &quot;2&quot;</th>
<th>Item &quot;3&quot;</th>
<th>Item &quot;4&quot;</th>
<th>Item &quot;5&quot;</th>
<th>Item &quot;6&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest (ALL)</td>
<td>19</td>
<td>7</td>
<td>33</td>
<td>41</td>
<td>11</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Posttest (ALL)</td>
<td>35</td>
<td>14</td>
<td>45</td>
<td>47</td>
<td>19</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Pretest (LOW)</td>
<td>6</td>
<td>4</td>
<td>22</td>
<td>34</td>
<td>7</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Posttest (LOW)</td>
<td>29</td>
<td>2</td>
<td>39</td>
<td>43</td>
<td>17</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Pretest (HIGH)</td>
<td>40</td>
<td>13</td>
<td>51</td>
<td>53</td>
<td>17</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Posttest (HIGH)</td>
<td>45</td>
<td>34</td>
<td>55</td>
<td>55</td>
<td>23</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Gr. 6, Spr. NLSMA</td>
<td>39</td>
<td>22</td>
<td>67</td>
<td>39</td>
<td>46</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Gr. 8, Spr. NLSMA</td>
<td>58</td>
<td>29</td>
<td>86</td>
<td>53</td>
<td>60</td>
<td>44</td>
<td></td>
</tr>
</tbody>
</table>

suggested that under-achievers are very vulnerable to the allurement of "distracters" on test IV. The following observations may provide evidence for such a claim. (See the appendix for the items of test IV.)

Items "3", "4", "10", and "11" are such that pupils might answer correctly without giving any heed to the problem accompanying the item. Item "8" allows only one answer
TABLE 14
DISPLAY OF RESULTS FROM "THE PRETEST-POSTTEST COMPARISON STUDY" AND "NLSMA" DATA FOR ITEMS "7" THROUGH "12" OF TEST IV

<table>
<thead>
<tr>
<th>Test</th>
<th>Subjects</th>
<th>Item &quot;7&quot;</th>
<th>Item &quot;8&quot;</th>
<th>Item &quot;9&quot;</th>
<th>Item &quot;10&quot;</th>
<th>Item &quot;11&quot;</th>
<th>Item &quot;12&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>(ALL)</td>
<td>9</td>
<td>15</td>
<td>7</td>
<td>25</td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>Posttes</td>
<td>(ALL)</td>
<td>12</td>
<td>27</td>
<td>12</td>
<td>37</td>
<td>35</td>
<td>23</td>
</tr>
<tr>
<td>Pretest</td>
<td>(LOW)</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>20</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Posttes</td>
<td>(LOW)</td>
<td>11</td>
<td>22</td>
<td>11</td>
<td>33</td>
<td>29</td>
<td>23</td>
</tr>
<tr>
<td>Pretest</td>
<td>(HIGH)</td>
<td>13</td>
<td>30</td>
<td>15</td>
<td>34</td>
<td>34</td>
<td>21</td>
</tr>
<tr>
<td>Posttes</td>
<td>(HIGH)</td>
<td>15</td>
<td>36</td>
<td>13</td>
<td>45</td>
<td>45</td>
<td>23</td>
</tr>
<tr>
<td>Gr. 6, Spr.</td>
<td>NLSMA</td>
<td>25</td>
<td>33</td>
<td>13</td>
<td>47</td>
<td>59</td>
<td>44</td>
</tr>
<tr>
<td>Gr. 8, Spr.</td>
<td>NLSMA</td>
<td>34</td>
<td>47</td>
<td>17</td>
<td>60</td>
<td>71</td>
<td>61</td>
</tr>
</tbody>
</table>

if the pupil is willing to at least look at the "answer" given for the accompanying problem. Ranking items in increasing order of difficulty by an appeal to pretest results for "(ALL)" yields the following: "4", "3", "10", "11", "1", "8", "12", "5", "7", "2", "9", "6". (The posttest results are used as "tie-breakers" in the ranking.) Items with particularly attractive distracters are items "2", "5", "8", "9".
"6", and "9",....attractive, that is, if the pupil either ignores the accompanying problem or responds with a naive disregard for "place value".

The above does not intend to suggest that the more "normal" subjects of the NLSMA samples are not attracted to distracters. Table 15 on the following page suggests that these "normal" subjects tend to rank items in very much the same way as do the under-achievers of this study.

The rows of table 14 rank items from easy to hard as determined by percentages of responses that were correct; with pretest results serving as tie-breakers for posttest results, posttest results serving as tie-breakers for pretest results, and with items appearing later in the test given preference as easier in case tie-breakers fail to break ties. The "Test-Group" entries of table 15 are defined as follows:

1. "Pr-7" gives pretest results over "(ALL)".
2. "Po-7" gives posttest results over "(ALL)".
3. "NL-6" gives results for NLSMA 6th grade, spring.
4. "NL-8" gives results for NLSMA 8th grade, spring.
5. "Pr-L" gives pretest results over "(LOW)".
6. "Po-L" gives posttest results over "(LOW)".
7. "Pr-H" gives pretest results over "(HIGH)".
8. "Po-H" gives posttest results over "(HIGH)".

The additional symbols used in table 15 are defined as follows: "-" indicates that correct responses are likely
TABLE 15
RANKINGS FOR ITEMS ON TEST IV AS DETERMINED BY PERCENTAGES OF CORRECT RESPONSES

<table>
<thead>
<tr>
<th>Test-Group</th>
<th>Item Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Easy......................... Hard</td>
</tr>
<tr>
<td>Pr-7</td>
<td>4, 3, 10, 11, 1, 8, 12, 5, 7, 2, 9, 6</td>
</tr>
<tr>
<td>Po-7</td>
<td>4, 3, 10, 11, 1, 8, 12, 6, 5, 2, 7, 9</td>
</tr>
<tr>
<td>NL-6</td>
<td>3, 11, 10, 5, 12, 1, 4, 8, 6, 7, 2, 9</td>
</tr>
<tr>
<td>NL-8</td>
<td>3, 11, 12, 10, 5, 1, 4, 8, 6, 7, 2, 9</td>
</tr>
<tr>
<td>Pr-L</td>
<td>4, 3, 10, 11, 5, 1, 8, 7, 12, 6, 2, 9</td>
</tr>
<tr>
<td>Po-L</td>
<td>4, 3, 10, 11, 1, 12, 8, 6, 5, 7, 9, 2</td>
</tr>
<tr>
<td>Pr-H</td>
<td>4, 3, 1, 11, 10, 8, 12, 5, 9, 2, 7, 6</td>
</tr>
<tr>
<td>Po-H</td>
<td>4, 3, 1, 11, 10, 8, 2, 12, 5, 6, 7, 9</td>
</tr>
</tbody>
</table>

for pupils who ignore the accompanying example; "o" indicates that some attention must be given to the accompanying example; ....or that none of the distracters appear particularly attractive; "+" indicates that there is at least one distracter that would be attractive to a naive treat-
Apologies for "beating a dead horse" may be in order for the last few pages. There is, after all, only one suggestion of those listed which appears solid.....and that is that the pupils of the seventh grade modified classes participating in the experimental did poorly relative to test IV, both on the pretest and on the posttest.

About two weeks after the posttest was given, the investigator administered test IV to ten of the under-achievers in order to decide if the low results stemmed largely from failures to read or to understand directions given for items. The investigator read each item to the pupils and then asked them to respond, without placing a strict time limit on the thirteen items. Test scores from that session were more than double the scores yielded by those pupils on posttest IV at the end of the ten week program.

The suggestion from the above is that test IV is somewhat ill-suited to determining the achievement levels for pupils in seventh grade modified classes.

Correlations with Reading Scores

Table 16 on the following page displays correlations for mathematics tests and reading tests. Correlations are computed by a Pearson's "r". Correlations at or above 0.1860 are statistically significant at the .05-level. On table 16, the identification of rows and columns is as
TABLE 16
CORRELATION MATRIX FOR MATHEMATICS
AND READING TEST RESULTS

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-----</td>
<td>.2885</td>
<td>.8695</td>
<td>.7406</td>
<td>.3909</td>
<td>.7300</td>
<td>.0596</td>
<td>.0806</td>
</tr>
<tr>
<td>(2)</td>
<td>.2885</td>
<td>-----</td>
<td>.5734</td>
<td>.2410</td>
<td>.4200</td>
<td>.3844</td>
<td>.2240</td>
<td>.2035</td>
</tr>
<tr>
<td>(3)</td>
<td>.8695</td>
<td>.5734</td>
<td>-----</td>
<td>.6693</td>
<td>.4533</td>
<td>.7978</td>
<td>.0717</td>
<td>.1646</td>
</tr>
<tr>
<td>(4)</td>
<td>.7406</td>
<td>.2410</td>
<td>.6693</td>
<td>-----</td>
<td>.3152</td>
<td>.8080</td>
<td>.0510</td>
<td>.1360</td>
</tr>
<tr>
<td>(5)</td>
<td>.3909</td>
<td>.4200</td>
<td>.4533</td>
<td>.3152</td>
<td>-----</td>
<td>.6612</td>
<td>.3994</td>
<td>.2942</td>
</tr>
<tr>
<td>(6)</td>
<td>.7300</td>
<td>.3844</td>
<td>.7978</td>
<td>.8080</td>
<td>.6612</td>
<td>-----</td>
<td>.1001</td>
<td>.1822</td>
</tr>
<tr>
<td>(7)</td>
<td>.0596</td>
<td>.2240</td>
<td>.0717</td>
<td>.0510</td>
<td>.3994</td>
<td>.1001</td>
<td>-----</td>
<td>.6591</td>
</tr>
<tr>
<td>(8)</td>
<td>.0806</td>
<td>.2035</td>
<td>.1646</td>
<td>.1360</td>
<td>.2942</td>
<td>.1822</td>
<td>.6591</td>
<td>-----</td>
</tr>
</tbody>
</table>

follows:
1. "(1)" is test III, the "computations" pretest.
2. "(2)" is test IV, the "algorithms" pretest.
3. "(3)" is the pretest composite of mathematics tests.
4. "(4)" is test III, posttest.
5. "(5)" is test IV, posttest.
6. "(6)" is the posttest composite of mathematics tests.
7. "(7)" is the reading "vocabulary" test.
8. "(8)" is the reading "comprehension" test.

The composite of mathematics tests includes tests II, III, and IV already defined in this chapter.
TABLE 17
DIFFERENCES BETWEEN CORRELATIONS FOR
READING SCORES AND MATHEMATICS
SCORES FROM PRETEST TO
POSTTEST

<table>
<thead>
<tr>
<th>Mathematics Test-Reading Test</th>
<th>Pretest &quot;r&quot;</th>
<th>Posttest &quot;r&quot;</th>
<th>Significance of difference between &quot;r's&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>III-Vocabulary</td>
<td>.0596</td>
<td>.0510</td>
<td>--</td>
</tr>
<tr>
<td>IV-Vocabulary</td>
<td>.2240</td>
<td>.3994</td>
<td>.10</td>
</tr>
<tr>
<td>COMP-Vocabulary</td>
<td>.0717</td>
<td>.1001</td>
<td>--</td>
</tr>
<tr>
<td>III-Comprehension</td>
<td>.0806</td>
<td>.1360</td>
<td>--</td>
</tr>
<tr>
<td>IV-Comprehension</td>
<td>.2035</td>
<td>.2945</td>
<td>--</td>
</tr>
<tr>
<td>COMP-Comprehension</td>
<td>.1646</td>
<td>.1822</td>
<td>--</td>
</tr>
</tbody>
</table>

The following hypotheses were stated for "the correlations with reading scores":

\[ H_0(\rho_{\text{pre-A,B}} = \rho_{\text{post-A,B}}) \]: There will be no difference between the correlation of pretest "A" with "B" and the correlation of posttest "A" with "B".

- A = III, IV, COMPosite (II, III, IV).
- B = Vocabulary, Comprehension.

Table 17 provides the data for judging the hypotheses. Statistical significance levels are given by a two-tail "t-test", the Hotelling Formula for Correlated Correlations.
The reading scores are from a city-wide testing conducted mid-way through the ten week period of the experimental program. The two reading tests are the "vocabulary" test and the "comprehension" test of the "California Comprehensive Reading Test", Level 3.

Most of the hypotheses of type \( \text{Ho}(T_{\text{pre-A,B}} \cdot T_{\text{post-A,B}}) \) for the correlations with reading scores appear to be supported by the data, with the probable exception of test IV and vocabulary, and at least suspicions for test IV and comprehension. That is, there is evidence to suggest that the experimental program raises the correlation between reading skills and performance on the algorithms test.

In the opinion of the investigator, the correlations for mathematics scores with reading scores are remarkably low. Perhaps a more fruitful check of the experimental program's ability to raise mathematics achievement levels by capitalizing on reading abilities would be to conduct a similar study using "average" pupils. Ideally, of course, such a study would be an experimental-control design which included reading tests with both pretest and posttest, to also check the possibility that the experimental mathematics program affects reading achievement levels.
CHAPTER V
SUNDRY OBSERVATIONS, COMMENTS, AND SUSPICIONS

The content covered by this chapter and not by others is mainly anecdotal, not anchored in claims of statistical significance, and perhaps tainted by bias. Readers with low tolerance levels for such are invited to ignore this chapter, and those with only moderate tolerance levels are invited to move directly to page 160. Any remaining reader is invited to draw the drapes, or hide this document behind a respectable journal, and proceed.

The larger part of this chapter is devoted to looking at each of the eight classes individually. Some statistical information will be provided for each school, but not in the form of exact figures. Indicated by such data will be: daily absence rate, pupil mobility rate, percentage of pupils new to Columbus Schools, percentage of pupils non-white, staff turnover rate, percentage of pupils above age in grade level, and the incidence of ADC cases to enrollment.

The decision to withhold exact information was made for the following two reasons. First, it could be in violation to the promise made to the Columbus Public School System that anonymity of both the schools and participating teachers would be preserved. Second, classes of seventh grade modified
pupils are not likely to be representative of the total school population.

Data displays (those of chapter IV) for the mathematics tests will also be provided for each class. No mention will be made of the psychological tests except in cases where the investigator happens to notice something which seems curious to him. In general, the statistical significance levels yielded by looking at the psychological measurements over single classes were low, ... too low to merit taking most such differences very seriously.

The same procedure was used in securing each of the eight classes. Before talking with a teacher, the principal of the school was given a sample of the materials and asked to decide whether or not he considered them worthy of a trial with seventh grade modified classes in his school. With his approval of the materials, the teacher was approached, given a set of the materials, and given a week to reach a decision. It was emphasized that the option not to participate in the program was truly available. The original list of schools which might be contacted was supplied by the Department of Evaluation and Research of the Columbus Public School System.

As it happened, no principal nor any teacher who was approached decided not to participate in the program. While this fact may speak of the investigator as a salesman and say nothing of the program, it does at least hold research advantages over a sample of eight agreeable to the program out of
twenty approached, ... which would leave the study vulnerable to attacks like those leveled against the Kinsey Report, a treatise on the sex practices of those willing to chat with strangers about the subject.

Also, teachers were encouraged not to begin the program until they were satisfied that children had completed the transition from summer's freedoms to days by bells, and had completed at least a brief review of the basic operations. In keeping with this request, teachers administered the pretest and began the program sometime during the third, fourth, or fifth week of the school year. A main reason for requesting a slight delay in beginning the program was to lower the probability that pretest results would be biased due to forgetting over the summer that might be quickly recovered by almost any classroom methodology.

All the teachers completed the program in nine to eleven weeks. Since this variation in time was not considered significant in the opinion of the investigator, all references to the time period for the program are given as "the ten week period" of the program.

The text of pages 133 through 160 describes the program by individual classes; that of pages 160 through 162 attempts to identify ground common to the eight uses of the program; that of pages 162 through 167 describes the program relative to the ten goals of chapter II; and that of pages 167 through 170 may be described as miscellany.
133

Classes Viewed Individually

Class (1)

1. Data on the School:
   a. classification: outer-city
   b. daily absence rate: below 8%
   c. pupil mobility rate: below 14%
   d. pupils new to Columbus Public Schools: below 7%
   e. pupils non-white: below 10%
   f. staff turnover rate: 30-40%
   g. pupils above age in grade level: below 15%
   h. incidence of ADC cases: below 10%

2. Data on the Teacher:
   a. sex: female
   b. age: 20-30 yrs.
   c. major area of study in college: mathematics
   d. previous years of teaching experience: 2
      i. at elementary level: 0
      ii. at secondary level: 2
   e. previous years experience with at least one class
      of under-achievers in math: 2

3. Results on the Math Tests:
   a. on test II, pretest and posttest means: 2.9, 3.9
   b. on test III, pretest and posttest means: 7.0, 9.0
   c. on test IV, pretest and posttest means: 1.6, 3.4
   d. on COMP, pretest and posttest means: 11.5, 16.4

(See page 90 for data displays.)
4. Experimental Mortality Information
   a. 5 took pretest only ....... pretest mean 9.8
   b. 5 took posttest only ....... posttest mean 12.6
   c. 19 took both tests ....... means: 11.5; 16.4

   (The main factor in mortality was a class scheduling change that took place at about the fourth week of the ten week program.)

5. Sundry Observations

   The investigator visited class (1) on two occasions during the ten week period of the program. Class (1) was the only class of the eight classes that saw the investigator during the program.

   According to the teacher, one of her main considerations in agreeing to use the unit was a shortage of textbooks. (This information was offered after the unit had been in use by her class for about four weeks.) In view of this, the investigator concludes that the teacher's initial agreement to participate in the program was not wildly enthusiastic. It was also revealed at that time that the school's principal, who saw the materials before the teacher saw them and who liked them very much, had applied subtle pressures to have the teacher agree to participate in the program.

   At the end of the second week of using the materials, the teacher reported the following:

   1. Pupils never leave their books in their lockers.
      (This phenomenon was viewed by the teacher as a modest miracle.)
2. Evidence that pupils are reading the materials outside of class is provided by their reports to the class of reactions by family members to the narrative.

3. A number of pupils asked the teacher for time to report to the class on counting habits of younger children,....without having had item eleven on page 289 of the workbook formally assigned by the teacher.

4. A number of the pupils requested to keep the books as personal property, in spite of the fact that the initial understanding was that books should be returned after the program was completed.

The only worrisome information received from the teacher came at the end of the fourth week. One of the boys in class had been annoying others by insisting that the class read the narrative and ignore both the workbook and the drill. As described by the teacher, he began each class session with a persistent and loud "Let's read today!" One of the girls, recognized by her peers as bright, cut the young man down one day by referring to the narrative as "a child's fairy tale."

The teacher had been told, prior to beginning the program, not to suppress such reactions, since the possibility of such attacks on the story had been a chief concern during both pilot studies,....and was still something of a
concern to the investigator.

The incident did not repeat itself. In the teacher's opinion, given several weeks later, the outburst was intended to be against the boy, not against the narrative.

At no point during the ten week period did the pupils appear to weary of the program. In fact, the chapter on division was rated by the teacher as being the most effective of the five chapters,...with the chapter on counting running a very close second.

Several times the teacher expressed the opinion that there were fewer managerial problems than she would have expected to encounter using a standard textbook.

She also emphasized that the class using the program was the first class she had taught that showed an interest in knowing the actual historical background for mathematical concepts. In her opinion, it was the fictitious narrative that had motivated such interest, since she had tried, largely in vain, to motivate an interest in history with former classes.

Class (2)

1. Data on the School:
   a. classification: outer-city
   b. daily absence rate: below 8%
   c. pupil mobility rate: below 14%
   d. pupils new to Columbus Public Schools: below 7%
   e. pupils non-white: below 10%
f. staff turnover rate: 20-30%

g. pupils above age in grade level: below 15%

h. incidence of ADC cases: below 10%

2. Data on the Teacher:
   a. sex: male
   b. age: 40-50 yrs.
   c. major area of study in college: elementary ed.
   d. previous years teaching experience: over 10
      i. at elementary level: over 10
      ii. at secondary level: 6
   e. previous years experience with at least one
      class of under-achievers: 4

3. Results on the Math Tests:
   a. on test II, pretest and posttest means: 4.2, 4.7
   b. on test III, pretest and posttest means: 8.2, 9.1
   c. on test IV, pretest and posttest means: 2.3, 4.0
   d. on COMP, pretest and posttest means: 14.7, 17.8
      (See page 90 for data displays.)

4. Experimental Mortality Information:
   a. 2 took pretest only........pretest mean 11.5
   b. 2 took posttest only........posttest mean 10.0
   c. 21 took both tests........means: 14.7, 17.8

5. Sundry Observations:
   The enthusiasm with which the teacher originally agreed to participate in the program was conspicuous, a bit too conspicuous, leaving no place to go but down. At the end of
the second week he reported that things just couldn't be going more smoothly. At the end of the third week he reported that his class had asked to put on a play enacting the early chapters of the book. (The investigator is not completely convinced that this data originated with the pupils.)

The first hint that an emotional peak had been attained came with a suggestion in about the fifth week that perhaps the story would be more interesting if the hero should expire and pass the business of multiplication and division to an offspring. This suggestion came along with his observation that some of the pupils had managed to sit through six years of elementary school without learning either the addition or the multiplication tables. It might be suggested that the disenchantment made its appearance at about the same time that some of the pupils began finding the content difficult.

When asked to rate the chapters at the conclusion of the program, the teacher suggested a descending order of effectiveness, rating the chapter on counting most effective and the one on division least effective. Of the eight teachers, he was the only one by whom the chapter on division was given a rating below one or two (of five chapters).

Since this teacher was the only one of eight who saw the unit as decreasing in effectiveness over the ten week period, it may be instructive to contemplate the following:

1. Class (2) showed the highest pretest scores of
the eight classes on "open sentences", "computation", and on the composite of the three math tests. The class ranked third from the top on the pretest on algorithms.

2. Early in the program, the teacher told the class that part of their job would be to decide on which grade level could most profitably use the experimental program. (Encountering difficulty after having reached a decision could be highly ego-threatening for a pupil.)

3. No class spent more time on chapter one (counting) than did class (2), and no class spent less time on chapter five (division).

It is the investigator's opinion that the class, along with the teacher, was "spoiled" during the early weeks of the program,...led to believe that mathematics had suddenly become joyous and easy. The jolt that came with difficulties from the content was likely intensified by the fact that the class had been led to believe that the task at hand included determining an appropriate grade level for the program's use. (It is not likely that anyone reached a decision that the materials belonged to some more advanced grade level.)

Further evidence that the pupils had been disillusioned is provided by the fact that this was the only class that showed a statistically significant drop (.02 by a two-tail
sign test) from pretest to posttest over the two "ideal self-concept" items. In other words, this was the only class that convincingly appeared to care less about doing well in mathematics at the time the program was completed. Perhaps this was due in large part to a sharp drop in teacher enthusiasm accompanied by an increasing difficulty in subject content.

Surprising to the investigator, class (2) retained its top ranking on the posttest composite of the math tests. This phenomenon could have to do with the low experimental mortality compared to that of other classes; or it might suggest that teacher enthusiasm curves have little to do with pupil achievement. Of particular interest, in view of his loss of enthusiasm for the materials over the ten weeks, is the fact that the teacher never reported pupil disdain for the story.

The investigator visited the teacher of class (2) a last time about two months after the completion of the program in an effort to determine whether the loss of enthusiasm shown over the ten week period had roots in the nature of the experimental unit or in disappointment with the particular class. At that time he strongly held the opinion that the experimental unit was superior to the more standard textbook in dealing with underachievers.

Class (3)

1. Data on the School:
   a. classification: outer-city
   b. daily absence rate: 8-15%
   c. pupil mobility rate: below 14%
d. pupils new to Columbus Public Schools: below 7%
e. pupils non-white: 10-30%
f. staff turnover rate: 10-20%
g. pupils above age in grade level: 15-25%
h. incidence of ADC cases: below 10%

2. Data on the Teacher:
   a. sex: female
   b. age: 20-30 yrs.
   c. major area of study in college: mathematics
   d. previous years of teaching experience: 0
      1. at elementary level: 0
      11. at secondary level: 0
   e. previous years experience with at least one class of under-achievers: 0

3. Results on the Math Tests:
   a. on test II, pretest and posttest means: 2.4, 3.4
   b. on test III, pretest and posttest means: 6.3, 8.5
   c. on test IV, pretest and posttest means: 1.2, 2.8
   d. on COMP, pretest and posttest means: 9.9, 14.6
      (See page 90 for data displays.)

4. Experimental Mortality Information:
   a. 5 took pretest only.......pretest mean 14.6
   b. 7 took posttest only.......posttest mean 10.9
   c. 16 took both tests.......means: 9.9: 14.6
      (The main factor in mortality patterns was the advancement of some pupils to regular math classes
during the program and the influx of other pupils who were having difficulty with the content of the regular classes.)

5. Sundry Observations:

After expressing a cautious willingness to participate in the program, the teacher of class (3) turned the conversation to managerial problems that might result from using the unit. At the time of this first contact, she appeared to hold the opinion that seventh grade modified classes were both unteachable and unmanageable. As further evidence that her worries were genuine, she waited a full week after asking for the materials before beginning the program with her class. Her reason given for the delay was that she wanted to know how other teachers using the unit fared with the early chapters. At the end of the additional week's delay, she began the program, ....without the information she had requested.

She reported the following upon concluding the chapter on counting:

1. Pupils do not leave their books in their lockers.
   (This information about not leaving books in the lockers was emphasized by three of the eight teachers as indicative of genuine enthusiasm on the part of pupils. It does suggest a curious way for measuring attitude, along with looking at whether or not pencils are sharpened, whether pupils are seated at the bell, etc.)

2. In order to control for excessive pupil volunteering
to read aloud, a procedure for taking turns was put into effect.

3. From their corner seats of banishment, two of the larger and more troublesome boys were seen diligently counting on their fingers during a reading aloud of the hero's difficulties with training scouts to count in chapter one. The teacher seemed almost amazed that the story had not been scorned by such pupils.

At about the fourth week into the program the English teacher who had the same pupils asked them what they were doing in math class. At that time the chapter on subtraction had just been completed. Her report of the session during the English class included the following:

1. The discussion consumed the entire hour, unyielding to her several attempts to get back to the English assignment.

2. The pupils' recall of detail from the story was remarkable. Of special interest were irregularities to what one might expect of primitive cave dwellers, such as pajamas, secretaries, night school, and scout unions.

3. Many of the pupils regarded the characters in the story as "stupid". As evidence they cited the "fix-the-answer-later method" and the conversation between the hero and the musician that led to the
"borrow method" for subtraction.

4. Some of the time was spent debating whether or not the tribe was American Indian or African.

(If nothing else, the report by the English teacher would indicate that the narrative has potential toward encouraging emotional involvement, one of its goals.)

After completing the chapter on subtraction, the mathematics teacher devoted one week to drill using selected pages from the regular text. She reported sending an average of two pupils per day to the principal's office for disciplinary reasons during that week. She also reported that managerial problems almost disappeared upon returning to the experimental unit,...meaning fewer than two pupils per week being sent to the principal's office.

According to the teacher, pupil interest in the program did not wane over the ten week period. As further evidence of her conviction that standard textbooks were inappropriate for under-achievers, she designed the course for the remainder of the year around other materials, including several of the CILAMP publications.

In her opinion, the chapters on counting and division were the most effective,...with the chapter on subtraction a very close third.

The investigator spent a half hour talking with class (3) about one month after the posttest. In his judgment, the recall of detail was surprising. The class was divided
almost evenly on which was considered more valuable, the workbook or the narrative. Also confirmed by this discussion was the observation by the English teacher that a number of the pupils regarded the hero as being somewhat stupid in his approaches to mathematical problems. This reaction is curious in light of the reactions of several "normal" sixth grade pupils who were asked to read the story and who saw those very approaches as "clever"....It may be that under-achievers have a need for sometimes seeing stupidity in others.

Class (4)

1. Data on the School:
   a. classification: outer-city
   b. daily absence rate: 8-15%
   c. pupil mobility rate: below 14%
   d. pupils new to Columbus Public Schools: below 7%
   e. pupils non-white: 10-30%
   f. staff turnover rate: 10-20%
   g. pupils above age in grade level: 15-25%
   h. incidence of ADC cases: below 10%

2. Data on the Teacher:
   a. sex: male
   b. age: 20-30 yrs.
   c. major area of study in college: mathematics
   d. previous years of teaching experience:
      i. at elementary level: 0
      ii. at secondary level: 7
e. previous years experience with at least one class of under-achievers: 

3. Results on the Math Tests:
   a. on test II, pretest and posttest means: 2.5, 3.2
   b. on test III, pretest and posttest means: 6.1, 8.6
   c. on test IV, pretest and posttest means: 1.8, 3.2
   d. on COMP, pretest and posttest means: 10.5, 15.1

4. Experimental Mortality Information:
   a. 8 took pretest only.......pretest mean 10.5
   b. 4 took posttest only.......posttest mean 9.3
   c. 14 took both tests.......means: 10.5; 15.1

   (Although there was some movement of pupils to and from regular classes, the main factor for class (4) was absenteeism.)

5. Sundry Observations:

   The teacher of class (4) easily agreed to participate in the program, not because of a belief that the unit would be effective, but because he held the opinion that it could surely do no worse than the regular text. Class (4) met during the last period of the school day. It also had a reputation for being the worst assignment a teacher could draw. In describing the class, one of the more seasoned teachers announced firmly that he "wouldn't take that (particular) class if they paid an extra twenty-five bucks a day!"

   The teacher of class (4) reported no indication that any changes were occurring in the class throughout the first two
chapters. The "fix-the-answer-later method" for addition had proved troublesome for some and had been largely ignored by others, who held the majority.

The first glimmer of change was reported during the fifth week, shortly after beginning the chapter on multiplication, at the point of comparing "tall" and "long" multiplication. The teacher reported for the first time that he believed pupils were learning, even though indifference still reigned without peer.

The first conversation that clearly showed a touch of teacher excitement came about halfway through the chapter on division. At that point the teacher reported the following:

1. Managerial problems were on the decrease.
2. Pupils were interested in the story's characters.
3. The "search for a better method" was an effective way for presenting division.

The teacher rated the chapter on division as the most effective, with the chapter on multiplication second.

A possible (probable) factor in the low degree of emotional involvement by pupils is that the teacher did not enjoy reading aloud to the class. The main part of the reading aloud was required of pupils. Most teachers, on the other hand, took over completely with the reading at points in the story where the mathematics was considered to be difficult for pupils.
Class (5)

1. Data on the School:
   a. classification: inner-city
   b. daily absence rate: above 8%
   c. pupil mobility rate: above 14%
   d. pupils new to Columbus Public Schools: below 7%
   e. pupils non-white: 70-100%
   f. staff turnover rate: 20-30%
   g. pupils above age in grade level: below 15%
   h. incidence of ADC cases: 15-40%

2. Data on the Teacher:
   a. sex: male
   b. age: 40-50 yrs.
   c. major area of study in college: elementary ed.
   d. previous years of teaching experience: over 10
      i. at elementary level: over 10
   ii. at secondary level: 0
   e. previous years experience with at least one class of under-achievers: 0

3. Results on the Math Tests:
   a. on test II, pretest and posttest means: 2.5, 2.5
   b. on test III, pretest and posttest means: 7.2, 8.4
   c. on test IV, pretest and posttest means: 2.0, 3.2
   d. on COMP, pretest and posttest means: 11.6, 14.1
   (See page 91 for data displays.)
4. Experimental Mortality Information:
   a. 7 took pretest only......pretest mean 11.3
   b. 5 took posttest only......posttest mean 10.8
   c. 17 took both tests......means: 11.6; 14.1

(Absenteeism and movement of pupils to and from the regular math class each account for about half of mortality.)

5. Sundry Observations:

The teacher was strongly convinced that remedial programs at the junior high level were at least four years too late,...which is certainly not an unreasonable point of view. As a consequence, teacher expectations for the program appeared to be low. He did, however, believe that pupils would find the program more enjoyable (endurable) than the regular textbook, a main factor in his decision to participate in the program. Another factor suspected by the investigator, not confirmed, was pressure from the principal, who saw the program as having potential.

The teacher was noticeably worried about his own ability to deal with the materials. A main concern, initially, was the nature of and completeness of the teacher’s manual. In addition, there was obvious relief in the fact that exercises labeled "optional" were intended to be such, with decisions belonging to the teacher.

A second worry expressed by the teacher was that reading levels of the pupils were extremely low. Evidence that this fear was well-founded is provided by the number of
pupils in class (5) who "mis-read" directions on test II (open sentences) on both pretest and posttest. Eight of the seventeen pupils responded on either the pretest, or posttest, or on both by giving the operation indicated by the sign that appeared in the open sentence.

Consultations with the teacher over the ten week period were warm, friendly, but somewhat unproductive regarding feedback on the use of the unit. Little information was volunteered; answers to direct questions about the program were almost always favorable to the program, but were also vague and unconvincing. The only persistently reported information was that pupils seemed to like the narrative and its characters.

Hallway eavesdropping by the experimentor suggests that the class was very much teacher-directed. The teaching was mainly expository, with little time given to open questions or to pupil interaction,...and consequently there was little occasion for informal feedback. Discipline problems were usually nipped (perhaps smashed) in the bud. In short, the classroom atmosphere might have warmed the heart of a pre-Dewey conservative,...but it failed to offer much information to the investigator beyond that of the two tests.  

Class (6)  
1. Data on the School:  
   a. classification: inner-city  
   b. daily absence rate: above 8%
c. pupil mobility rate: above 14%
d. pupils new to Columbus Public Schools: below 7%
e. pupils non-white: 70-100%
f. staff turnover rate: 20-30%
g. pupils above age in grade level: below 15%
h. incidence of ADC cases: 15-40%

2. Data on the Teacher:
a. sex: male
b. age: 40-50 yrs.
c. major area of study in college: elementary ed.
d. previous years of teaching experience: over 10
   i. at elementary level: over 10
   ii. at secondary level: 0
e. previous years experience with at least one class of under-achievers: 0

3. Results on the Math Tests:
a. on test II, pretest and posttest means: 1.2, 1.3
b. on test III, pretest and posttest means: 6.9, 7.9
c. on test IV, pretest and posttest means: 1.7, 3.4
d. on COMP, pretest and posttest means: 9.8, 12.6

4. Experimental Mortality Information:
a. 11 took pretest only......pretest mean 9.5
b. 8 took posttest only......posttest mean 7.3
c. 11 took both tests.....means: 9.8; 12.6

(Absenteeism, movement of pupils to and from other classes, movement of pupils to and from the school area all contribute to the mortality.)
5. Sundry Observations:

Class (6) was the only class participating in the program that was "special modified,"...that is, comprised of pupils whose achievement records in reading indicated an unreadiness for the demands of seventh grade modified classes. The teacher of class (6) suggested that many of the pupils appeared to be unwilling to try reading anything at all; and that they also tended to ignore verbal instructions in favor of waiting for an example to be displayed. As a consequence, the teacher did almost all of the reading aloud of the narrative.

In the teacher's opinion, pupils enjoyed the story, but had difficulty following the discussions of algorithms that were not already familiar by past experience. The first evidence of such difficulty came with the treatment of the "fix-the-answer-later method" for addition. The teacher also reported later that "tall multiplication" had provided its share of trauma.

None of the optional exercises concerned with different number bases were assigned to the class. The general pattern was to read a part of the story; discuss and explain it directly, with only a moderate reliance on the discussion questions from the workbook; and then to devote the lion's share of classroom time to drill, often accompanied by clarifications relating the drill to the developments in the narrative.
About two weeks after the completion of the program, the investigator talked with ten members of class (6). When asked what they thought of things like "tall multiplication", about half of the pupils agreed that working with such algorithms was a waste of time because people already knew a better method and that it was also likely to mix up peoples' thinking so that it made long division seem harder than ever. The other half of the group believed such algorithms helped toward an understanding of what never was clear to them in grade school.

Although half of the pupils believed they understood more because of the program, only two of them thought it was important to understand "why" an algorithm yielded the answers it did. These were also the only two pupils who admitted that they would like to enter the regular math class if given a chance....The others insisted that the regular math sequence demanded more work than should be required of a mere mortal.

Relative to the math tests, the most pronounced difference between the special modified and modified classes appears to be with regard to response patterns on test II (open sentences). Nine of the eleven pupils responded to all items on both tests by simply naming the operation indicated by the operation sign appearing in the open sentence.

Two possible explanations are suggested for this
extreme pattern. First, it may be that the directions were too difficult for pupils so handicapped in reading. The teacher's description of the pupils would indicate that such might have been the case. A second possible explanation, the one preferred by the investigator, is that the pupils decided, by one quick glance at example "0" to test II, that they understood the task at hand; and that they consequently ignored both example "00" and the words that accompanied the examples. Reasons for embracing the second suggested explanation include the following:

1. The mean actually drops from pretest to posttest over test II, which would not be the expected pattern if an understanding of mathematics and reading ability were the only contributing factors.

2. The gain from pretest to posttest is statistically significant for another test, test IV; and only verbal instructions, no examples, were given for test IV.

Class (7)

1. Data on the School:
   a. classification: inner-city
   b. daily absence rate: 8-15%
   c. pupil mobility rate: above 14%
   d. pupils new to Columbus Public Schools: above 7%
   e. pupils non-white: below 10%
   f. staff turnover rate: 20-30%
g. pupils above age in grade level: 15-25%
h. incidence of ADC cases: 15-40%

2. Data on the Teacher:
   a. sex: male
   b. age: 30-40 yrs.
   c. major area of study in college: mathematics
   d. previous years of teaching experience: 8
      i. at elementary level: 0
      ii. at secondary level: 8
   e. previous years experience with at least one class of under-achievers: 8

3. Results on the Math Tests:
   a. on test II, pretest and posttest means: 3.1, 3.6
   b. on test III, pretest and posttest means: 8.0, 9.4
   c. on test IV, pretest and posttest means: 2.7, 3.9
   d. on COMP, pretest and posttest means: 13.8, 16.9

4. Experimental Mortality Information:
   a. 6 took pretest only………pretest mean 11.7
   b. 4 took posttest only………posttest mean 14.5
   c. 18 took both tests………means: 13.8; 16.9

(Absenteeism and movement of pupils to and from the school area are the main factors in mortality.)

5. Sundry Observations:

Class (7) was the class used in the experimental-control study. In an effort to remove extraneous factors from that study, the investigator was seen by the teacher only three times: when materials were first described and
delivered; when posttests were delivered for use; and when posttests were collected. The following comments derive from eavesdropping and from the single conference with the teacher after the program had been completed.

The teacher of class (7) appeared to have that rare ability which can allow a class discussion to approach the brink of disorder without reaching it. He also appeared to have the ability to be extremely firm (but with love) when pupils attempted to disrupt class proceedings. In short, he appeared to control the class very well without having to resort to despotism.

After the program, it was the teacher's opinion that pupils enjoyed the story more than he had expected them to enjoy it. The level of humor was considered appropriate and adequately diverse. He made special mention of the fact that different pupils reacted to quite different types of humor and to different parts of the story. As an example, he referred to one boy whose rather sullen manner was broken by the hero's argument for "STAMPEDE" as the last number. (The argument appears on page 189 of the narrative.) While this young man found it necessary to chuckle intermittently over the next few minutes, other members of the class treated the account as an outline of a rather serious mathematical matter.

Class (8)

1. Data on the School:
a. classification: inner-city
b. daily absence rate: above 8%
c. pupil mobility rate: above 14%
d. pupils new to Columbus Public Schools: above 7%
e. pupils non-white: 10-30%
f. staff turnover rate: 20-30%
g. pupils above age in grade level: 15-25%
h. incidence of ADC cases: 15-40%

2. Data on the Teacher:
a. sex: female
b. age: 20-30 yrs.
c. major area of study in college: elementary ed.
d. previous years of teaching experience: 1½
   i. at elementary level: 1½
   ii. at secondary level: 0
e. previous years experience with at least one class of under-achievers in math: 1

3. Results on the Math Tests:
a. on test II, pretest and posttest means: 2.4, 3.7
b. on test III, pretest and posttest means: 6.8, 9.1
c. on test IV, pretest and posttest means: 2.8, 4.4
d. on COMP, pretest and posttest means: 12.1, 17.2

4. Experimental Mortality Information:
a. 10 took pretest only........pretest mean 9.9
b. 8 took posttest only........posttest mean 11.8
c. 13 took both tests.........means: 12.1; 17.2
(The main factor in the mortality was a school day schedule change during the program. A second factor, also considerable, was the movement of pupils to and from the school area.)

5. Sundry Observations:

It took the teacher of class (8) less than ten minutes to reach a decision to participate in the program. When handed a copy of the narrative, she sat down and began reading at once. Before reaching the end of chapter one she closed the book and announced rather resolutely, "I definitely want to use it....Nothing can be worse than another regular textbook for these kids."

The teacher was quite enthusiastic at the end of the second week. The following observations were offered:

1. In two weeks, no pupil had left his books in his locker....The exlamatory manner of her making the observation suggested that some school record may have been set.

2. For the first time in the school year, the pupils showed an interest in mathematics.

3. The class discussion items in the workbook were particularly effective....Even some of the toughest boys joined the class in discussing the counting habits of smaller children.

4. There was never a shortage of volunteers for reading the story aloud in class. To control for this excess, reading assignments were made by going up and down the rows.
5. On several occasions, pupils requested discussion sessions during the reading of the narrative.

The teacher enthusiasm that characterized the meeting at the end of the second week remained undaunted over the ten week period. Chapters on counting and division were rated as most effective, with multiplication a close third. In the teacher's opinion, "tall multiplication" and "tall division" were extremely valuable to the pupils.

The distinguishing feature of the class (8) treatment of the materials was that the teacher strongly insisted on considerable drill with each of the methods given for the basic operations. At the end of the program she asked the pupils which methods they liked best, with the following results:

1. "Tall multiplication" was considered a better method than "long multiplication" by a majority of the pupils.

2. The class was divided evenly in choosing between "tall division" and "long division".

3. The "fix-the-answer-later method" for addition was soundly rejected by the class.

Evidence of her insistence on work with different algorithms was provided by the fact that class (8) was the only class that displayed both "tall multiplication" and "tall division" on the posttest, part III.

In the teacher's opinion, expressed after the completion
of the program, the program was superior to traditional approaches in raising achievement levels; was much more enjoyable for both the teacher and the pupils; and considerably reduced the amount of effort required to manage the class.

Comments Based on Post-Program Interviewing of the Teachers of the Eight Classes

At the conclusion of the program the investigator conducted slightly formal interviews with each of the teachers, making a special effort to discuss matters which seemed to be of concern or interest on the basis of eavesdropping and the more informal sessions with the teachers. Reported in the following paragraphs are those statements or feelings which received agreement from six or more of the teachers.

Six or more teachers agreed to each of the following as points of weakness or as troublesome elements in the program:

1. More problems should be included which demand that the pupil be able to identify steps in algorithms.
2. Drill problems should be included under the same binding with the rest of the student workbook.
3. Both the chapter on addition and the one on multiplication tended to "drag" a bit. If possible, both should be streamlined.
4. The program would be more effective if it were used
a grade of two earlier, shortly after pupils had completed work with the four basic operations on non-negative integers.

5. It would be helpful if the teacher's manual included some historically accurate discussions of the evolution of the concepts treated by the narrative.

Six or more of the teachers agreed to each of the following as identifying strong points of the experimental program:

1. The provisions in the unit for an avoidance of anxiety-producing stimuli were effective. Unexpectedly effective were the class discussion questions which asked pupils to evaluate methods for presenting content or which asked pupils to hypothesize about the thinking of others.

2. Classroom managerial problems were fewer than would be expected with the use of a regular textbook. (Female teachers seemed to emphasize this point more strongly than did males.)

3. The phenomenon of pupils volunteering to read aloud in class was the most surprising observation associated with the use of the experimental program.

4. The vocabulary was often too difficult for some pupils' reading abilities, but it would be ill-advised to risk losing the attraction of the story
by lowering the vocabulary. (Some of the teachers felt it impossible to find a vocabulary level which would be comfortable to all the pupils.)

5. Introducing the operations of multiplication and division by beginning with difficult (or at least long) problems was seen as an effective technique both toward forcing some of the pupils to look at more difficult problems and toward relieving the fears some pupils had of long problems.

6. The unit, while being more enjoyable than the regular text, was also a more effective instructional device than more traditional textbooks.

Comments on the Attainment of the Ten Goals Stated for the Experimental Program (See Chapter 2)

It would be pretentious to suggest that eavesdropping and rather informal interviews with the teachers provided respectable measurements for the ten goals stated for the experimental unit. On the other hand, the experimentor holds the opinion that some evidence of different variables' effectiveness, or ineffectiveness, may be gleaned from such methods for seeking information. Again, readers not in sympathy with this opinion are free to regard the following as "a striving after wind."

Controlling "Closure"

The most apparent ally to controlling closure is "the
search for a better method", enhanced, in the opinion of
most teachers, by beginning with difficult problems in
first presenting a topic. In the investigator's opinion
the controls for closure were effective, but not complete­
ly so. Interviews with two of the classes and with their
teachers after the program was completed would indicate
that less than half of the pupils showed evidence of less
over-compartmentalization of learning than before. These
pupils were also aware of the changes within themselves.
However, in spite of limited success, the opinion held is
that the program more effectively controls closure than
would a regular text.

Controlling "Pacing"

During the first four weeks of the program, all but one
of the teachers apologized at least twice for not seeming
to be able to move through the materials more quickly. Al­
most all teachers expressed concern at one time or another
over how far the other classes had gotten and in what time
period. The investigator usually dismissed such questions
by insisting that his concern was to have each teacher pick
a pace agreeable to the particular class.

Two explanations for such teacher feelings seem reason­
able. First, it may be that the provisions built into the
unit were singly cause for the slowing of the pace. Second,
it may be that the variety in presentation and the class
discussion questions sharpened teacher awareness of pupil
deficiencies regarding "elementary" content. At any rate, whether by design or by accident, the program did appear to have a degree of control over pacing.

Further evidence of such control is offered by the fact that all teachers said that they spent far more time with the non-negative integers than they would have spent using a regular text, and that the additional time was often spent on concepts they might have assumed the pupils to possess.

An alternative explanation, ever present to this study, is that teacher and pupil behaviors stem from the effects of "new" materials or differential treatment.

Avoiding Anxiety-Producing Stimuli

Both "safety in numbers" via class discussions and invitations to role-playing were seen by all teachers as effective means toward relieving pupils of unpleasantries of the past. Most teachers held the opinion that pupils would have resisted discussing topics like "counting" without provisions similar to those built into the unit.

Encouraging Emotional Involvement

For whatever it may or may not contribute toward achievement in mathematics, the experimenter concludes that the emotional involvement displayed by pupils during the program was both qualitatively and quantitatively different from what would typify a trek through yet another standard text. Supported by statements of most teachers and by eavesdropping, it is suggested that such involvement was both more fervid
and less ego-related than would be expected for a more traditional approach to the content. Much of the emotional involvement was evidenced in relation to pupils evaluating the unit's presentations as "stupid" or as "clever",...the type of involvement that simply does not come easily to being confronted with a cold explanation followed by two pages of problems.

Regarding pupils' emotional involvement, the most surprising phenomenon to the investigator is that pupils who saw the hero's methods as "stupid" did not subsequently reject the story.

Entertaining

No teacher saw the experimental program as being less entertaining than a regular textbook.....In fact, not one of the teachers saw regular textbooks as being entertaining at all, while each of the teachers saw the experimental program as being "funny" at least part of the time.

Dealing With the Need for and Danger in Repetition

It is the opinion of the investigator that neither the informal interviews with teachers nor the tests chosen for the study are endowed with sensitivities adequate toward evaluating the effectiveness of the program's design with respect to managing repetition.

Controlling Ego-involvement

Already discussed on the preceding page under "the
encouragement of emotional involvement", teachers seemed to agree that the two strategies for lowering ego-involvement were effective. The problem in trying to determine the degree of and consequences of ego-involvement is that one must deal in symptoms. The study described in this context clearly lacks the sophistication that would be required to match symptoms with causes. On the other hand, in keeping with the liberties already taken in this chapter, the investigator suggests that efforts to lower threats to self-esteem without losing emotional involvement of a less ego-rooted nature were effective.

Managing Variety in Presentation and in Task

As stated in chapter II, one goal of the experimental unit was to provide ample variety in presentation without compromising the task at hand. As with the discussion of "the need for and danger in repetition," no effort will be made to define the role played by variety in changing achievement scores.

It may be in order, however, to mention that the two teachers participating in the program who approached being "variety advocates" saw the variety displayed by the unit as adequate.

Providing for Individual Differences.

Although moderate provisions for dealing with individual differences were incorporated into the unit, very little effort was made by any of the teachers to accommodate such
differences. This observation does not intend to say that the teachers were not concerned with such differences,.... since it is likely that placing one teacher with over twenty pupils in a given room at a fixed time each day is just not favorable to making provisions for individual differences, especially where managerial problems abound.

Providing for Pupil Interactions

In classes where teachers allowed it, as most did, there was an abundance of pupil interaction,...some of it dealing with the task at hand, and some upheld by pupils as an end unto itself. Class discussion questions that invited pupils to evaluate methods of presentation were, according to the teacher, most conducive to pupil interaction.

Several of the teachers mentioned that explanations offered by pupils to their peers were sometimes "awful," but none of the teachers held the opinion that the pupils subjected to such rounds of pedagogical fumbling were worse off for the experience. To the contrary, three of the teachers saw value in such interactions,...enough to sometimes appoint one of the pupils as teacher of the day, to sometimes have pupils work in small groups over the drill exercises, and to sometimes insist that any pupil who had completed his assignment should then appoint himself mentor to someone less fortunate than himself.

Miscellany

In the opinion of the investigator, the most surprising
feature of the program was the report by most of the teachers that pupils wanted to read aloud in class. The degree of surprise comes largely from the fact that seventh grade modified classes are formed primarily on the basis of pupils' reading deficiencies. No explanation will be offered....The phenomenon remains a puzzlement to the investigator.

A second phenomenon, not anticipated, is that women teachers seemed more enthusiastic about the materials at the conclusion of the program than did men. This may simply indicate that women are more comfortable with reading stories to children than are men. An alternative explanation is suggested by the fact that the women teachers more strongly attached significance both to the program's facilitating classroom management and to its apparent ability to help pupils remember to take their books along to class....The alternative explanation then is that the program suppresses behaviors that provide managerial problems for women and not (or less so) for men.

A suggestion appeared to be made by comparing the patterns of teacher enthusiasm and pupils' responses to the "ideal self-concept" items of psychological scales. The only statistically significant drop over these items from pretest to posttest was provided by the only class whose teacher lost enthusiasm for the program over the ten-week period. Looking at patterns in other classes, even though none of the pretest-posttest changes were statistically
significant, further suggested to the investigator that "ideal self-concept" in pupils may be largely a function of the teachers apparent commitment to the task at hand,... ....and may be somewhat independent of achievement, responses on other psychological scales used in this study, or the extent of gain in mathematical achievement.

Since the effect of "new" material or differential treatment may be a strong factor in this study, it may be in order to declare an opinion regarding its possible usefulness to education....If it can be a factor, it would seem reasonable for educators to consciously exploit effects whenever possible. (Except, of course, when the task at hand is basic research.)

Such practices are likely to clutter attempts to carry on "clean" studies. Some obvious difficulties to employing such effects are the following:....It may be that a change in environment (or treatment) will not be noticed by pupils; it may happen that announcements that a treatment is unique or different will not be believed; and it may happen that differential treatment will be perceived when no attempt to provide it is made. In other words, determining the presence of such effects, whether or not their presence was intended, may require the use of a polygraph recorder,.....and the use of such a machine is so likely to help the pupil decide whether or not he is being treated differentially.

Perhaps related to the above is the problem of "fadism".
Where (if) such effects are a main, but unrecognized factor, the chances are present that men of honor will advocate methodology which depends almost exclusively on prior practices for its effectiveness; that is, depends on prior practices for its being perceived as "new" or different....Any "differential treatment", practiced over time, is likely to become commonplace.

This warning seems in order for anyone who might be tempted to conclude that the methods of content presentation employed in this study should replace the more traditional textbook format......A plainsman weary of the mid-west's freeways is likely to find the hair-pinned crossing of the Rockies an emotional, alerting, and instructional venture......Natives who travel them often may find them a frustrating bottleneck; they may even take to flying.
CHAPTER VI
SUMMARY STATEMENTS

This final chapter is divided into two parts, the first dealing with data-based statements and the second dealing with statements derived from the less formal feedback. The first part may be seen as a summary of chapter IV and the second part as a summary of chapter V.

Data-Based Statements

One of the outstanding characteristics of the study was the degree of and nature of the experimental mortality. Of the eight classes participating in the experimental program: 54 subjects took the pretest but were not available for the posttest; 43 subjects not in the classes at pretest time took the posttest; and 129 subjects took both the pretest and the posttest. The pretest mean on the mathematics tests for those who took only the pretest was 10.76; the posttest mean for those who took only the posttest was 10.70; and for those who took both tests the pretest mean was 12.0 and the posttest mean was 15.8.

A more detailed discussion of the experimental mortality appears in the early pages of chapter IV. The main point, in the opinion of the investigator, is that the nature of the mortality tends to portray gain scores as lower than
might be expected with more stable class enrollments.

In the following pages, the terms "claim" and "suggest" will be used to indicate varying degrees of commitment to statements. In general, a "claim" will identify a statement that the investigator would be willing to defend rather tenaciously, while a "suggestion" will identify a statement which the investigator would be inclined merely to believe in relative silence,....with possible reservations.

The Pretest-Posttest Comparison

Regarding the psychological scales, included in the testing with very limited purposes, the investigator concludes that the experimental program did not lower "attitudes" toward mathematics. The data suggests changes in a positive direction from pretest to posttest on the psychological scales, especially over the items under "easy vs. hard" and over those under "fun vs. dull". However, in view of both inadequate comparative data and of deficiencies in research design, the investigator refrains from allowing "acceptable" levels of statistical significance to prompt him to make strong statements. (For a more detailed treatment, see pages 86 through 98.)

Regarding the mathematics tests, achievement gains are claimed for each of the three tests: test II, the "open sentences" test; test III, the "computations" test; and test IV, the "algorithms" test.

In the opinion of the investigator, gains on test II
may be questioned; but gains on test III and test IV, those more directly related to the immediate concerns of the experimental program, are claimed based upon convincing objective data, ... by either looking at individual classes, or at an inner-city and outer-city breakdown, or at classes as units of sampling.

To the extent that the study described by this paper might be thought of as a "feasibility study", the data firmly establishes the materials and methodologies of the experimental program as a candidate for further investigation.

The primary weakness, of course, in looking narrowly at gains from pretest to posttest is the lack of comparative data. The following two sections of this chapter address themselves to that weakness. (For a more detailed treatment of "the pretest-posttest comparison" study, the reader is referred to pages 86 through 98 of chapter IV.)

The Experimental-Control Study

Regarding the psychological scales, the data encourages several suggestions, perhaps no firm claims. The suggestions are as follows:

1. The experimental program is considered by pupils to be both "more fun" and "easier" than the more traditional control treatment.

2. The experimental treatment tends to raise "ideal self-concept" more than does the control treatment, and the control treatment tends to raise "actual
self-concept" more than does the experimental treatment.

The above two statements should be viewed as candidate hypotheses for future study. They are included with the summary statements for two reasons. First, they are suggested by the data derived from this study. Second, they seem very reasonable to the investigator. (For further and more detailed discussion of the above, see pages 89 through 107 of chapter IV.)

Regarding the mathematics test, the following claims are made for the experimental-control study:

1. The experimental treatment is not less effective than the control treatment relative to any of the three mathematics tests.

2. The experimental treatment is more effective than the control treatment relative to test IV, the "algorithms" test.

3. The experimental treatment is more effective than the control treatment relative to the composite of the three mathematics tests,...mainly because of its effectiveness relative to test IV.

As mentioned before in this paper, there is the possibility that the apparent effectiveness of the experimental program is actually an effect of "new" materials or differential treatment. Detecting such intermediate variables is
beyond the power of the research design of this study, and is therefore willed to the future. (For a more detailed discussion of the above claims, see pages 98 through 107 of chapter IV.)

The Comparisons with NLSMA Data

Statements made in conjunction with "the comparisons with NLSMA data" are not based on levels of statistical significance for prestated hypotheses, but are rather ex-post-facto in nature. The license to make such comparisons is seen as implied by the preceding two treatments of the research results. (No attempt was made to make comparisons for results on the psychological scales, for reasons outlined earlier in this paper.)

Relative to test II, the "open sentences" test, included more as a check for "transfer" than as a direct concern of the experimental program; three suggestions are made:

1. The pupils of the seventh grade "modified" classes appear to be at least two years behind chronological age in achievement level.

2. The responses to test II do not indicate that the seventh graders' responses are qualitatively different from patterns displayed by NLSMA.

3. Relative to test II, the seventh graders appeared to have gained in achievement level by more than one half a grade level over the ten-week period of
the experimental program.

Relative to test III, the "computations" test, the following three suggestions are made:

1. "Under-achievement" is more clearly evidenced as task complexity increases. That is, there appears to be a qualitative (as well as a quantitative) difference between response patterns yielded by "under-achievers" and those yielded by more "normal" pupils.

2. The experimental program appears to be effective in dealing with both the qualitative and quantitative differences between response patterns yielded by "under-achievers" and those yielded by "normal" pupils. That is, the tendency for under-achievement to be more clearly visible as task difficulty increases was less evidenced by posttest results than by pretest results.

3. Relative to test III, the seventh graders appeared to have gained in achievement level by more than one half a grade level over the ten-week period of the experimental program.

Relative to test IV, the "algorithms" test, the following suggestions are made:

1. "Underachievers" do very poorly on test IV, perhaps for two reasons. First, they appear to be very vulnerable to the allurement of distracters.
Second, they experience difficulty with the reading demanded by the items of test IV.

2. More than with the other two tests, "underachievement" shows itself most convincingly on test IV. The data does not allow for pronouncements on possible qualitative differences between "under-achievers" and "normal" pupils, mainly because performance was too low to allow for such comparisons. Also, because of the low scores on test IV, it is impossible to describe gains of a grade level of performance.

A more detailed treatment of "the comparisons with NLSMA data" is provided by pages 107 through 126 of chapter IV. Correlations with Reading Scores

In general, correlations between reading scores and scores from the mathematics tests were very low; for the most part, in view of the sample's probable failure to meet the assumptions of the Pearson "r", too low for allowing pronouncements of any kind. The apparent exceptions are for correlations between test IV (both pre- and post-) and reading (both vocabulary and comprehension).

Although not strongly, the data does suggest that correlations for test IV and the reading test were higher at posttest than at pretest for reading vocabulary, and perhaps for reading comprehension. (For a more detailed discussion,
see pages 126 through 129 of chapter IV.)

Concluding Remarks

The investigator saw no deficiencies in the experimental program that would tend to rank it as inferior to a more traditional presentation of the same content to the given population. Contrariwise, the experimental program, whether of itself or by employing the possible effects of "new" or differential treatment, is seen to be a superior treatment relative to the mathematics tests employed. The primary factor in this effectiveness is seen to be in relation to test IV, the "algorithms" test, which could be viewed in part as a test of "understanding". Of particular interest regarding the above claim is that the experimental program is very heavily "verbal", both because of the lengthy narrative and because of a dependence on interaction between pupils via "class discussion" exercises.

The above claim may appear to be in direct contradiction to statements by some that under-achievers need a nearly "word-free" mathematics curriculum. In the opinion of the investigator, however, there is perhaps not a genuine contradiction. Regarding the apparent contradiction, a distinction needs to be made between "free verbalization along with stimuli presentation" and "precise verbalization intended for efficient communication of higher-order concepts". This second type of verbalization is almost completely ignored by
the experimental program, more so than in some materials which intend to lighten the verbal load.

A concern to any research effort in schools is that of generalizability. Looking narrowly at the guidelines for respectable sampling, no claims can be made. However, the investigator holds the opinion that under-achievers outside of Columbus are enough like those of the Columbus sample of eight classes to suggest generalizability.

**Statements Based on Informal Feedback**

A listing of weaknesses and strengths of the experimental program as perceived by the teachers of the eight classes may be found on pages 160 through 162 of chapter V. To the investigator, the most surprising of such perceptions was that there was much volunteering by pupils to read the narrative aloud in class, so much that in several classes the teachers turned to making reading assignments as a control.

A listing of "goals" for the experimental program with comments on attaining those goals may be found on pages 162 through 167 of chapter V. Of those (ten) goals, it is the investigator's opinion, supported by comments from teachers, that the following were among the more clearly attained:

1. Controlling "pacing".
2. Controlling "ego-involvement".
3. Avoiding anxiety-producing stimuli.
4. Providing for pupil interaction.
5. Entertaining.

Although all teachers, with the possible exception of one, were convinced at the program's conclusion that it was more effective in dealing with seventh grade modified classes than more traditional materials and approaches; most were of the opinion that the unit would find its highest "pay-off" as a review unit for fifth or sixth graders in "normal" classes. The following study, therefore, may be in order for the future, assuming agreement to the proposition that the experimental program is a worthy competitor with more traditional approaches....Using the unit which will evolve from revisions based on the feedback from this study, a similar study might be conducted, with the following differences:

1. Samples would be drawn from the late fifth or early sixth grade and from "normal" classes.
2. A more extensive set of psychological scales would be employed....Such items should be more sensitive when used with students enjoying stronger reading skills.
3. Reading tests (vocabulary and comprehension) would be included both with the pretest and with the posttest....Such a testing schedule would also allow looking for a possible effect of the program on reading achievement.
4. A larger control group would be sought to allow for more conclusive information from both the psychological scales and the mathematics tests.

5. If possible, sampling would allow for an unblushing use of parametric statistics. Such samples should be more easily obtained with regular classes than with the smaller number per school of classes for under-achievers.

6. To the degree that the added expenditure of student-hours can be justified, more comprehensive mathematics tests would be employed.

The more ambitious goal for future study would be to test the efficacy of employing an anecdotal style of mathematical content presentation. The problem, however, in establishing or refuting methodologies is that almost any method can be managed well or poorly. It is for that reason that claims and suggestions in this context deal narrowly with "the experimental program" and not with its methodology or format as abstractions, even though the investigator's feelings about them may be strong....There are, after all, few disasters like that of employing poorly-conceived "discovery methods" in an effort to arrive at what is properly a "definition" in mathematics; an anecdotal style of presentation, or any other of the ingredients of the experimental program, could quite easily be used as fruitlessly, or could be used to greater advantage than was the case with the experimental program.
APPENDIX A

THE NARRATIVE FOR THE EXPERIMENTAL UNIT

NOTE: As of March 11, 1971, the copyright for appendixes "A" and "B", registration number "A 186331", is assigned to the Charles E. Merrill Publishing Company of Columbus, Ohio.

Requests for the materials should be addressed to the Senior Editor, Mathematics; Chas. E. Merrill Publishing Co.; 1300 Alum Creek Dr.; Columbus, Ohio 43216
PLEASE NOTE:


UNIVERSITY MICROFILMS
APPENDIX B
THE STUDENT WORKBOOK FOR
THE EXPERIMENTAL UNIT

NOTE: As of March 11, 1971; the copyright for appendixes "A" and "B", registration number "A 186331", is assigned to the Charles E. Merrill Publishing Company of Columbus, Ohio.

Requests for the materials should be addressed to the Senior Editor, Mathematics; Chas. E. Merrill Publishing Co.; 1300 Alum Creek Dr.; Columbus, Ohio 43216
APPENDIX C

PSYCHOLOGICAL SCALES AND
MATHEMATICS TESTS USED
FOR THE STUDY

NOTE: Most test items in appendix "C" are from the NLSMA test batteries. Use of or reproduction of those items requires permission in writing from the Director, SMSG - Cedar Hall, Stanford University, Stanford, Calif.
PART I

There are no right or wrong answers in Part I. Just answer them as honestly as you can.

Below are two sample items:

0. Playing football is fun.

☐□☐□☐□☐□☐□☐□

strongly agree agree mildly mildly disagree disagree strongly disagree

--

00. It is hard to get up on Monday morning.

☐□☐□☐□☐□☐□☐□

strongly agree agree don't know disagree strongly disagree

Decide which one of the ways tells best how you feel about the statement. Then mark the box above your choice with an "x" .... [x]

Work carefully and quickly. Do not spend a long time on any one question. Please answer all items and give only one answer to each.

You will have 8 minutes for this section. There are two pages to this section. Begin when you are told to do so.
1. No matter how hard I try, I cannot understand arithmetic.

[ ] strongly agree [ ] agree [ ] don't know [ ] disagree [ ] strongly disagree

2. Arithmetic is a subject which is more difficult to understand than any other subject.

[ ] strongly agree [ ] agree [ ] don't know [ ] disagree [ ] strongly disagree

3. The subject I enjoy least is arithmetic.

[ ] strongly agree [ ] agree [ ] don't know [ ] disagree [ ] strongly disagree

4. I can not understand how some students think arithmetic is fun.

[ ] strongly agree [ ] agree [ ] don't know [ ] disagree [ ] strongly disagree

5. Arithmetic is boring.

[ ] strongly agree [ ] agree [ ] don't know [ ] disagree [ ] strongly disagree

6. Arithmetic is fun.

[ ] strongly agree [ ] agree [ ] don't know [ ] disagree [ ] strongly disagree

Go on to the next page
7. I feel upset in arithmetic class.

- [ ] strongly agree
- [ ] agree
- [ ] mildly agree
- [ ] mildly disagree
- [ ] disagree
- [ ] strongly disagree

8. I am discouraged with my arithmetic school work.

- [ ] strongly agree
- [ ] agree
- [ ] mildly agree
- [ ] mildly disagree
- [ ] disagree
- [ ] strongly disagree

9. I find it hard to talk in front of my arithmetic class.

- [ ] strongly agree
- [ ] agree
- [ ] mildly agree
- [ ] mildly disagree
- [ ] disagree
- [ ] strongly disagree

10. I like to be called on in arithmetic class.

- [ ] strongly agree
- [ ] agree
- [ ] mildly agree
- [ ] mildly disagree
- [ ] disagree
- [ ] strongly disagree

11. I wish it were easier for me to talk in front of my arithmetic class.

- [ ] strongly agree
- [ ] agree
- [ ] mildly agree
- [ ] mildly disagree
- [ ] disagree
- [ ] strongly disagree

12. I wish I were trying harder in arithmetic.

- [ ] strongly agree
- [ ] agree
- [ ] mildly agree
- [ ] mildly disagree
- [ ] disagree
- [ ] strongly disagree

STOP—DO NOT GO ON UNTIL TOLD TO DO SO
PART II

In each of the questions on the next page one of the numbers is missing. The place where it goes is shown by a box.

In this case you do not have to find the missing number.

You just tell how you would find it.

Below are two sample items:

0. \[48 \div 12 = \square\]

\[\square\text{add} \quad \square\text{multiply}\]

\[\square\text{subtract} \quad \times\text{divide}\]

00. \[\square + 7 = 10\]

\[\square\text{add} \quad \square\text{multiply}\]

\[\times\text{subtract} \quad \square\text{divide}\]

The answer to "0" is "divide".

To find the missing number in "00", you would have to subtract 7 from 10. Here the answer is "subtract".

You will have 3 minutes for this part. There is one page.

Begin when you are told to do so.
1. $71 + 24 = \square$

2. $17 + \square = 64$

3. $13 \times \square = 104$

4. $\square - 23 = 46$

5. $\square \div 12 = 7$

6. $\square + 23 = 655$

7. $12 \times 16 = \square$

8. $144 \div \square = 16$

STOP DO NOT TURN THE PAGE UNTIL TOLD TO DO SO
1. $61 + 34 = \square$

2. $15 + \square = 73$

3. $14 \times \square = 126$

4. $\square - 32 = 65$

5. $\square + 13 = 6$

6. $\square + 31 = 566$

7. $14 \times 12 = \square$

8. $120 \div \square = 15$

STOP DO NOT TURN THE PAGE UNTIL TOLD TO DO SO
PART III

On the next page are addition, subtraction, multiplication, and division problems. Do your work on that page. (If you find that there is not enough room, use the back side of this page.)

You will have 7 minutes to work the problems.

Work neatly. You do not have to re-write the problems.

Begin when you are told to do so.
STOP

DO NOT GO ON UNTIL YOU ARE TOLD TO DO SO
PART IV

In this part there are thirteen questions about computations.

Answer as many of the questions as you can. Do not spend too much time on any one question.

You should only guess if you can rule out some of the choices. DO NOT guess wildly.

You will have 13 minutes for this part.

Begin when you are told to do so.

(If you need space for scratch work, use the back sides of the pages in this section.)

As you did for the other parts of this test, you must put an "x" in the box which matches the statement you choose for your answer. Choose only one answer for each question.
1. Egbert begins a subtraction problem this way:

He finishes it correctly in two more subtraction steps.
In the first of these he could subtract:

(a) 2 from 8
(b) 8 from 12
(c) 1 from 8
(d) 3 from 7
(e) 8 from 11

(Put an "x" in the box near your choice.)

(Questions 2, 3, and 4 refer to the following problem):

In finding the answer to 37 x 52,
The steps could be written as:

step 1 → \[ \frac{37 \times 52}{74} \]
step 2 → \[ \frac{185}{1924} \]

2. In "step 2" 185 is placed where it is to stand for:

(a) \[ 5 \times 37 \]
(b) \[ (1 \times 100) + (8 \times 10) + 5 \]
(c) \[ 30 \times 52 \]
(d) \[ (1 \times 1000) + (9 \times 100) + (2 \times 10) + 4 \]
(e) \[ 50 \times 37 \]
1. Egbert begins a subtraction problem this way:

\[ \begin{array}{c}
\text{He finishes it correctly in two more subtraction steps.} \\
\text{In the first of these he could subtract:} \\
\text{(a) 2 from 9} \\
\text{(b) 9 from 12} \\
\text{(c) 1 from 9} \\
\text{(d) 3 from 8} \\
\text{(e) 9 from 11} \\
\end{array} \]

(Questions 2, 3, and 4 refer to the following problem:

In finding the answer to \(47 \times 62\), the steps could be written as:

\[
\begin{array}{c}
\text{step 1} \rightarrow 47 \\
\text{step 2} \rightarrow 282 \\
\text{step 3} \rightarrow 2914
\end{array}
\]

2. In "step 2" \(282\) is placed where it is to stand for:

\[
\begin{array}{c}
\text{(a) } 6 \times 47 \\
\text{(b) } (2 \times 100) + (8 \times 10) + 2 \\
\text{(c) } 40 \times 62 \\
\text{(d) } (2 \times 1000) + (9 \times 100) + (1 \times 10) + 4 \\
\text{(e) } 60 \times 47
\end{array}
\]
3. In "step 1" 74 is placed where it is to stand for:

(a) $(7 \times 100) + (4 \times 10)$  
(b) $7 + 4$  
(c) $2 \times (30 + 7)$  
(d) $7 \times 4$  
(e) $5 \times 37$

\[
\begin{array}{c}
\text{step 1} \rightarrow \frac{37}{52} \\
\text{step 2} \rightarrow \frac{185}{74} \\
\text{step 3} \rightarrow 1924
\end{array}
\]

4. In "step 3" 1924 stands for:

(a) $74 + 185$  
(b) $185 \times 74$  
(c) $30 \times 50$  
(d) $1850 + 74$  
(e) $37 \times 2$

5. A student made the same error in each of the following subtraction problems:

\[
\begin{array}{cccc}
48 & 35 & 274 & 417 \\
-9 & -9 & -96 & -226 \\
\hline
49 & 36 & 288 & 291
\end{array}
\]

If he makes the same error in 195 - 76 his answer will be:

(a) 121  
(b) 129  
(c) 119  
(d) 185  
(e) None of these
3. In "step 1" 94 is placed where it is to stand for:

(a) \((9 \times 100) + (4 \times 10)\)  
(b) \(9 + 4\)  
(c) \(2 \times (40 + 7)\)  
(d) \(9 \times 4\)  
(e) \(6 \times 47\)

\[
\begin{array}{c}
\text{step 1} \\
\text{step 2} \\
\text{step 3}
\end{array}
\]

\[
\begin{array}{c}
47 \\
\times 62
\end{array}
\]

\[
\begin{array}{c}
94 \\
282 \\
2914
\end{array}
\]

4. In "step 3" 2914 stands for:

(a) \(94 + 282\)  
(b) \(282 \times 94\)  
(c) \(60 \times 40\)  
(d) \(2820 + 94\)  
(e) \(47 \times 2\)

5. A student made the same error in each of the following subtraction problems:

If he makes the same error in \(184 - 65\) his answer will be:

\[
\begin{array}{ccccccc}
48 & 35 & 274 & 417 \\
-9 & -9 & -96 & -226
\end{array}
\]

\[
\begin{array}{ccccccc}
49 & 36 & 288 & 291
\end{array}
\]

(a) \(121\)  
(b) \(129\)  
(c) \(119\)  
(d) \(185\)  
(e) None of these
6. In the problem to the right, \(4\) is placed where it is to stand for:

(a) \(20 \times 2 = 40\)  
(b) \(8 \times 5 = 40\)  
(c) \(2 + 2 = 4\)  
(d) \(10 \times 4 = 40\)  
(e) 4 ones

7. In the problem to the right, \(14\) is placed where it is to stand for:

(a) \(10 + 4\)  
(b) \(8 + 6\)  
(c) \(63 + 87\)  
(d) \(60 + 80\)  
(e) 14 hundreds

8. Look at the work to the right.
This is the work for which of these problems?

(a) \(4 \times 6 \times 50 \times 30 = ?\)  
(b) \(2240 + 168 = ?\)  
(c) \(34 \times 56 = ?\)  
(d) \((8 \times 56) + (60 \times 56) = ?\)  
(e) \(36 \times 54 = ?\)
6. In the problem to the right, 4 is placed where it is to stand for:

(a) \(20 \times 2 = 40\) \(\square\) (a)
(b) \(8 \times 5 = 40\) \(\square\) (b)
(c) \(2 + 2 = 4\) \(\square\) (c)
(d) \(10 \times 4 = 40\) \(\square\) (d)
(e) 4 ones \(\square\) (e)

7. In the problem to the right, 13 is placed where it is to stand for:

(a) \(10 + 3\) \(\square\) (a)
(b) \(8 + 5\) \(\square\) (b)
(c) \(54 + 86\) \(\square\) (c)
(d) \(50 + 80\) \(\square\) (d)
(e) 13 hundreds \(\square\) (e)

8. Look at the work to the right.
This is the work for which of these problems?

(a) \(4 \times 7 \times 50 \times 60\) \(\square\) (a) \(\times ?\)
(b) \(2280 + 342\) \(\square\) (b) \(\downarrow \quad 28\)
(c) \(64 \times 57\) \(\square\) (c) \(\downarrow \quad 200\)
(d) \((8 \times 57) + (120 \times 57)\) \(\square\) (d) \(\downarrow \quad 420\)
(e) \(67 \times 54\) \(\square\) (e) \(\downarrow \quad 3000\)
(Questions 9, 10, and 11 refer to the following problem.)

9. In "step 1" 208 is placed where it is to stand for:
   (a) $4 \times 52$
   (b) $(2 \times 100) + 8$
   (c) $46 \times 52$
   (d) $52 \times 40$
   (e) NONE of these

\[ \begin{array}{c}
52 \overline{2393} \\
-208 \\
\hline
205 \\
\end{array} \]

10. In "step 2" 313 is placed where it is to stand for:
   (a) $2393 - 2080$
   (b) $4 \times 52$
   (c) $239 - 208$
   (d) $6 \times 50$
   (e) $93 - 80$

\[ \begin{array}{c}
52 \overline{2393} \\
-208 \\
-313 \\
\hline
0 \\
\end{array} \]

11. In "step 3" 312 is placed where it is to stand for:
   (a) $46 \times 52$
   (b) $2393 \div 52$
   (c) the remainder
   (d) $6 \times 52$
   (e) a factor of 2393

\[ \begin{array}{c}
52 \overline{2393} \\
-208 \\
-313 \\
\hline
312 \\
\end{array} \]
Questions 9, 10, and 11 refer to the following problem:

9. In "step 1" 106 is placed where it is to stand for:

   (a) $2 \times 53$
   (b) $(1 \times 100) + 6$
   (c) $24 \times 53$
   (d) $53 \times 20$
   (e) NONE of these

   \[
   \begin{array}{c}
   \underline{2} \\
   53 \overline{1273} \\
   \underline{106} \\
   \end{array}
   \]

10. In "step 2" 213 is placed where it is to stand for:

   (a) $1273 - 1060$
   (b) $2 \times 53$
   (c) $127 - 106$
   (d) $4 \times 50$
   (e) $73 - 60$

   \[
   \begin{array}{c}
   \underline{2} \\
   53 \overline{1273} \\
   \underline{106} \\
   \underline{213} \\
   \end{array}
   \]

11. In "step 3" 212 is placed where it is to stand for:

   (a) $24 \times 53$
   (b) $1273 + 53$
   (c) the remainder
   (d) $4 \times 53$
   (e) a factor of 1273

   \[
   \begin{array}{c}
   \underline{24} \\
   53 \overline{1273} \\
   \underline{106} \\
   \underline{213} \\
   \underline{212} \\
   \underline{1} \\
   \end{array}
   \]
12. In the subtraction problem to the right, which of the following is the mathematical description of the "borrowing" or regrouping necessary?

(a) $783 = (6 \text{ hundreds}) + (18 \text{ tens}) + (3 \text{ ones})$
(b) $783 = (6 \text{ hundreds}) + (17 \text{ tens}) + (13 \text{ ones})$
(c) $783 = (6 \text{ hundreds}) + (17 \text{ tens}) + (3 \text{ ones})$
(d) $783 = (6 \text{ hundreds}) + (18 \text{ tens}) + (13 \text{ ones})$
(e) $783 = (7 \text{ hundreds}) + (9 \text{ tens}) + (13 \text{ ones})$

13. A student does a multiplication problem as you see to the right:

\[
\begin{array}{c}
\phantom{000} \\
\phantom{000} \\
\phantom{000} \\
\phantom{000}
\end{array}
\begin{array}{c}
32 \\
\times 43 \\
\hline
120 \\
80 \\
90 \\
\hline
296
\end{array}
\]

The answer is incorrect because:

(a) It is always necessary to start multiplying the units rather than the tens.
(b) The addition is done incorrectly.
(c) The 120 is wrong.
(d) The 90 should not be directly under the 80.
(e) NONE of these.

STOP

IF YOU FINISH BEFORE TIME IS CALLED, CHECK YOUR WORK ON THIS SECTION ONLY. DO NOT WORK ON ANY OTHER SECTION IN THE TEST.
12. In the subtraction problem to the right, which of the following is the mathematical description of the "borrowing" or regrouping necessary?

(a) $672 = (5 \text{ hundreds}) + (17 \text{ tens}) + (2 \text{ ones})$
(b) $672 = (5 \text{ hundreds}) + (16 \text{ tens}) + (12 \text{ ones})$
(c) $672 = (5 \text{ hundreds}) + (16 \text{ tens}) + (2 \text{ ones})$
(d) $672 = (5 \text{ hundreds}) + (17 \text{ tens}) + (12 \text{ ones})$
(e) $672 = (6 \text{ hundreds}) + (8 \text{ tens}) + (12 \text{ ones})$

13. A student does a multiplication problem as you see to the right:

\[
\begin{array}{c}
\phantom{1}0 \\
4 \times 32 \\
\hline
120 \\
\phantom{1}90 \\
\phantom{1}80 \\
\hline
296
\end{array}
\]

The answer is incorrect because:

(a) It is always necessary to start multiplying the units rather than the tens.
(b) The addition is done incorrectly.
(c) The 120 is wrong.
(d) The 80 should not be directly under the 90.
(e) NONE of these.

STOP

IF YOU FINISH BEFORE TIME IS CALLED, CHECK YOUR WORK ON THIS SECTION ONLY. DO NOT WORK ON ANY OTHER SECTION IN THE TEST.
Mr. William H. Nibbelink
3067 St. John's Court
Columbus, Ohio 43202

Dear Mr. Nibbelink:

This is to authorize you to use items from various MSMA test batteries for the purposes stated in your letter of August 21, 1970.

Yours very truly,

E. G. Begle

EGB:ed