



OVERDUE FINES ARE 25¢ PER DAY
PER ITEM

Return to book drop to remove
this checkout from your record.

JUN 01 1987
JUN 11 1987

A SUMMATIVE EVALUATION
OF A JUNIOR HIGH SCHOOL
COMMUNICATION ARTS PROGRAM

By

Leona Mae Wiest Barnhart

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Secondary Education and Curriculum

1979

© Copyright by
LEONA MAE WIEST BARNHART
1979

ABSTRACT

A SUMMATIVE EVALUATION OF A JUNIOR HIGH COMMUNICATION ARTS PROGRAM

By

Leona Mae Wiest Barnhart

Purpose

The purpose of this study was to do a summative evaluation of the Communication Arts Pilot (CAP) program at Traverse City (Michigan) Junior High School. Specifically, the purpose was to determine the effect of the program on the reading and associated abilities of students who were in grades seven to nine, 1972-78. The CAP program was introduced by the Communication Arts Department of the Traverse City Junior High School.

Answers were sought to these questions:

1. Was there a difference between the effect of a new program (CAP) and the program it replaced, as measured by selected standardized tests? Four reading and associated ability areas were analyzed: reading comprehension, vocabulary, spelling, and language.
2. Was the effect of CAP the same for all groups who participated in the program?
3. Did the program have the same effect on students of all reading abilities?

CAP was described as a mastery learning type program with two major components: reading and writing. Students are pre- and posttested on reading and writing skills.

Procedure

Samples for the study were selected from the sixth grade classes of 1970, 1971, 1972, 1973, and 1974. Each sample became one group in the study, except Sample 2 (1971) which became Groups 2 and 6.

Assignment to CAP or non-CAP for Groups 2 and 6 was random. The other groups were CAP (Groups 2, 3, 4, and 5) or non-CAP (Group 1) dependent only on the year the student began seventh grade English.

Tests administered at five different grade levels were used in the study: Stanford Achievement Intermediate II (sixth grade), Gates-MacGinite Reading Test (seventh and ninth grades), and National Educational Developmental Test (tenth grade).

The large design framework of the study was similar to Design 15, the Recurrent Institutional Cycle Design (Campbell, 1963, p. 57-61). Two factors allowed for the use of stronger designs when appropriate with their more powerful statistics:

1. Random assignment of Groups 2 and 6, treatment and control.

2. Naturally occurring X.

Campbell and Stanley (1963) Designs 4, 6, and 10 were used.

Statistics used, where appropriate, were the t-test, analysis of variance, and analysis of covariance. Regression lines were used to indicate trends.

Conclusions

Research Question 1: There was no statistically significance, as measured by standardized tests, between the effect of CAP on the reading and associated abilities of students and the effect had by the program it replaced.

Research Question 2: There was a statistically significant difference between the effect of CAP, as measured by standardized test, on the different groups who participated in the program.

CAP appeared to be the most effective with Group 3, the second group to participate in CAP and the first to have CAP for two years.

Research Question 3: CAP did not have a differential effect for students of varying abilities. The regression lines were positive. The trend was linear.

Summary

Whether CAP should be continued, discontinued, or modified would depend on the goals of the school district. CAP is doing as well as the program it replaced, as measured by

Leona Mae Wiest Barnhart

standardized tests. It does appear to be influencing a decline in the effectiveness of reading comprehension, as measured by standardized tests. CAP does not appear to be influential in changing the relative performance level of students.

DEDICATION

To the wellspring of my joy;

Tom, my husband,

and our children,

Tom, Terry, and LeAnne

ACKNOWLEDGMENTS

Sincere appreciation and gratitude are extended to many who have provided encouragement and support, not only during this study, but throughout my academic career. Among those acknowledged are:

Dr. Peggy M. Riethmiller, committee chair and major professor, for the kindness, encouragement, guidance, and assistance provided throughout my doctoral program.

Dr. Walter W. Scott, committee member, for his unselfish gift of time and self.

Dr. Samuel S. Corl and Dr. Daniel Jacobson, committee members, for their careful and perceptive guidance. Dr. Charles A. Blackman for serving as a substitute at the defense of the dissertation.

Dr. Barbara Aswad, Anthropology Department, Wayne State University, for encouraging me to pursue graduate study.

Iota State, Delta Kappa Gamma Society International for the award of a scholarship to me in 1978 and 1979.

Women of Wayne, Wayne State University, for encouraging women to return to school and for granting scholarships to me in 1967 and 1969.

Traverse City Area Public School District for granting permission to conduct this research. A special thank you to Dr. Vernon L. Oxender, superintendent, R. Edward Johnston, director of personnel and curriculum, Maxine MacInnis, Dr. Richard Ayling, Paul Needham, Shirley Forton, Ella Moore, and Beverly Thornburg.

My family and friends who have provided love and encouragement.

Mary A. Ryan, my mother, for her constant faith, love and guidance. Rochus Wiest, my father, for

his love and belief in me. William J. Ryan for his love and cheerful support.

Mae MacNeill Barnhart and Joseph D. Barnhart, parents-in-law, for their love and support.

Thomas M. Barnhart, my husband, who is without equal. Thomas M. Barnhart II, Terence M. Barnhart, and LeAnne J. Barnhart, our children, who are a joy to my heart and a continual source of encouragement.

TABLE OF CONTENTS

List of Tables	ix
List of Figures	x
List of Graphs	xii
 CHAPTER I: THE PROBLEM	
Background	1
The Communication Arts Pilot (CAP) Program	3
Purpose of the Study	7
Importance of the Study	7
Assumptions of the Study	8
Limitations of the Study	9
Definitions of Terms	10
Frequently Used Abbreviations	11
Overview of the Study	11
 CHAPTER II: REVIEW OF THE LITERATURE	
Introduction	12
Need and Value of Educational Evaluations	12
Summary	15
Summative Evaluation	15
Summary	20
 CHAPTER III: DESIGN OF THE STUDY	
Introduction	22
Restatement of Purpose and Research Questions	22
Summative Evaluation	23
Description of Communication Arts Pilot (CAP)	25
Population	26
Samples	28
Test Instruments	30
Data Needed	34
Data Collection	34
Description of Communication Arts Pilot (CAP)	34
Other Data	35
Tests	36
Design and Analysis	37
Where This Study Expands Design 15	47
Hypotheses	48

Research Question 1	48
Research Question 2	51
Research Question 3	52
Summary	53

CHAPTER IV: RESEARCH FINDINGS

Introduction	56
Research Question 1: Hypotheses and Analysis . .	58
Null Hypothesis 1.1	59
Alternate Hypothesis 1.1	59
Null Hypothesis 1.2	60
Alternate Hypothesis 1.2	61
Null Hypothesis 1.3	62
Alternate Hypothesis 1.3	63
Null Hypothesis 1.4	64
Alternate Hypothesis 1.4	65
Null Hypothesis 1.5	66
Alternate Hypothesis 1.5	66
Null Hypothesis 1.6	67
Alternate Hypothesis 1.6	68
Null Hypothesis 1.7	69
Alternate Hypothesis 1.7	69
Null Hypothesis 1.8	74
Alternate Hypothesis 1.8	75
Null Hypothesis 1.9	80
Alternate Hypothesis 1.9	80
Null Hypothesis 1.10	85
Alternate Hypothesis 1.10	86
Null Hypothesis 1.11	87
Alternate Hypothesis 1.11	87
Null Hypothesis 1.12	90
Alternate Hypothesis 1.12	90
Null Hypothesis 1.13	92
Alternate Hypothesis 1.13	92
Null Hypothesis 1.14	95
Alternate Hypothesis 1.14	96
Research Question 2: Hypotheses and Analysis . .	101
Null Hypothesis 2.1	102
Alternate Hypothesis 2.1	102
Null Hypothesis 2.2	103
Alternate Hypothesis 2.2	103
Null Hypothesis 2.3	104
Alternate Hypothesis 2.3	104
Null Hypothesis 2.4	106
Alternate Hypothesis 2.4	106
Null Hypothesis 2.5	107
Alternate Hypothesis 2.5	107
Null Hypothesis 2.6	108
Alternate Hypothesis 2.6	108
Null Hypothesis 2.7	109
Alternate Hypothesis 2.7	109
Null Hypothesis 2.8	110

Alternate Hypothesis 2.8	110
Null Hypothesis 2.9	111
Alternate Hypothesis 2.9	111
Null Hypothesis 2.10	113
Alternate Hypothesis 2.10	114
Null Hypothesis 2.11	116
Alternate Hypothesis 2.11	116
Null Hypothesis 2.12	118
Alternate Hypothesis 2.12	119
Null Hypothesis 2.13	120
Alternate Hypothesis 2.13	121
Null Hypothesis 2.14	123
Alternate Hypothesis 2.14	123
Null Hypothesis 2.15	125
Alternate Hypothesis 2.15	126
Null Hypothesis 2.16	127
Alternate Hypothesis 2.16	128
Null Hypothesis 2.17	129
Alternate Hypothesis 2.17	129
Null Hypothesis 2.18	130
Alternate Hypothesis 2.18	130
Null Hypothesis 2.19	131
Alternate Hypothesis 2.19	132
Null Hypothesis 2.20	132
Alternate Hypothesis 2.20	133
Null Hypothesis 2.21	134
Alternate Hypothesis 2.21	134
Null Hypothesis 2.22	135
Alternate Hypothesis 2.22	135
Null Hypothesis 2.23	136
Alternate Hypothesis 2.23	136
Null Hypothesis 2.24	138
Alternate Hypothesis 2.24	139
Null Hypothesis 2.25	139
Alternate Hypothesis 2.25	140
Null Hypothesis 2.26	141
Alternate Hypothesis 2.26	142
Null Hypothesis 2.27	142
Alternate Hypothesis 2.27	143
Research Question 3: Hypotheses and Analysis . .	143
Null Hypothesis 3.1	149
Alternate Hypothesis 3.1	149
Null Hypothesis 3.2	150
Alternate Hypothesis 3.2	150
Summary	158
Research Question 1 and Research Findings .	158
Research Question 2 and Research Findings .	159
Research Question 3 and Research Findings .	160

CHAPTER V: CONCLUSIONS AND RECOMMENDATIONS

Introduction	161
------------------------	-----

Conclusions	162
Research Question 1	162
Research Question 2	165
Research Question 3	177
Implications for Traverse City Public Schools	177
Implications for Public Schools in General	180
Recommendations for Further Research	181
Summary	182
Appendix A	183
Appendix B	184
Appendix C	191
Appendix D	206
Appendix E	210
Bibliography	215

LIST OF TABLES

Groups (CAP-nonCAP): Diagram of Treatment and Observations (Tests) of the Study	46
Comparability of Designs	51
Groups (CAP-nonCAP), Tests, and When Administered . .	54
Groups (CAP-nonCAP), Tests, and When Administered . .	58

LIST OF FIGURES

The Relationship Between Evaluation Activities and the Stages of a Developing Program	24
Untitled	77
Untitled	78
Untitled	82
Untitled	83
Untitled	89
Untitled	93
Untitled	95
Untitled	112
Untitled	113
Untitled	115
Untitled	115
Untitled	117
Untitled	118
Untitled	119
Untitled	120
Untitled	122
Untitled	122
Untitled	124
Untitled	125
Untitled	127
Untitled	131

Untitled	138
Untitled	141

LIST OF GRAPHS

4:1.1	70
4:1.2	71
4:1.3	72
4:1.4	73
4:1.5	97
4:1.6	98
4:1.7	99
4:1.8	100
4:2.1	144
4:2.2	145
4:2.3	146
4:2.4	147
2:4.5	148
4:3.1	152
4:3.2	153

CHAPTER I

THE PROBLEM

Background

Questions about the effectiveness of educational programs have been frequent in education. Rudolph Flesch brought to the fore in the mid-1950's the question of effectiveness of programs on the reading ability of students (Why Johnny Can't Read, 1955).

The questioning of educational programs continued when, in 1957, Sputnik appeared. The American public was shocked. The federal government provided funds to change the challenged curriculum. Changes were brought about, primarily in the mathematics and the sciences (Goodlad, 1966, p. 11).

In 1964 the Scholastic Aptitude Test (SAT) and the American College Testing Assessment (ACT) scores, which had been increasing since the early 1940's, began to decline ("Declines in Standardized Test Scores among Secondary School Students," 1976, pp. 4-5). Again, questions were raised about students and programs. Were the high school students of the 1960's and 1970's taking the SAT and ACT tests less able than the students of the '40's and '50's who had taken them? If the students were not less able,

then were the school programs to be considered less effective in areas represented on the examinations?

More books decrying the schools appeared: How Children Fail by John Holt, 1964; Death at an Early Age by Jonathan Kozol, and 36 Children by Herbert Kohn, both in 1967. The nation's children were not receiving the education parents expected.

The public concern about the problem of reading ability of students had grown since Flesch's 1955 publication to the point that Richard M. Petre was to write in the December, 1972, volume of Journal of Reading:

Never before have the mass media reported so many conflicting stories about how "well" or how "poorly" Americans read. According to these reports, Americans appear to believe that they can hardly read, that large numbers of people are illiterate, that reading failures have reached epidemic proportions, and that school systems neglect reading to such an extent that a national crisis exists (p. 200).

Reading as a "subject" was being analyzed by many within the educational community. Publishers, reading specialists, and teaching staffs were three groups who attempted to delineate reading skills. The skills identified by publishers often varied. (For example, see the teacher's edition of More Sounds and Syllables: The Ginn Enrichment Program by Theodore Clymer, Thomas C. Barrett, and Lori E. Burmeister, 1969, pp. 3, 8; and the teacher's edition of Discovering Treasure, 1973, pp. 320-325.) Objectives for identified reading skills were written at all levels of education: federal, state, and local (Gussion, 1979, pp.

1-2). They were frequently placed in a sequential order (Strang, 1967, pp. 131-144).

English programs at the secondary level traditionally had been literature-oriented. English teachers had been trained in literature and composition, and English programs reflected that training (Dunning, 1975, p. 9). Reading as a "subject" was not taught.

The Communication Arts Pilot (CAP) Program

The junior high school communication arts department of a local school district which shifted the format of its program is the focus of this study. Its program was shifted from a traditional format to one emphasizing the teaching of reading and writing skills.

The "traditional" junior high school program may be described as one using a classroom grammar text, a classroom literature text, heterogeneous groupings, and vague objectives. The "new" junior high school program may be described as one using no classroom text, heterogeneous skill groupings, and specific objectives.

The junior high school is a part of the Traverse City Public School District, Traverse City, Michigan. The 1978-79 district enrollment was 9,516 students. There are one senior high, one junior high, and fourteen elementary schools in the district. The junior high school contains grades seven to nine. The 1978-79 enrollment was 2,202 students. Each grade level had approximately 700 plus students.

The junior high department chairperson described the Communication Arts Pilot (CAP) program as a mastery learning type program (Needham, 1979). It has been at various stages of implementation since 1972. In that year the program was introduced and conducted at the seventh grade level by a team of two classroom teachers and one reading teacher. CAP was expanded to the whole seventh grade communication arts program in 1973-74. In 1974 it was expanded to the eighth grade. In 1976 the Ninth Grade Project was started. This was to be a transitional program between CAP and APEX (Appropriate Placement for Excellence), the high school communication arts program. CAP, 1979, is still considered a pilot program. The program has been funded at the local level through Board of Education action.

The department chairperson of the Traverse City Junior High School Communication Arts Department stated two reasons for the development of CAP: to provide data for answering the public concern over declining reading scores and to stabilize or improve student reading achievement (Needham, 1979).

A second explanation for the development of CAP was presented in a 1974 Master's thesis. In September, 1972, it was determined that students in the seventh grade communication arts classes ranged in reading ability from grades two to twelve, as measured by the Gates-MacGinitie Reading Test. Two teachers expressed concern that they were not effectively teaching all the students within the

reading ability range. The program developed from that concern (Needham, 1974).

The rationale for development of CAP provided in a department memo distributed to parents and other interested persons, "CAP--Communication Arts Pilot" (undated), is similar to the explanation reported in the thesis.

Also provided in the memo was a statement of the goal of CAP:

To individualize instruction by:

1. Using pretesting to place students in temporary instructional skill groups, focusing on a particular writing or reading skill within a hierarchy of objectives.
2. Utilizing as many materials and approaches as possible to teach the skill while responding to style of learning and interests.
3. Using post testing (frame sheets) to determine mastery (90%) and placement in the next skill group of the writing or reading design.

Two tests are systematically administered to junior high communication arts students: the Gates-MacGinite Reading Test and the Croft Reading Comprehension Test. Administration of the tests was conducted by the communication arts staff.

The Gates-MacGinite Reading Test has been administered to all students in the seventh and ninth grade English classes since 1870. The seventh grade scores are used by the junior high counselors and the seventh and eighth grade communication arts staff for placement and grouping of

students. The ninth grade scores are used by the ninth grade counselors and the senior high counselors in the development of a student's high school program. The ninth grade communication arts staff uses the ninth grade test scores for placement and grouping of students.

The Croft Reading Comprehension Test is administered in the seventh grade. It is used for student placement into skill groups and as a pre-post test.

The CAP program has two components, reading and writing. The reading skills component was derived from the Croft Reading Comprehension Skills material. The CAP reading objectives are Croft reading objectives. The writing component of the program, on the other hand, is and has been teacher-developed.

Students are pre-post tested on specific skills in reading and writing. Instruction is with large groups (class size is usually thirty-three). Students who do not pass the posttest for a skill may go through a skill cycle again. Students may go through a skill cycle up to three times in attempting to reach mastery. The number of opportunities varies depending on the availability of staff, number of students needing the skill cycle, and the time of year.

A skill cycle has six steps:

1. Performance objectives specified
2. Criterion test administered
3. "Teach" step implemented (attention, presentation, reinforcement, practice) - explain
4. Posttest

5. Application - give a sheet explaining kinds
- length, extent
6. Recycle or on to the next ("Update on CAP
Program," 1978)

Purpose of the Study

The purpose of this study was to do a summative evaluation of the Communication Arts Pilot (CAP) program at Traverse City (Michigan) Junior High School. Specifically, the purpose was to determine the effect of the program on the reading and associated abilities of students who were in grades seven to nine, 1972-78. The CAP program was introduced by the Communication Arts Department of Traverse City Junior High School.

The major research questions explored in this study were:

1. Is there a difference, as measured by standardized tests, between the reading and associated abilities of students who participated in CAP and students who did not?
2. Is the effect of CAP upon different student groups consistent over time, as measured by standardized tests?
3. Is there a difference in the effect of CAP for students of varying reading abilities?

Importance of the Study

Ralph Tyler states that evaluation is an important means of communicating school success to the public:

Increasingly, we must expect to use evaluation procedures to determine what changes are actually taking place in students and where we are achieving our curriculum objectives and where

we must make still further modifications in order to get an effective educational program (1949, p. 125).

The Traverse City Public Schools implemented a communication arts program in the junior high school in 1972. Money has been spent, human resources used, and for seven years students have participated in the program. What changes have occurred in the students? Malcolm Provus (1973) has stated that the "purpose of program evaluation is to determine whether to improve, maintain, or terminate a program" (p. 172). This study will provide data to help make that decision about the Communication Arts Pilot (CAP) program.

In addition, the data and analysis presented in this study may be useful in three other areas of inquiry:

1. Mastery learning as an effective method for improving reading achievement at the junior high school level
2. General studies of junior high school students' reading achievement
3. Impact of educational innovations on student outcomes

Assumptions of the Study

1. The tests used in the study are part of the Traverse City Public School's testing program. They are standardized tests used to measure student achievement, reading achievement, and general aptitude. The school district is familiar with the tests. They have been used since prior

to 1971. It is assumed that school district personnel find the tests acceptable and the results reliable.

2. The tests used in the study were administered by the Traverse City Public School staff. It is assumed the tests were administered in a consistent manner and according to the test developer's instructions.

3. The tests were scored by outside agencies, either the testing companies or the Intermediate School District. It is assumed the tests were accurately scored.

4. The test results were provided to the school district on computer printouts. It is assumed that the printouts were accurate.

5. The test developers have conducted research to determine the content validity of their instruments. It is assumed the research was conducted in an adequate manner. It is also assumed the determination of content validity is accurate.

Limitations of the Study

1. The researcher's lack of control over tests used, administration of the tests, scoring, recording and storing of results are all limitations inherent in an ex post facto study. The limitations are those any district would encounter that wished to do a program evaluation for which they had not preplanned.

2. A study over time must contend with variables which may have a differential effect on the different

groups being studied. Information to help evaluate the impact of three such variables are provided:

a. Demographic changes--demographic data for the Traverse City area have been provided.

b. Differences in educational experiences of the six groups in the study--information about changes in the curriculum and materials of three Traverse City junior high departments--social science, science, and mathematics--have been provided.

c. Impact of history experienced by the six groups in the study--the seventh grade reading test results of the Michigan Educational Assessment Program for the Traverse City Public Schools and six other districts of the northwest region of Michigan are provided. It is assumed life experiences, such as mass media, are similar for the seven districts. It can be observed whether the position of Traverse City students relative to the students in the other six districts remained the same.

Definition of Terms

The following definitions are accepted throughout the study:

Class: all the students in a grade level for a particular year.

Skill Cycle: a teaching procedure with six steps (objectives stated, pretest, teach, posttest, student

application, recycle). Further described on page six of this paper.

Summative Evaluation: a non-interfering outcome evaluation of a stable program to help determine whether a program should be improved, continued, or discontinued.

Frequently Used Abbreviations

CAP: Communication Arts Pilot program

Gates: Gates-MacGinite Reading Test

Stanford or Stan: Stanford Achievement Test Intermediate II

DAT: Differential Aptitude Test

NEDT: National Educational Development Test

Overview of the Study

Chapter II consists of a presentation of the literature which led to decisions made in the study.

Chapter III includes the design of the study with a description of the population, the samples, instruments used, the data needed, collection of the data, statistical procedures, and analysis of the data to be used.

Chapter IV contains an analysis of the data collected.

Chapter V includes a summary, findings, conclusions, implications of the study for Traverse City Junior High School, and recommendations for further research.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The purpose of this study was to do a summative evaluation of a program innovation in a local school district. The review of the literature was directed to answering two questions:

1. Is an evaluation of an educational program perceived by educators as a valuable and needed activity?
2. When is it appropriate to do a summative evaluation?

Need and Value of Educational Evaluations

Many educators perceive evaluation of educational programs as imperative. The following are quotations from educators who speak to this point.

Careful evaluation has not been made of the innovations of the past, nor is it being made today. This failure to assess the effects of innovations against their total outcomes has been perhaps the cause of the fact that in American education curriculum revision proceeds by replacing one scheme with another and one "approach" with another, not necessarily because objective evidence has demonstrated the merits of the one or the failures of the other, but merely because the new scheme or approach somehow has gained attention, is in "fashion" for the time being, or is championed by forceful leaders (Taba, 1962, p. 315).

There has been agreement, both within and without the ranks of education, that systematic investigation has much to offer. Indeed, there is agreement that massive, lasting changes in education cannot be made except on the basis of deep objective inquiry (Cronback and Suppes, 1969, p. 12).

Accurate assessment of educational outcomes is essential for sound planning and effective stimulation of growth in our educational structure. Assessment has always been an integral aspect of curriculum development and is a major responsibility of curriculum workers. This responsibility is especially critical in a time of awakened public concern, massive federal commitment and widespread professional reappraisal of our educational endeavors (Combs, 1967, p. v).

Evaluation is one of the most widely discussed but little used processes in today's educational systems...despite these tacit trends toward accountability, only a tiny fraction of the educational programs operating at any level have been evaluated in any but the most cursory fashion, if indeed at all. Verbal statements about education and accountability? An abundance. Genuine evaluation of educational programs? Unfortunately rare (Worthen and Sanders, 1973, p. 1).

Until quite recently school systems have rarely been equipped or even motivated to attempt to evaluate the effects of their learning and teaching activities. There seems to have been an assumption that an innovation is good per se, because it is new and unaccustomed. This idea is further reinforced if the innovation lasts, if it appears not to be doing a poorer job than the practice it replaced, and if it does not, at the same time, disturb too much the other activities which are in progress in the school (Moorish, 1978, p. 154).

As public programs have become increasingly massive and expensive, it is not surprising that taxpayers and some of those who authorize expenditures of money have begun to ask questions about how much it buys. So, a new enterprise called program evaluation has come on the scene. Even though its record of results is still modest, program evaluation offers enough promise that it is being mandated and incorporated in the planning

for both large and small human-service interventions (Anderson and Ball, 1978, p. 2).

The need for and desirability of objective evaluation seems to have been a consistent concern of educators. The concerns expressed, over a sixteen year time span, have not changed.

What kind of evaluations are considered desirable? Cronbach and Suppes (1969) suggest that the type of investigation needed in evaluation is one of disciplined inquiry.

Disciplined inquiry has a quality that distinguishes it from other sources of opinion and belief. The disciplined inquiry is conducted and reported in such a way that the argument can be painstakingly examined...fundamental to disciplined inquiry is its central attitude, which places a premium on objectivity and evidential test (pp. 15, 18).

Moorish (1978) states that it is impossible to subject educational innovations to the same rigors of scientific investigation found in the natural sciences. He further states that the evaluations of the non-rigorously developed programs also lack rigor.

New educational practices can seldom be justified on a scientific basis before being tried out; and they are certainly not all subjected to careful and detailed evaluation (p. 59).

...if an innovation is really to be seen to be successful, it needs not merely the casual visit of an external observer, but the careful analysis of the outside expert or moderator. A really scientific attempt to evaluate any innovation is certainly a necessity (p. 157).

Summary

It was apparent from the literature that educational evaluation is desirable. It was also apparent that evaluation should be rigorous and objective. It was the intent, in this study, to do a rigorous, objective evaluation of a program innovation.

Summative Evaluation

The types of evaluation methods available were explored in order to select one appropriate for this study. Worthen and Sanders (1973) provide a chart comparing eight evaluation models. They are listed by the name of the person or group associated with each model. Below are the eight authors and the purposes Worthen and Sanders ascribe to each model.

1. Robert E. Stake

To describe and judge educational programs based on a formal inquiry process.

2. Michael Scriven

To establish and justify merit or worth. Evaluation plays many roles.

3. Malcolm Provus

To determine whether to improve, maintain, or terminate a program.

4. Robert L. Hammond

To find out whether innovation is effective in achieving expressed objectives.

5. Daniel L. Stufflebeam

To provide relevant information to decision makers.

6. Marvin C. Alkin

To report summary data useful to decision makers in selecting among alternatives.

7. Accreditation

To identify deficiencies in the education of teachers and students relevant to content and procedures; self-improvement.

8. Ralph W. Tyler

To determine the extent to which purposes of a learning activity are actually being realized (pp. 210-215).

Anderson and Ball (1978) recognize the preceding models but suggest that it is a mistake for an evaluator to slavishly use one pattern. The purpose of the evaluation should determine which evaluation strategy is used. Anderson and Ball provide a list of six evaluation purposes:

1. To contribute to decisions about program installation.
2. To contribute to decisions about program continuation, expansion, and "certification."
3. To contribute to decisions about program modification.
4. To obtain evidence to rally support for a program.
5. To obtain evidence to rally opposition to a program.
6. To contribute to the understanding of basic psychological, social, and other processes (pp.3,4).

Marvin C. Alkin and Carol T. Fitz-Gibbon in Methods and Theories of Evaluating Programs (1975) stated that confusion was present in understanding evaluation methods.

This is particularly true regarding the difference between formative and summative evaluation.

Formative evaluation is directed toward decisions relating to program management whereas summative evaluation is conducted with a little intervention in the program as possible...The theory behind summative evaluation is that it is done to assess the impact of a stable program in order to provide information leading to the decision to continue, discontinue, expand, or curtail a program (p. 7).

The definition provided by Alkin and Fitz-Gibbon is very similar to the second purpose statement in Anderson and Ball. Anderson and Ball (1978) provide a table with the kinds of investigation one might use when performing an evaluation for any one of the six purposes. Experimental and quasi-experimental investigations are deemed appropriate for the second purpose.

...evaluation directed toward determining program impact requires experimental designs that allow such inferences (Anderson and Ball, 1978, p. 5).

Fitz-Gibbon and Morris (1978) in How to Design a Program Evaluation state:

Typically, design has been associated with summative evaluation (p. 11).

Historically, designs were developed as methods for conducting scientific experiments... Since designs serve the interest of producing defensible results, and...such production is primarily the interest of the summative evaluator (p. 12).

Moorish (1978) makes a strong statement regarding experimentation:

The precise manner of evaluating any innovation is by trying it out on an experimental basis

and by comparing the results with those of a control group which is not using it (p. 157).

Moorish does suggest that while experimentation may be viewed by some as a questionable activity, it is essential.

Experimentation, it is often suggested, is quite all right on things...but not on people, particularly children. And, moreover, if the projected change is not a good or successful one --whatever the connotation given to those words --then the children may have suffered "untold harm" as a result. It is never suggested how any innovation in school can be introduced without actually experimenting in some way upon children, yet any other form of experimentation would be unreal and incapable of any viable evaluation or assessment (p. 57).

Alkin and Fitz-Gibbon (1975) in their definition of summative evaluation, stated that impact evaluation should be conducted on a stable program. The purpose is to avoid the "Hawthorn" effect. As Morrish (1975) states:

...type of problem with experimental evaluation is what is variously known as the "halo," "Hawthorne," or "placebo" effect. This...is the fact that the conditions which surround any experiments--particularly those involving social and human behaviors--tend to distort the results ...Both teachers and pupils will tend to perform better, administration will be avidly interested in the project, and more amenable classroom arrangements will be provided. Probably, therefore, the best time really to assess an innovation is when it ceases to be an innovation and when the project can no longer evoke special enthusiasms, energies or resources (p. 158).

Anderson and Ball (1978) state that test instruments are appropriate for the designs used with the second purpose listed earlier. Lee J. Cronbach in "Course Improvement Through Evaluation" (1963), reprinted in Worthen and Sanders (1973), states his view regarding measuring instruments:

In course evaluation, we need not be much concerned about making measuring instruments fit the curriculum. However startling this declaration may seem, and however contrary to the principles of evaluation for other purposes, this must be our position if we want to know what changes a course produces in the pupil...If you wish only to know how well a curriculum is achieving its objectives, you fit the test to the curriculum; but if you wish to know how well the curriculum is serving the national interest, you measure all outcomes that might be worth striving for (p. 54).

Anderson and Ball (1978) refer to a "global need."

There is controversy, as implied in Cronbach's statement, about whether two curricula with different purposes can be compared. Scriven (1967) maintains that it is valid to compare different curricula.

Certainly there are cases when the most meaningful question to be addressed in evaluation is whether one treatment or delivery system accomplishes a certain effect better than an alternative system--especially if the systems differ profoundly in content or in such factors as ease of execution, popularity, or potential for negative side effects (Anderson and Ball, 1978, p. 28).

The questions of "what" and "how" of evaluation seem to be dependent on many factors. Some of these are:

1. The purpose of the evaluation
2. The questions being asked in the evaluation
3. The use to which the evaluation information will be put

Anderson and Ball (1978) in their discussion of trends in evaluation mention one trend which they even hesitate to refer to as a trend. They view it as an area of evaluation that could potentially provide valuable information to decision makers. They refer to the methodology as "person

treatment interaction." It is an analysis of the differentiation effect of a program. Very little work has been done in this area. Anderson and Ball suggest that this kind of evaluation, although complex, may not be pursued because it may not be viewed as "good" to allow differentiation in operating programs.

Summary

The purpose of this study was to do a summative evaluation of a program innovation. In this study the program innovation was a new junior high school language arts program.

The literature suggests that a summative evaluation should utilize a "design," similar to ones used in scientific research. The data used should be objective; test scores are deemed appropriate. The program should be stable when it is evaluated, in order to avoid the "Hawthorne" effect. It might be valuable to determine the differentiating effect of a program.

This study attempts to utilize each one of these suggestions. A design was used. Objective data from standardized tests were used. The program being evaluated was begun in 1972 and, therefore, may be considered stable. An attempt was made to determine the differentiating effect of the program.

The literature implied that program evaluation is an essential component of education. In this study the rigors

of objective evaluation have been adhered to as closely as possible.

CHAPTER III

DESIGN OF THE STUDY

Introduction

The design of the study is the focus of Chapter III. The purpose of the study and the research questions will be restated here, and a description of the Communication Arts Pilot (CAP) program will be provided. The population, samples, and instruments used will be described. The data needed and the collection of the data will be outlined. The design, hypotheses, and treatment of the data will be explained.

Restatement of Purpose and Research Questions

The purpose of this study was to do a summative evaluation of the CAP program at Traverse City (Michigan) Junior High School. Specifically, the purpose was to determine the effect of the program on the reading and associated abilities of students who were in grades seven to nine, 1972-78. The CAP program was introduced by the Communication Arts Department of Traverse City Junior High School.

The research questions explored in this study were:

1. Is there a difference, as measured by standardized tests, between the reading and associated abilities of students who participated in CAP and students who did not?

2. Is the effect of CAP upon different student groups consistent over time, as measured by standardized tests?

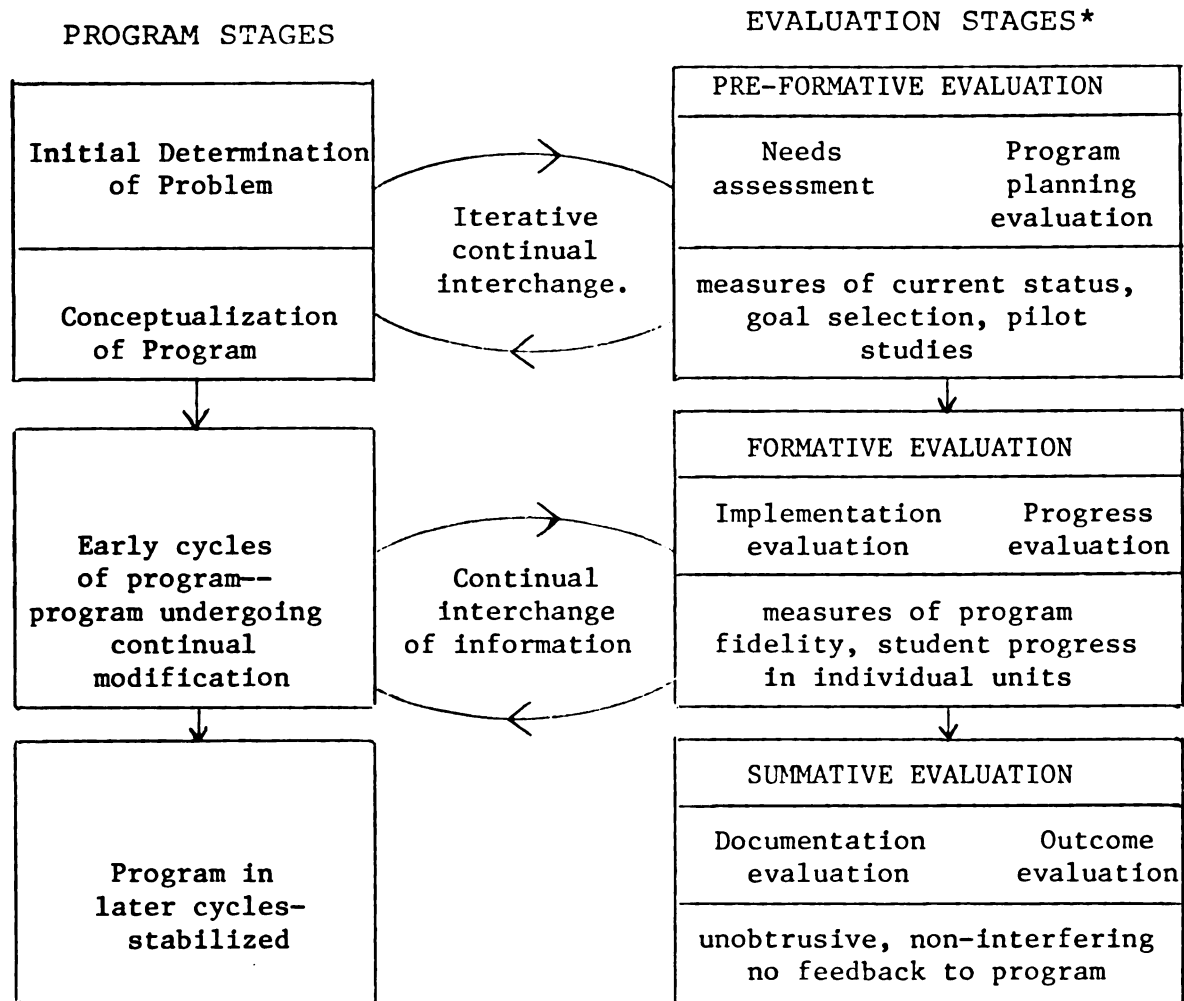
3. Is there a difference in the effect of CAP for students of varying reading abilities?

Summative Evaluation

The definition of summative evaluation provided by Alkin and Fitz-Gibbon (1975, pp. 4-5, 7) describes the type of evaluation conducted in this study with the emphasis on outcome evaluation.

- unobtrusive, non-interfering, no feedback to program (see Figure 1, Alkin and Fitz-Gibbon, 1975, p. 7).
- performed when a program must be judged as a whole in order to decide to drop it or continue it.
- conducted with as little intervention in the program as possible. Ideally, summative evaluation is unobtrusive, and at least outcome data are collected via the tests which will be routinely employed in the program. The theory behind summative evaluation is that it is done to assess the impact of a stable program in order to provide information leading to the decision to continue, discontinue, expand, or curtail a program. It is important that summative evaluation is not undertaken "until a program is stable."
- there are two distinct evaluation activities during the conduct of a summative evaluation: documentation and outcome evaluation. Documentation involves collecting information which shows how the program was actually implemented ...outcome evaluation, which measures the results or outcomes obtained by the program (Alkin and Fitz-Gibbon, 1975, pp. 4-5, 7).

Figure 1: The relationships between evaluation activities and the stages of a developing program



*At each stage of the evaluation process, the following four phases occur:

- (1) ascertaining the decision areas of concern
- (2) collecting and analyzing data in order to
- (3) report summary information to decision makers

Description of
Communication Arts Pilot (CAP)

The Communication Arts Pilot (CAP) program has been described by the junior high department chairperson as a mastery learning type program (Needham, 1979). It has been at various stages of implementation since 1972. That year it was introduced and conducted in the seventh grade by two classroom teachers and one reading teacher. The following year, 1973-74, the whole seventh grade communication arts program was CAP. In 1974 CAP was expanded to the eighth grade.

CAP has been funded by Traverse City Board of Education action. In the 1978-79 school year, CAP was still considered a "pilot" program, according to the department chairperson. A department paper showing the implementation process may be found in the appendix.

The goal of CAP ("CAP--Communication Arts Pilot," undated) is:

To individualize instruction by:

1. Using pretesting to place students in temporary instructional skill groups, focusing on a particular writing or reading skill within a hierarchy of objectives.
2. Utilizing as many materials and approaches as possible to teach the skill while responding to style of learning and interests.
3. Using post-testing (frame sheets) to determine mastery (90%) and placement in the next skill group of the writing or reading design.

The CAP program has two components, reading and writing. The CAP reading objectives are those provided in the Croft Reading Comprehension Skills material (see Appendix). The writing component of the program is and has been teacher-developed.

Students are pre- and posttested on specific reading and writing skills. Instruction is in large groups (class size is usually thirty-three). Students who do not pass a posttest for a skill may go through a skill cycle up to three times in attempting to reach mastery. The number of opportunities varies depending on the availability of staff, numbers of students needing the skill cycle, and the time of year.

The skill cycle has six steps:

1. Performance objectives specified
2. Criterion test administered
3. "Teach" step implemented (attention, presentation, reinforcement, practice) - explain
4. Posttest
5. Application - give a sheet explaining kinds - length, extent
6. Recycle or on to the next ("Update on CAP Program," 1978)

Population

The population of the study included three categories of Traverse City Junior High School students:

1. Those who had participated in the Communication Arts Pilot (CAP) program from September, 1972, to June, 1977.
2. Students of seventh grade English teachers who did not participate in CAP in 1972-73, the first year of implementation of CAP.

3. Students in the seventh grade English classes of 1971-72, one year prior to any implementation of CAP.

Students who had reading improvement in seventh and/or eighth grade of junior high school during the years spanned by the study were not included in the population. The Reading Improvement program was a separate English program prior to and after implementation of CAP; therefore, the exclusion does not make the CAP and non-CAP groups different.

The population of the study is assumed to be normally distributed. The five sixth-grade classes from which the samples were drawn all have a large N:

1970	N = 644
1971	N = 668
1972	N = 717
1973	N = 714
1974	N = 691

A choice was made not to follow students past grade ten for two reasons. Grade ten is the last grade in the Traverse City Public Schools when a test is administered to all students. Tests are administered in the eleventh and twelfth grades, Preliminary Scholastic Aptitude Test (PSAT), Scholastic Aptitude Test (SAT), and American College Testing Assessment (ACT); but they are taken primarily by students planning to attend college.

The second reason for not including grades eleven and twelve is the student drop-out problem. Students must attend school until they are sixteen years of age. For most

students sixteen years is reached during the tenth grade. At that time, students may drop out of school if they choose. Since this study was of a junior high school program, as many students as possible who had participated in the program should be included in the population from which the samples were drawn. If grades eleven and twelve were included, high school dropouts would have to be eliminated from the population of the study. For this study they were not eliminated. However, they were excluded from the study if they left school before being administered all of the test instruments being used in this study.

Samples

Initially five samples were selected from the population. A systematic sampling procedure was used. The first individual was selected at random using a random number table in Elementary Statistical Methods by Helen M. Walker and Joseph Lev (pp. 372-4). Subsequent individuals were selected at regular intervals. This is a standard procedure for selecting a sample from a list where there is no reason to believe that a bias exists in the arrangement of the list. The lists were in alphabetical order, and there was no reason to assume there was bias.

The lists from which the sample selections were drawn were the Traverse City Public Schools' computer printouts of the test results of the Stanford Achievement Test Intermediate II. The test was administered to all the sixth grade

students in the Traverse City School District in 1970, 1971, 1972, 1973, and 1974. 150 individuals were selected from each list.

Individuals were eliminated from a sample if they (1) did not have complete test scores for the five tests used in the study, and/or (2) had reading improvement in seventh and/or eighth grade.

Sample 1 was selected from the 1970 sixth grade class of Traverse City Public Schools. It is referred to as Group 1. None of the students participated in CAP. It is a control group. N=87.

Sample 2 was selected from the 1971 sixth grade class of Traverse City Public Schools. It was further divided into two groups: those students who had had CAP and those who had not. The group which participated in CAP is referred to as Group 2. N=43. The group which did not participate in CAP is referred to as Group 6. N=52. Group 6 is a control group.

Sample 3 was selected from the 1972 sixth grade class of Traverse City Public Schools. It is referred to as Group 3. All students participated in CAP. N=87.

Sample 4 was selected from the 1973 sixth grade class of Traverse City Public Schools. It is referred to as Group 4. All students participated in CAP. N=87.

Sample 5 was selected from the 1974 sixth grade class of Traverse City Public Schools. It is referred to as Group 5. All students participated in CAP. N=87.

The groups included in the study were:

<u>Group</u>	<u>Description</u>	<u>Years</u>	<u>Status</u>
1	Control group; 1970 6th grade	1970-74	Graduated 1977
6	Control group; 1971 6th grade	1971-75	Graduated 1978
2	1971 6th grade 7th grade CAP	1971-75	Graduated 1978
3	1972 6th grade 7th & 8th grade CAP	1972-76	1978-79 12th grade
4	1973 6th grade same as #3	1973-77	1978-79 11th grade
5	1974 6th grade same as #3	1974-78	1978-79 10th grade

The N of each group is large:

Group 1	N = 87
Group 2	N = 43
Group 3	N = 87
Group 4	N = 87
Group 5	N = 87
Group 6	N = 52

...if N is reasonably large, the sampling distribution will tend to be normal; and this will be true even if the samples come from a decidedly skewed population...second, the mean of the sampling distribution of sample means will, in the long run, equal M, the population mean. This second property of the sample mean permits it to be called an unbiased estimate of its parameter, the population mean (Klugh, 1974, p. 204).

Test Instruments

Test instruments used in this study were: (1) Stanford Achievement Test, Intermediate Level II, Forms Y, A, and B; (2) Gates-MacGinite Reading Test: Survey E; (3) .

Differential Aptitude Tests, Forms L and S; and (4) the National Educational Development Tests, Level 2.

The four tests are reviewed in the Sixth Mental Measurement Yearbook (1965) and the Seventh Mental Measurement Yearbook (1972), edited by Oscar Krisen Buros. The tests have been widely used and accepted by educators.

The Stanford Achievement Test, Intermediate Level II, is an achievement test given to students in grades 5.5-6.9. The developers consider it to be a power test.

Three forms of the Stanford were used with groups in this study. Groups 1, 2, 6, and 3 used Form Y, 1964 edition. Group 4 used Form A, and Group 5 used Form B of the 1973 edition.

The 1964 edition, Form Y, had nine scores: word meaning, paragraph meaning, spelling, language, arithmetic computation, arithmetic concepts, arithmetic application, social studies, and science.

The 1973 edition, Forms A and B, had eleven scores: vocabulary, reading comprehension, word study skills, mathematics concepts, mathematics computations, mathematics application, spelling, language; social science, science, and listening comprehension.

Two new tests were included in the 1973 edition: word study skills and listening comprehension. A table of equivalence of forms for corresponding tests of the successive editions has been prepared. It can be found in the Stanford Research Report #5 (1975).

Information about major research programs, norms, validity, reliability, and other technical information can be found in the Stanford Achievement Test, Manual Part V, Technical Data Report (1975) and in Stanford Research Reports.

The Gates-MacGinite Reading Test: Survey E is a general reading test for grades 7-9. It was first published in 1939. The test consists of three parts: speed and accuracy, vocabulary, and comprehension. The developers consider the test a power test.

Information regarding the selection of items, reliability, and other technical data can be found in the Technical Manual, Gates-MacGinite Reading Tests (1965).

The Differential Aptitude Test is an aptitude test for grades 8-12 and adults. The first edition was published in 1947, the fourth edition (used with Group 1 of this study) in 1966, and the fifth edition (used with Groups 2, 3, 4, 5, and 6) in 1973.

The Differential Aptitude Test has nine scores: verbal reasoning, numerical ability, total verbal reasoning and numerical ability, abstract reasoning, clerical speed and accuracy, mechanical reasoning, space relations, spelling, and grammar.

Forms L (1966) and S (1973) were the forms from each of the editions used in this study. Form S of the 1973 edition was largely based on Form L of the 1966 edition. Tables of interform correlations between Forms S and M are given in

the Fifth Edition Manual for Differential Aptitude Tests Forms S and T (1973), pp. 77-78.

Information about norms and profiles, equivalence of forms, validity, reliability, correlation with other tests, and other technical information can be found in the Fifth Edition Manual for Differential Aptitude Tests Forms S and T (1973).

The National Educational Development Test, Level 2, is an achievement battery given to students in grades 9-10. It was first developed in 1959. It is a secured test with two new forms issued annually. The battery has six scores: English usage, mathematics usage, social studies reading, natural sciences reading, word usage, and total.

Information about the technical data can be found in the Interpretative Manual for Grades 9-10 and the Technical Report for each year of the test.

The tests described are used in the Traverse City Public Schools and thus have been used in this study. Following is a chart with the tests and the months when they are administered:

<u>Grade</u>	<u>Month</u>	<u>Test</u>	<u>Type of Test</u>
6	March	Stanford Achievement Test, Intermediate II	Achievement
7	Sept.	Gates-MacGinite Reading Test	Reading
8	Oct.	Differential Aptitude Test	Aptitude
9	Sept.	Gates-MacGinite Reading Test	Reading
10	Oct.	National Educational Differential Test	Achievement

Data Needed

Data needed for this study were three-fold:

1. Data to describe the CAP program.
2. Data to describe other factors which may have impacted on test results and program effect:
 - a. Demographics.
 - b. Other educational experiences.
 - c. Impact of history experienced by students in the same geographical area.
3. Standardized test results.

Data Collection

An April, 1979, interview was held with the Acting Director of Personnel and Director of Curriculum and Instruction of the Traverse City Public Schools to secure permission to do the study and to use test data collected by the school system. Permission was granted. A letter dated April 22, 1979, was provided to the district stating that random sampling would be used in the study and that students would not be identified.

Description of Communication Arts Pilot (CAP)

Descriptive data of the CAP program was acquired through interviews with three primary informants: the junior high school communication arts chairperson (an original teacher participant in CAP), the seventh and eighth grade reading teachers who had been involved with CAP from

the beginning, and through documents provided by these persons.

CAP is herein after known as "X." "X" statistically has no meaning except as format for the whole program.

Other Data

Demographics. Demographic data were acquired from two sources:

1. Population Characteristics: Ten Counties, developed by Nancy C. Hayward, Director, Data Research Center, Traverse City, MI (June, 1979)
2. Traverse City Public Schools' fourth Friday counts for years covered by the study (1970-78)

Other Educational Experiences. Information about curriculum and materials used in other subject areas at the Traverse City Junior High School was requested to determine changes in educational experiences of the groups studied.

Interviews were held with the department chairpeople of the science, social science, and mathematics departments. They described the courses at the junior high school, which courses were required and which elective, and materials used since 1970. The department chairs said there were no curricular changes in their departments since 1969-71.

Information regarding materials was also acquired from the curriculum office of the Traverse City Public Schools.

Impact of History Experienced by Students in the Same Geographic Area. Michigan Educational Assessment Program results, years 9 and 10, were used as an indicator of

whether the effect of general life experiences had shown up on test results that all students in the area had taken. It was assumed that position of achievement on the tests would remain similar if there was no outside influence acting on the students. Results of the Reading Test administered to seventh grade students (statewide) were requested for seven school districts: Traverse City, Charlevoix, Manton, Elk Rapids, Sutton's Bay, Petoskey, and Mesick. Traverse City and Charlevoix provided their reports; the rest were requested from the Michigan Department of Education.

The method of reporting the data has varied from year to year. The format has been consistent since year nine of the program. The data for the four years previous to year nine were included in the year nine report. Only data for those years appearing on the year nine and year ten reports were used in this study. To have sought information prior to 1974, which appeared on the year nine report, would have entailed effort beyond what the information would provide.

Tests

The tests used in the study were all administered by the staff of the Traverse City Public Schools. Scoring was done by the test publishers, except the Gates-MacGinite Reading Test, which was scored by the Traverse Bay Area Intermediate District. Computer printouts of the results were

provided to the Traverse City Public Schools by the scoring agents.

The test data collected for the study were obtained, when available, from the computer printouts.

Stanford. The computer printouts for the Stanford Achievement Test were located at the central office of the Traverse City Public Schools. The samples for the study were selected using the lists of the printouts. A systematic sampling procedure was used and is described in the "Sample" section of this chapter.

Selection of Sample 3 and collection of the Stanford data for Sample 3 were begun on May 4, 1979. 150 subjects were drawn for the sample. Stanford Form Y had been used. It was administered in the spring of 1973 to the sixth grade students of the district. Selection of Sample 3 and the collection of Stanford data for the sample were completed on May 7, 1979.

Selection of Sample 2 and collection of the Stanford data for this sample were begun on May 7, 1979. 150 subjects were drawn for the sample. Stanford Form Y had been used. It was administered in the spring of 1972 to the sixth grade students of the district. Selection of Sample 2 and the collection of Stanford data for the sample were completed on May 7, 1979.

Selection of Sample 4 and collection of the Stanford data for it were begun May 7, 1979. Again, 150 subjects

were drawn for the sample. The Stanford Form A had been used. It was administered in the spring of 1974 to the sixth grade students of the district. Selection of Sample 4 and the collection of Stanford data for it were completed on May 8, 1979.

Selection of Sample 5 and the collection of the Stanford data for Sample 5 were begun on May 8, 1979. 150 subjects were drawn for the sample. Stanford Form B had been used. It was administered during the spring of 1975 to the sixth grade students of the district. Selection of Sample 5 and the collection of Stanford data for it were completed on May 8, 1979.

Selection of Sample 1 and the collection of the Stanford data for Sample 1 were begun on May 8, 1979. Stanford Form Y had been used. It was administered Spring, 1971, to the sixth grade students of the district. 150 subjects were drawn for the sample. Selection of Sample 1 and the collection of Stanford data for it were completed on May 16, 1979.

Grade equivalents and percentiles were provided on the computer printouts and recorded for the study. The data provided in Form Y, used with Samples 1, 2 (which includes Groups 2 and 6 of the study), and 3, were for: word meaning, paragraph meaning, spelling, language, arithmetic computation, arithmetic concepts, arithmetic application, social studies, science, and total test.

The data provided in Forms A and B, used with Samples 4 and 5, respectively, were for: vocabulary, reading comprehension, word study skills, mathematics concepts, mathematics computations, mathematics application, spelling, language, social science, science, listening comprehension, total auditory, total reading, total math, and total test.

Manuals for the Stanford Achievement Test, Intermediate Level II were provided by the curriculum office of the Traverse City Public Schools.

National Educational Development Test. The National Educational Development Test (NEDT) and Differential Aptitude Test (DAT) data were made available at the Traverse City Senior High in computer printouts, student CA 60's, or student permanent records.

Collection of the NEDT data for Sample 3 was begun on May 16, 1979. The test was administered in the fall of 1976 to the tenth grade students of the district. Only the scores for the subjects in the sample were sought. If a subject did not have scores for this test, s/he was dropped from the sample. The collection of NEDT data for Sample 3 was completed on May 16, 1979.

Collection of the NEDT data for Sample 4 was begun on May 17, 1979. The test was administered during the fall of 1977 to the tenth grade students of the district. Only the scores for the subjects in the sample were sought. If

a subject did not have scores for this test, s/he was dropped from the sample. The collection of NEDT data for Sample 4 was completed on May 17, 1979.

Collection of the NEDT data for Sample 5 was begun on May 17, 1979. The test was administered in Fall, 1978, to the tenth grade students of the district. Only the scores for the subjects in the sample were sought; consequently, if a subject did not have scores for this test, s/he was dropped from the sample. The collection of NEDT data for Sample 5 was completed on May 24, 1979.

The collection of NEDT data for Sample 1 began on May 24, 1979, the test having been administered in the fall of 1974 to tenth grade students in the district. Only the scores of subjects in the sample were sought. If a subject did not have scores for this test, s/he was dropped from the sample. The collection of NEDT data for Sample 1 was completed on May 24, 1979.

Collection of NEDT data for Sample 2 was begun on June 4, 1979. The test was administered during the fall of 1975 to tenth grades students of the district. Since only the scores for the subjects in the sample were sought, if a subject did not have scores for this test, s/he was dropped from the sample. The collection of NEDT data for Sample 2 was completed on June 12, 1979.

National and local percentiles and standard scores were provided on the NEDT computer printouts and recorded for the

study. The data provided were for: composite, English usage, math usage, social studies reading, natural science reading, and word usage.

The Interpretative Manual for Grades 9-10, 1970-71 and 1978-79 were provided by the counseling and guidance department of the school district.

Differential Aptitude Test. Collection of the Differential Aptitude Test (DAT) data for Sample 3 was begun on May 5, 1979. The test was Form S and was administered in the fall of 1974 to the eighth grade students of the district. Only the scores for the subjects in the sample were sought. If a subject did not have scores for this test, s/he was dropped from the sample. The collection of the DAT data for Sample 3 was completed on June 16, 1979.

Collection of the DAT data for Sample 4 was begun on May 17, 1979. The test was Form S and was administered in the fall of 1975 to the eighth grade students of the district. Since only the scores for the subjects in the sample were sought, if a subject did not have scores for this test, s/he was dropped from the sample. The collection of DAT data for Sample 4 was completed on May 17, 1979.

Collection of the DAT data for Sample 5 began on May 25, 1979. The test was Form S and was administered Fall, 1976, to the eighth grade students of the district. Only the scores for the subjects in the sample were sought. If

a subject did not have scores for this test, s/he was dropped from the sample. The collection of the DAT data for Sample 5 was completed on May 25, 1979.

The collection of the DAT data for Sample 2 was begun on May 31, 1979. The test was Form S and was administered during the fall of 1973 to the eighth grade students of the district. Only the scores for the subjects in the sample were sought; consequently, if a subject did not have scores for this test, s/he was dropped from the sample. The collection of the DAT data for Sample 2 was completed on June 12, 1979.

Collection of the DAT data for Sample 1 was begun on June 4, 1979. The test was Form L and was administered in the fall of 1972 to the eighth grade students of the district. Only the scores for the subjects in the sample were sought. If a subject did not have scores for this test, s/he was dropped from the sample. Collection of the DAT data for Sample 1 was completed on June 12, 1979.

Raw scores and percentiles were provided on the computer printouts and/or student record stickers on the permanent records and were recorded for the study. The data provided in Form L, used with Sample 1, were for: verbal reasoning, numerical ability, verbal reasoning plus numerical ability, abstract reasoning, clerical speed and accuracy, mechanical reasoning, space relations, spelling, and grammar.

The data provided in Form S, used with Samples 2, 3, 4, and 5 were for the same tests as named for Form L above.

The Fifth Edition Manual for the Differential Aptitude Tests Forms S and T (1974) was provided by the department of guidance and counseling of the school district.

Gates-MacGinite Reading Test. Computer printouts of Gates-MacGinite Reading Test: Survey E, administered to ninth grade students in the district, were available for Samples 2, 3, 4, and 5. One was not available for Sample 1.

No computer printouts of the Gates administered to seventh grade students in the district were available. Administrative sources felt the printouts might have been destroyed when the students left the junior high school.

Grade equivalent scores of the Gates administered in the seventh grade were available in permanent records for those students who had participated in CAP: Group 2 of Sample 2, Sample 3, Sample 4, and Sample 5. No information was available for the Gates seventh grade test for Sample 1 or Group 6 of Sample 2.

Collection of the Gates data, administered in both the seventh and ninth grades, was begun on May 5, 1979, and was completed on June 16, 1979.

Standard scores, percentiles, and grade equivalents were provided on the computer printouts for the Gates administered in the ninth grade and were recorded for the study. The data provided was for: vocabulary, comprehension, and composite.

Grade equivalents were available for the Gates administered in the seventh grade and were recorded for the study. The data provided was for vocabulary and comprehension.

Design and Analysis

The large design/framework of the study is similar to Design 15, the Recurrent Institutional Cycle Design: a "Patched-Up" Design (Campbell, 1963):

The design is appropriate to those situations in which a given aspect of an institutional process is, on some cyclical schedule, continually being presented to a new group of respondents. such situations include schools, indoctrination procedures, apprenticeships, etc. If in these situations one is interested in evaluating the effects of such a global and complex X as an indoctrination program, then the Recurrent Institutional Cycle Design probably offers as near an answer as is available from the designs developed thus far (p. 60).

This design combines the "longitudinal" and "cross-sectional" approaches commonly employed in developmental research...The cross-sectional study itself confounds maturation with selection and mortality. The longitudinal study confounds maturation with repeated testing and with history...The combination, perhaps with repeated cross-sectional comparisons at various times, seems ideal (p. 61).

...feature often characteristic of such designs is that the effect of X is demonstrated in several different manners. This is obviously an important feature where each specific comparison would be equivalent by itself (p. 57).

The model, as it would appear in most schools:

Class A	X	O ₁	-----	
Class B ₁		R	O ₂	X O ₃
Class B ₂		R		X O ₄

Class C			O ₅	X

(p. 60)

Table 1 shows how the model is applied to this study.

"X" represents the CAP program.

This study expands Design 15 in two important ways. As shown in Table 1, Group 2 and Group 6 were randomly assigned to treatment or no treatment. The study was approached as an experimental, quasi-experimental study. Designs which are used in both categories were used in this study.

Random assignment is a necessary factor in true experiments:

...If you are told the selection was more or less random, keep questioning to find out what the "less random" part was...you could treat your data as though it came from a true control group design (Fitz-Gibbon, 1978, p. 49).

...randomization is conceived to be a process occurring at a specific time and is the all-purpose procedure for achieving pretreatment equality of groups, within known statistical limits (Campbell, 1963, p. 6).

Table 1. Groups (CAP-nonCAP):
Diagram of Treatment and Observations
(Tests) of the Study

	Group	1970	1971	1972	1973	1974	1975	1976	1977	1978
Class A	1	O ₁		O ₃		O ₅				
Class B ₁	6		O ₁ R		O ₃	O ₄	O ₅			
Class B ₂	2		O ₁ R	O ₂ X	O ₃	O ₄	O ₅			
Class C	3			O ₁	O ₂ X	O ₃ X	O ₄	O ₅		
Class D	4				O ₁	O ₂ X	O ₃ X	O ₄	O ₅	
Class E	5					O ₁	O ₂ X	O ₃ X	O ₄	O ₅

O ₁	= Stanford Achievement Intermediate II
O ₂	= Gates-MacGinitie Reading Test (administered in seventh grade)
O ₃	= Differential Aptitude Test
O ₄	= Gates-MacGinitie Reading Test (administered in ninth grade)
O ₅	= National Educational Development Test
X	= Communication Arts Pilot (CAP)
R	= Random Assignment

Students were assigned to CAP or non-CAP classes in a random fashion. They were assigned by computer. The problem of elective classes was not present for two reasons: there are very few electives in the seventh grade, and the CAP classes were taught all hours of the day, except the teachers' conference hour. The combination of these factors makes it very unlikely a student would not have been able to be assigned to a CAP class. The researcher is satisfied the assignment was random for Groups 2 and 6 of the study.

Where This Study Expands Design 15

The other important factor is X, as referred to in this study, CAP. X was not introduced by the researcher. Ordinarily, this would preclude experimental or quasi-experimental designs; but, as Campbell states:

Designs 7, 10, 12, 13 (but not 13a) and 14 would be applicable both for naturally occurring X's and for X's deliberately introduced by the experimenter (p. 64).

Campbell and Stanley further state:

Where we have pretests and where clear-cut determination of who were exposed and who were not is available, then Designs 10 and 14 may be convincing even without the randomization. But for a design lacking a pretest (imitating Design 6) to occur naturally requires very special circumstances, which almost never happen. Even so, in keeping with our general emphasis upon the opportunistic exploitation of those settings which happen to provide interpretable data, one should keep his eyes open for them. Such settings will be those in which it seems plausible that exposure to X was lawless, arbitrary, uncorrelated with prior conditions (Campbell, 1963, p. 65).

The term longitudinal was used with this study based on Good's definition of the term:

...a study that follows a case or group of cases over a period of time; includes genetic studies, follow-up studies, growth studies, and experimental growth studies; its purpose may be to gather normative data on growth, to plot trends (as of attitudes) or to observe the effects of special factors (1973, p. 565).

Hypotheses

The hypotheses tested and the models and statistics used in this study were:

Research Question 1

Is there a difference, as measured by standardized tests, between the reading and associated abilities of students who participated in CAP and students who did not?

Null Hypothesis: no difference will be found, as measured by standardized test scores, between the reading and associated abilities of students who participated in CAP and students who did not.

Symbolically: $H_0: M_1 = M_2$

Legend: M_1 = CAP group mean

M_2 - non-CAP group mean

Model: Campbell and Stanley's Design 4 and 6 will be used for Groups 2 and 6 of the study. Random assignment is assumed.

Design 4. Pretest-posttest control group design (Campbell, 1963, pp. 13-22).

R	O	X	O
---	---	---	---

R	O		O
---	---	--	---

R = Random assignment

O = Observation

X = Treatment

O = Observation (p. 13)

Stanford test results are used as the pretest, and tests subsequent to X are used as posttests.

Statistics: analysis of covariance

...analysis of covariance with pretest scores as the covariant are usually preferable to simple gain-score comparisons (Campbell, 1963, p. 23).

Design 6. Posttest-only control group design (Campbell, 1963, pp. 25-27).

R		X	O
---	--	---	---

R			O	(p. 25)
---	--	--	---	---------

While the pretest is a concept deeply embedded in the thinking of research workers in education and psychology, it is not actually essential to true experimental designs. For psychological reasons, it is difficult to give up "knowing for sure" that the experimental and control groups were "equal" before the differential experimental treatment. Nonetheless, the most adequate all-purpose assurance of lack of initial biases between groups is randomization (p. 25).

Statistics: t-test, analysis of covariance.

Design 6 is perhaps the only setting for which this test is optimal...covariance analysis...can be used, thus providing an increase in the power of the significance test very similar to that provided by a pretest. Identicalness of pretest and posttest is not essential...whether such a pseudo-pretest design should be classified as Design 6 or Design 4 is of little moment. It would have the advantages of Design 6 in avoiding an experimenter-introduced pretest session and in avoiding

the "giveaway" repetition of identical or highly similar unusual content (Campbell, 1963, p. 26).

...if appropriate antecedent variants are available, they should certainly be used for blocking or leveling, or as covariants. This recommendation is made for two reasons: first, the statistical tests available for Design 4 are more powerful than those available for Design 6...second, the availability of pretest scores makes possible examination of the interaction of X and pretest ability level, thus exploring the generalizability of the finding more thoroughly (p. 26).

Design 10 is used to make comparisons between CAP and non-CAP groups other than the Group 2-Group 6 comparison; i.e., CAP and non-CAP combinations:

2,3,4,5 - 1,6

3 - 1

3,4,5 - 1

Design 10: the Nonequivalent Control Group Design (Campbell, 1963, pp. 47-50).

O X O

O O (p. 47)

----- = no pre-experimental sampling equivalence

...the addition of even an unmatched or non-equivalent control group reduces greatly the equivocality of interpretation over what is obtained in Design 2, the One-Group Pretest-Posttest Design. The more similar the experimental and the control groups are in their recruitment, and the more this similarity is confirmed by the scores on the pretest, the more effective this control becomes (pp. 47-48).

Statistics: analysis of covariance

The Designs 4, 6, and 10 of Campbell and Stanley (1963) compare with Designs 1, 2, and 3 of Fitz-Gibbon and Morris (1978) and Designs 3, 4, and 7 presented in Borg and Gall (1971). These comparisons can be seen in Table 2.

Table 2. Comparability of Designs

	<u>Campbell</u>	<u>Morris</u>	<u>Borg</u>	Sources of Invalidity*	
				<u>Inter- nal</u>	<u>Exter- nal</u>
Pretest-posttest control group	Design 4	Design 1	Design 3	none	Inter- action of Test- ing, X
Posttest-only control group	Design 6	Design 2	Design 4	mor- tality	none
Unequivalent control group	Design 10	Design 3	Design 7	Inter- action of se- lec- tion & matur- ation, etc.	Inter- action of se- lec- tion & test- ing & X

*Borg, 1973, p. 376; adapted from Campbell, 1963.

Research Question 2

Is the effect of CAP upon different student groups consistent over time, as measured by standardized tests?

Null Hypothesis: no difference will be found between the effect of CAP on one group of students than on another group of students.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group (Groups 2, 3, 4, 5)

M_2 = Any other CAP group (Groups 2, 3, 4, 5)

Statistics: same as used for Research Question 1

Design 2, the One-Group Pretest-Posttest Design (Campbell, 1963, p. 7) would ordinarily fit this hypothesis, but there was more powerful information available from the use of Designs 6 and 10, previously described. Graphs and charts will be used to present the information for this hypothesis.

Research Question 3

Is there a difference in the effect of CAP for students of varying reading abilities?

Null Hypothesis: no difference will be found in the effect of CAP for students of varying reading abilities.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group (Groups 2, 3, 4, 5)

M_2 = any other CAP group (Groups 2, 3, 4, 5)

Statistics: information acquired from statistics used in Research Question 1.

Scattergrams were developed from the high and low values of Gates-MacGinitie, administered in grades 7 and 9. Graphs were used to indicate trends. Regression lines were used to indicate difference in effect.

The Statistical Package for the Social Sciences (SPSS) (Nie, Hull, Jenkins, Steinbrenner, and Brent, 1975) was used

for data analysis. The analysis was done at the computer center at Michigan State University.

Summary

The purpose of this study was to do a summative evaluation of the Communication Arts Pilot (CAP) program of the Traverse City (Michigan) Junior High School. Specifically, the study was designed to determine the effect of the program on the reading and associated abilities of students, as measured by standardized tests, who were in grades seven to nine in 1972-78. The program was introduced by the communication arts staff of the Traverse City Junior High School.

Samples for the study were selected from the sixth grade classes of 1970, 1971, 1972, 1973, and 1974. Each sample became one group in the study except Sample (1971) which became Groups 2 and 6.

Assignment to CAP or non-CAP for Groups 2 and 6 was random. Groups 1, 3, 4, and 5 were non-CAP (Group 1) or CAP (Groups 3, 4, 5), dependent only on the year they began seventh grade English.

Tests administered at five different grade levels were used in the study: Stanford Achievement Intermediate II (sixth grade), Gates-MacGinite Reading Test (seventh and ninth grades), Differential Aptitude Test (eighth grade), and National Educational Development Test (tenth grade).

CAP was described as a mastery learning type program with two components: reading and writing. The objectives for the reading component are from the CROFT Reading

Table 3. Groups (CAP-non-CAP),
Tests and When Administered

<u>Group</u>	<u>N</u>	<u>Test and Grade</u>				
		Stanford <u>6</u>	Gates <u>7</u>	DAT <u>8</u>	Gates <u>9</u>	NEDT <u>10</u>
1	87	1970		1972		1974
6	52	1971		1973	1974	1975
2	43	1971 X	1972	1973	1974	1975
3	87	1972 X	1973 X	1974	1975	1976
4	87	1973 X	1974 X	1975	1976	1977
5	87	1974 X	1975 X	1976	1977	1978

Stanford = Stanford Achievement Test Intermediate II

Gates = Gates-MacGinite Reading Test

DAT = Differential Aptitude Test

NEDT = National Educational Development Test

X = Communication Arts Pilot (CAP)

Comprehension Skills material. The writing component is teacher-developed.

The students are pre- and posttested on reading and writing skills. Up to three opportunities may be provided for a student to reach mastery on a skill.

The large design framework of the study is similar to Design 15, the Recurrent Institutional Cycle Design: a "Patched-up" Design (Campbell, 1963, pp. 57-61). Two factors allowed for the use of stronger designs when appropriate with their more powerful statistics:

1. Random assignment of Groups 2 and 6, treatment and control.
2. Naturally occurring X.

Campbell and Stanley speak to these factors and their use whenever the opportunity is presented (Experimental and Quasi-Experimental Designs for Research, 1963).

Statistics appropriate to the described designs were applied to the data. The SPSS programs were used.

Chapter IV contains a presentation of the data collected. An analysis of the data is provided.

Chapter V contains the conclusions of the study. Recommendations will be made.

CHAPTER IV

RESEARCH FINDINGS

Introduction

The purpose of this study was to do a summative evaluation of the Communication Arts Pilot (CAP) program at Traverse City (Michigan) Junior High School. Specifically, the purpose was to determine the effect of the program on the reading and associated abilities of students who were in grades seven to nine, 1972-78. The CAP program was introduced by the Communication Arts Department of the Traverse City Junior High School.

The research questions explored in this study were:

1. Is there a difference, as measured by standardized tests, between the reading and associated abilities of students who participated in CAP and students who did not?
2. Is the effect of CAP upon different student groups consistent over time, as measured by standardized tests?
3. Is there a difference in the effect of CAP for students of varying reading abilities?

Samples for the study were selected from the sixth grade classes of 1970, 1971, 1972, 1973, and 1974. Each sample became one group in the study except Sample 2 (1971) which became Groups 2 and 6.

Assignment to CAP or non-CAP for Groups 2 and 6 was random: Groups 1 (1970), 3 (1972), 4 (1973), and 5 (1974) were non-CAP (Group 1) or CAP (Groups 3, 4, 5), dependent only on the year they began seventh grade English.

Tests administered at five different grade levels were used in the study: Stanford Achievement Intermediate II (sixth grade), Gates-MacGinitie Reading Test (seventh and ninth grades), Differential Aptitude Test (eighth grade), and National Educational Development Test (tenth grade).

CAP was described as a mastery learning type program with two components: reading and writing. The objectives for the reading component were those of the CROFT Reading Comprehension Skills material. The writing component was teacher-developed.

The students were pre- and posttested on reading and writing skills. Up to three opportunities were provided for a student to reach mastery on a skill.

Table 3 is reprinted here to provide information about groups, number in each group, tests, when tests were administered to each group, when and if students participated in CAP (indicated by the X). Blanks indicate information was not available for that group.

On the following pages of this chapter, the data are presented and analyzed. The major divisions of the chapter are based on the three research questions.

Table 3. Groups (CAP-non-CAP),
Tests and When Administered

<u>Group</u>	<u>N</u>	<u>Test and Grade</u>				
		Stanford <u>6</u>	Gates <u>7</u>	DAT <u>8</u>	Gates <u>9</u>	NEDT <u>10</u>
1	87	1970		1972		1974
6	52	1971		1973	1974	1975
2	43	1971 X	1972	1973	1974	1975
3	87	1972 X	1973 X	1974	1975	1976
4	87	1973 X	1974 X	1975	1976	1977
5	87	1974 X	1975 X	1976	1977	1978

Stanford = Stanford Achievement Test Intermediate II

Gates = Gates-MacGinite Reading Test

DAT = Differential Aptitude Test

NEDT = National Educational Development Test

X = Communication Arts Pilot (CAP)

Research Question 1: Hypotheses and Analysis

Research Question 1: is there a difference, as measured by standardized tests, between the reading and associated abilities of students who participated in CAP and students who did not?

Groups 2 and 6 are the groups referred to in Hypotheses 1.1 to 1.7 as CAP and non-CAP. Students were randomly assigned to Groups 2 and 6. Group 2 was the experimental group (CAP). Group 6 was the control group (non-CAP). Randomly assigned control group designs were used with these groups, Design 4 and 6 (Campbell, 1963).

Null Hypothesis 1.1

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group mean (Group 2)

M_2 = non-CAP group mean (Group 6)

Significance Level = .05

Alternative Hypothesis 1.1

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = CAP group mean (Group 2)

M_2 = non-Cap group mean (Group 6)

Significance Level = .05

Scores are reported in two forms, percentiles and grade equivalents.

Percentiles

The t test was used to statistically test the difference between the group mean vocabulary scores of CAP and non-CAP groups. The t-value was -.102. The two tailed probability level of .919 was not statistically significant at the .05 level.

	N	Posttest--Gates (9) Vocabulary (PR)				sig. at .05
		Mean	SD	t-value	t-probability	
Group 2	43	45.67	24.89	-.102	.919	no
Group 6	52	46.21	29.16			

Grade Equivalents

The t test was used to statistically test the difference in the group mean vocabulary scores between the CAP and non-CAP groups. The t-value was -.052. The two tailed probability level of .959 was not statistically significant at the .05 level.

	N	Posttest--Gates (9) Vocabulary (GE)				sig. at .05
		Mean	SD	t-value	t-probability	
Group 2	43	8.97	2.33	-.052	.959	no
Group 6	52	9.00	2.87			

Analysis of covariance was used with the Stanford Achievement Intermediate II vocabulary test as the covariant. The F value was .801 with a significance of .373. It was not statistically significant at the .05 level of significance.

The null hypothesis 1.1 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 1.2

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the comprehension section of the Gates-

Symbolically: $H_0 : M_1 = M_2$

Legend: $M_1 = \text{CAP group mean (Group 2)}$

$M_2 = \text{non-CAP group mean (Group 6)}$

Significance level = .05

Alternate Hypothesis 1.2

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the comprehension section of the Gates-MacGinitie Reading Test, administered in the ninth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: $M_1 = \text{CAP group mean (Group 2)}$

$M_2 = \text{non-CAP group mean (Group 6)}$

Significance level = .05

Scores are reported in two forms, percentile and grade equivalent.

Percentiles

The t test was used to statistically test the difference between the group mean comprehension scores of CAP and non-CAP groups. The t-value was -.049. The two tailed probability level of .961 was not statistically significant at the .05 level.

		Posttest--Gates (9) Comprehension (PR)				sig. at .05	
		N	Mean	SD	t-value		t-probability
Group 2	43	57.26	30.42		-.049	.961	no
Group 6	52	57.56	31.83				

Grade Equivalent

The t test was used to statistically test the difference between the group mean comprehension scores of CAP and non-CAP groups. The t-value of .083 with a probability of .934 was not statistically significant at the .05 level.

		Posttest--Gates (9) Comprehension (GE)				sig. at .05
		N	Mean	SD	t-value	t-probability
Group 2	43	10.02	2.75	.083	.934	no
Group 6	52	9.98	3.13			

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F value was .316 with a significance of .575. It was not statistically significant at the .05 level of significance.

The null hypothesis 1.2 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 1.3

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP

students on the spelling section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group mean (Group 2)

M_2 = non-CAP group mean (Group 6)

Significance level = .05

Alternate Hypothesis 1.3

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the spelling section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = mean of CAP group (Group 2)

M_2 = mean of non-CAP group (Group 6)

Significance level = .05

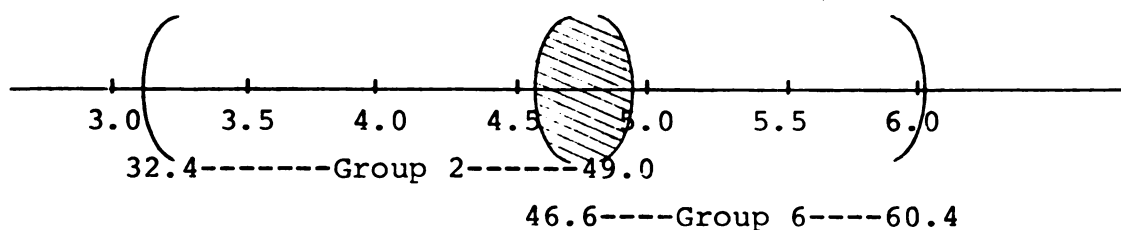
Scores are reported in percentiles.

The t test was used to statistically test the difference between the group mean spelling scores of CAP and non-CAP groups. The t-value was -2.263. The two-tailed probability level of .024 was significant at the .05 level.

	N	Posttest--DAT (8) Spelling (PR)				sig. at .05
		Mean	SD	t-value	t-probability	
Group 2	43	40.67	26.90	-2.263	.024	yes
Group 6	52	53.50	24.79			

Analysis of covariance was used with the Stanford Achievement Intermediate II Spelling Test as the covariant. The independent variable was "group." The F value was 1.879 with a significance of .174. It was not statistically significant at the .05 level of significance.

Figure 1 shows the ninety-five percent confidence intervals of Group 2 and Group 6 for the DAT Spelling Test.



There was a statistically significant difference (.024) on the t test, the difference between group means of CAP and non-CAP groups. There was not a significant difference (.174) for the F value (1.879) using the analysis of covariance. The analysis of covariance test is the stronger statistical test (Campbell, 1963). The null hypothesis 1.3 was not rejected.

Null Hypothesis 1.4

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the Language Usage section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group mean (Group 2)

M_2 = non-CAP group mean (Group 6)

Significance level = .05

Alternate Hypothesis 1.4

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the Language Usage section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = CAP group mean (Group 2)

M_2 = non-CAP group mean (Group 6)

Significance level = .05

Scores are reported in percentiles.

The t test was used to statistically test the difference between the group mean language usage scores of CAP and non-CAP groups. The t-value was .130. The two-tailed probability level of .897 was not significant at the .05 level of significance.

	N	Posttest--DAT (8) Language Usage (PR)				sig. at .05
		Mean	SD	t-value	t-probability	
Group 2	43	59.00	24.75	.130	.897	no
Group 6	52	58.37	23.17			

Analysis of covariance was used with the Stanford Achievement Intermediate II Language Test as the covariant.

The F value was 1.111 with a significance of .295. It was not statistically significant at the .05 level of significance.

The null hypothesis 1.4 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 1.5

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the verbal reasoning section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group mean (Group 2)

M_2 = non-CAP group mean (Group 6)

Significance level = .05

Alternate Hypothesis 1.5

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the verbal reasoning section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = CAP group mean (Group 2)

M_2 = non-CAP group mean (Group 6)

Significance level = .05

Scores are reported in percentiles.

The t test was used to statistically test the difference between the group mean verbal reasoning scores of CAP and non-CAP groups. The t-value was .197. The two-tailed probability level of .844 was not statistically significant at the .05 level.

		Posttest--DAT (8) Verbal Reasoning (PR)				sig. at .05
	N	Mean	SD	t-value	t-probability	
Group 2	43	63.32	26.54	.197	.844	no
Group 6	52	62.27	27.80			

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F value was .570 with a significance of .452. It was not statistically significant at the .05 level of significance.

The null hypothesis 1.5 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 1.6

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the English usage section of the National Educational Development Tests administered in the tenth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group mean (Group 2)

M_2 = non-CAP group mean (Group 6)

Significance level: .05

Alternate Hypothesis 1.6

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the English usage section of the National Educational Development Tests administered in the tenth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = CAP group mean (Group 2)

M_2 = non-CAP group mean (Group 6)

Significance level = .05

Scores are reported in percentiles.

The t test was used to statistically test the difference between the group mean verbal reasoning scores of CAP and non-CAP groups. The t-value was -.201. The two-tailed probability level of .841 was not statistically significant at the .05 level of significance.

		Posttest--NEDT (10) English Usage (PR)				sig. at .05
N		Mean	SD	t-value	t-probability	
Group 2	43	43.35	27.13	-.201	.841	no
Group 6	52	44.56	32.24			

Analysis of covariance was used with the Stanford Achievement Intermediate II Language Test as the covariant. The F ratio was .284 with a significance of .595. It was not statistically significant at the .05 level.

The null hypothesis 1.6 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 1.7

No difference will be found between the group mean scores of CAP students and the group mean scores of non-CAP students on the word usage section of the National Educational Development Tests administered in the tenth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group mean (Group 2)

M_2 = non-CAP group mean (Group 6)

Significance level = .05

Alternate Hypothesis 1.7

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the word usage section of the National Educational Development Tests administered in the tenth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = CAP group mean (Group 2)

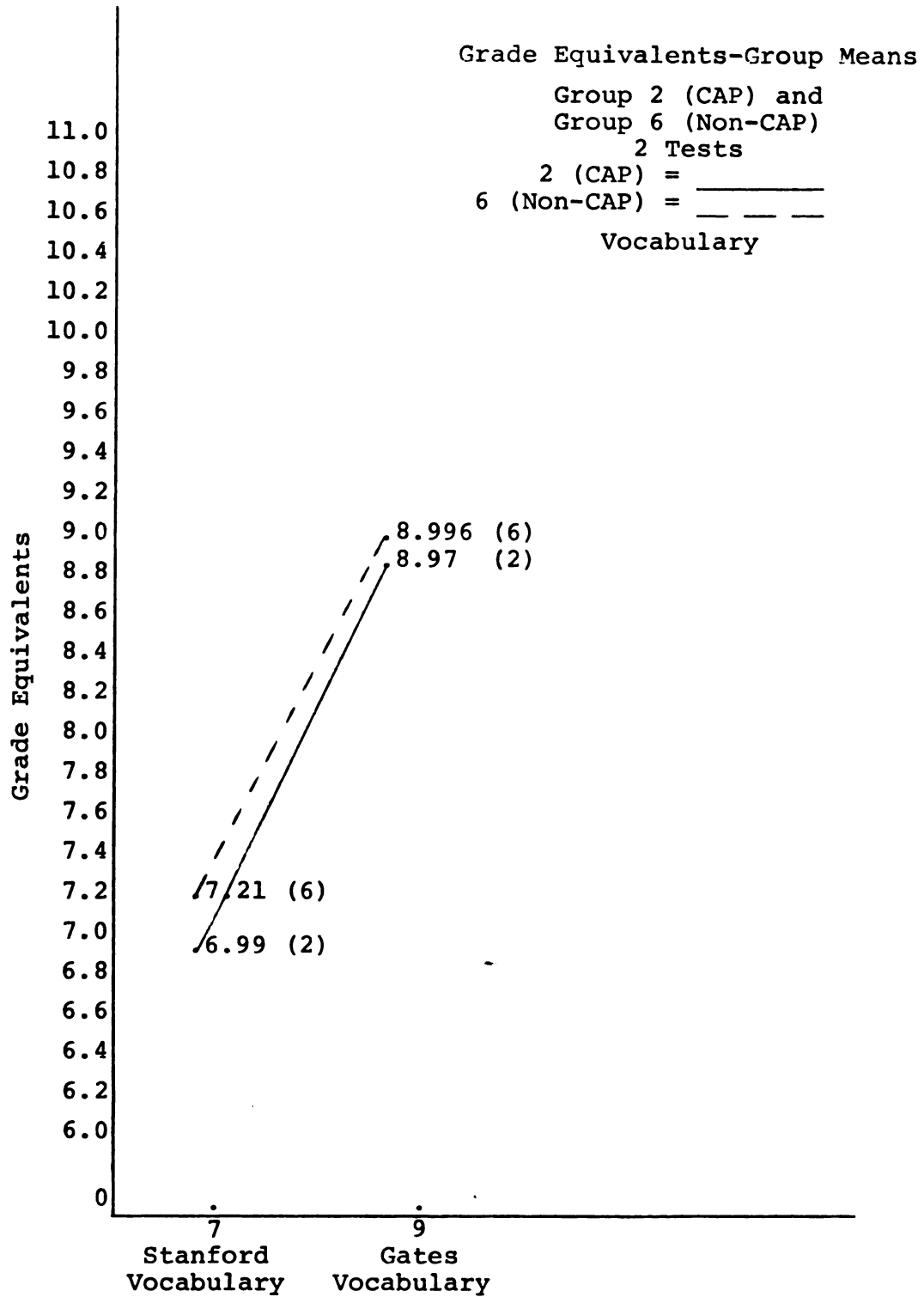
M_2 = non-CAP group mean (Group 6)

Significance level - .05

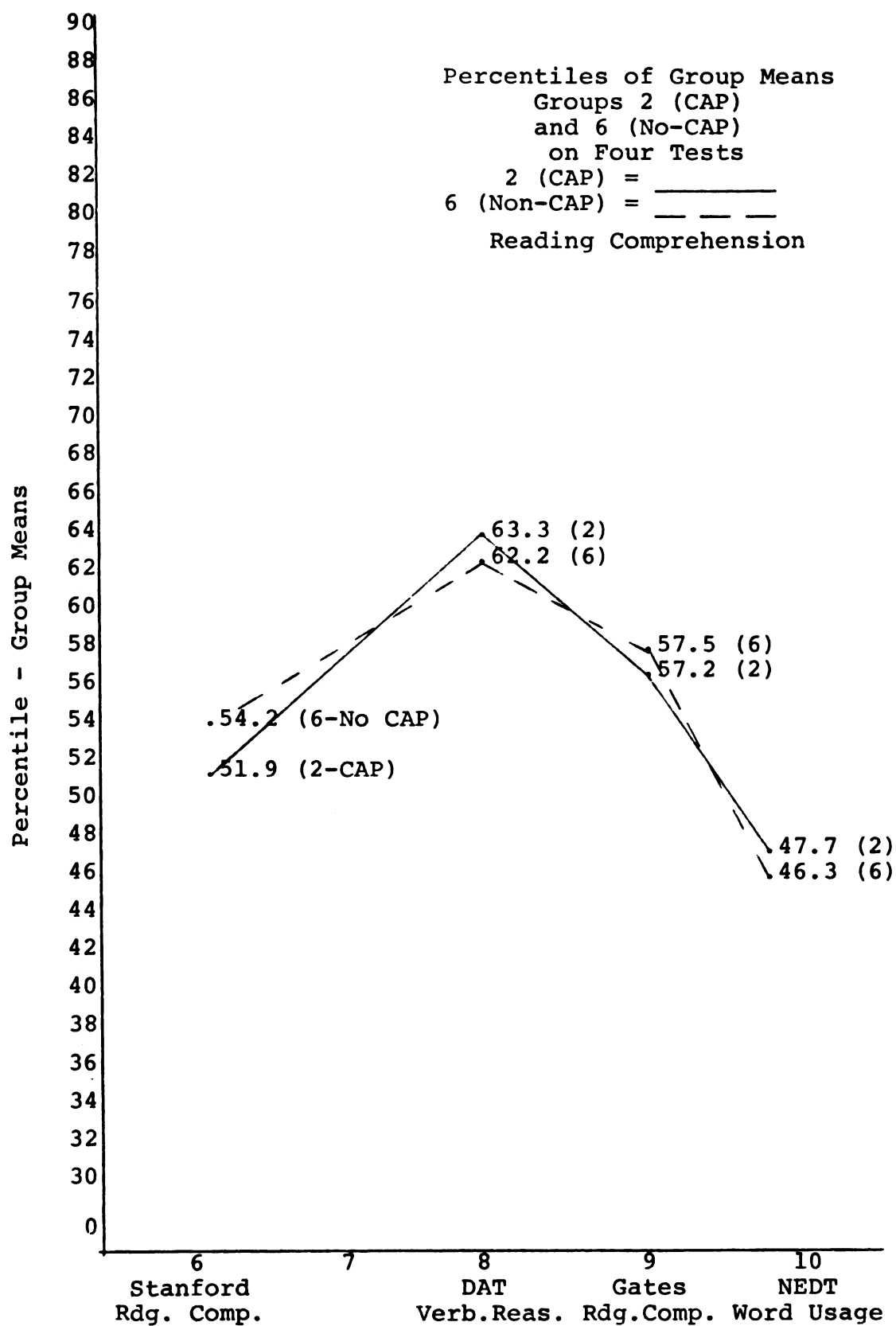
Scores are reported in percentiles.

The t test was used to statistically test the difference between the group mean word usage scores of CAP and non-CAP groups. The t-value was .249. The two-tailed probability level of .803 was not statistically significant at the .05 level of significance.

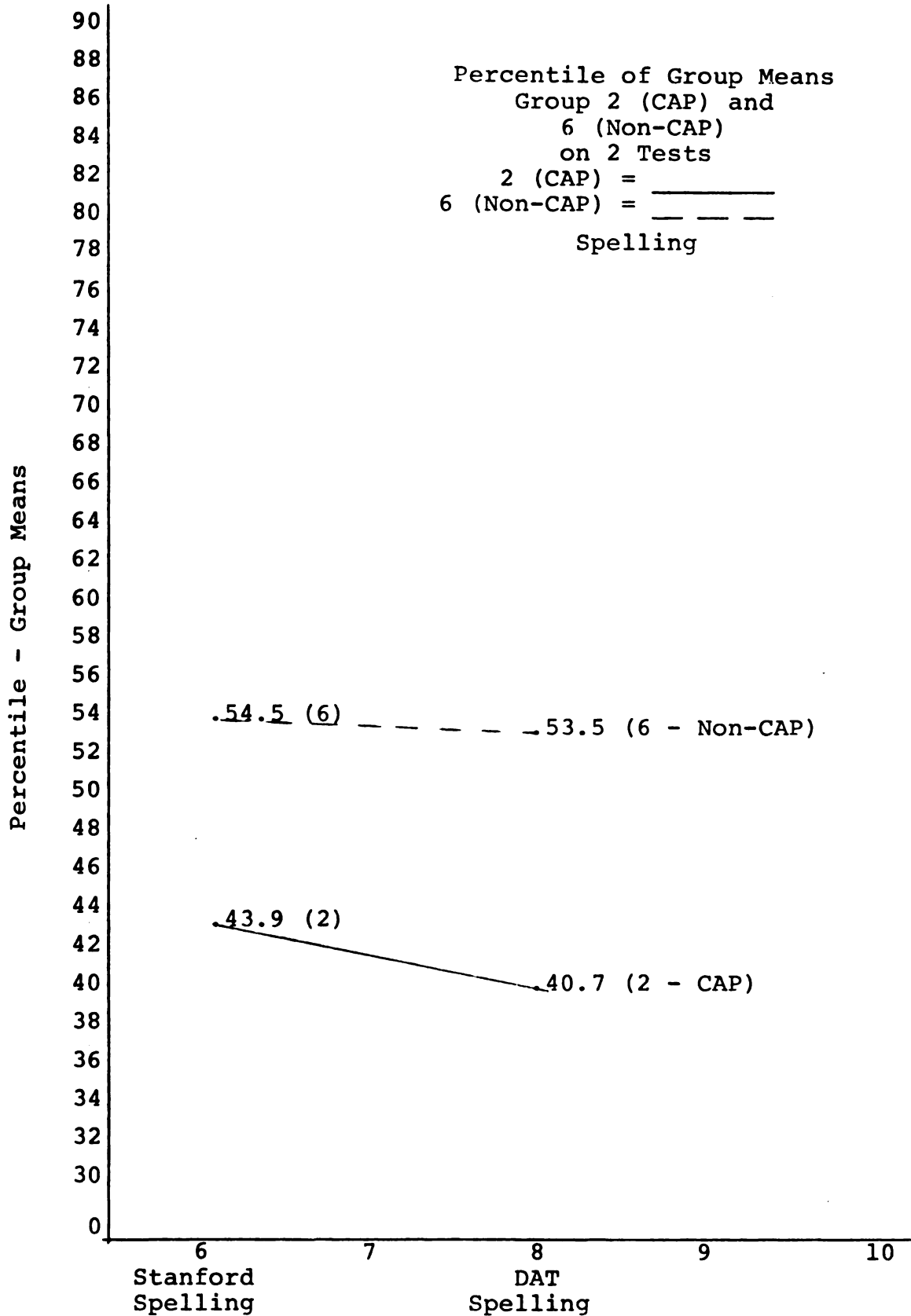
GRAPH 4:1.1



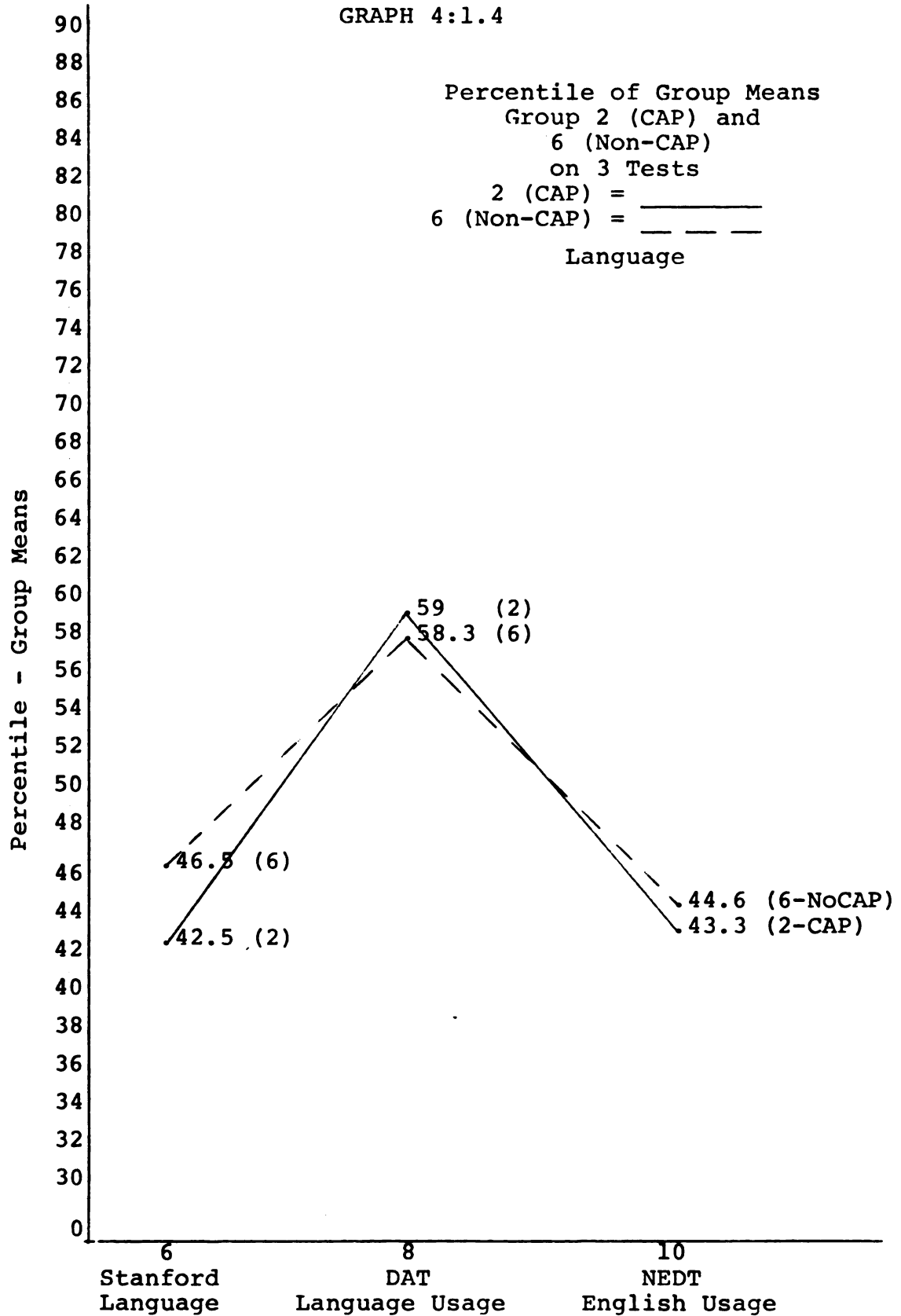
GRAPH 4:1.2



GRAPH 4:1.3



GRAPH 4:1.4



	N	Posttest--NEDT (10) Word Usage (PR)				sig. at .05
		Mean	SD	t-value	t-probability	
Group 2	43	47.79	27.60	.249	.803	no
Group 6	52	46.37	28.34			

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F ratio was .934 with a significance of .336. It was not statistically significant at the .05 level.

The null hypothesis 1.7 cannot be rejected. The .05 level of significance was not reached.

Groups 2, 3, 4, and 5 and Groups 1 and 6 are the groups referred to in Hypotheses 1.8 to 1.14 as CAP and non-CAP. Students were not randomly assigned to the groups except students in Groups 2 and 6. Information was not available for Group 1 in all instances (see Table 3). Where it was not available, the analysis was conducted using only Group 6 as the non-CAP group. An indication is made when that occurs. A non-equivalent control group design was used with these groups, Design 10 (Campbell, 1963).

Null Hypothesis 1.8

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP

students in the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: $M_1 = \text{CAP group mean (Groups 2,3,4,5)}$

$M_2 = \text{non-CAP group mean (Group 6)}$

Significance level: .05

Alternate Hypothesis 1.8

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: $M_1 = \text{CAP group mean (Groups 2,3,4,5)}$

$M_2 = \text{non-CAP group mean (Group 6)}$

Significance level: .05

Scores are reported in percentiles and grade equivalents.

Percentiles

The t test was used to statistically test the differences between the group mean vocabulary scores of CAP and non-CAP groups. The t value was -10.240. The two-tailed probability level of .0000 was statistically significant at the .05 level of significance.

	Posttest--Gates Vocabulary (PR)					sig. at .05
	N	MEAN	SD	t-value	t-probability	
Groups 2,3,4,5	304	47.28	24.89	-10.240	.0000	yes
Group 6	52	46.21	29.16			

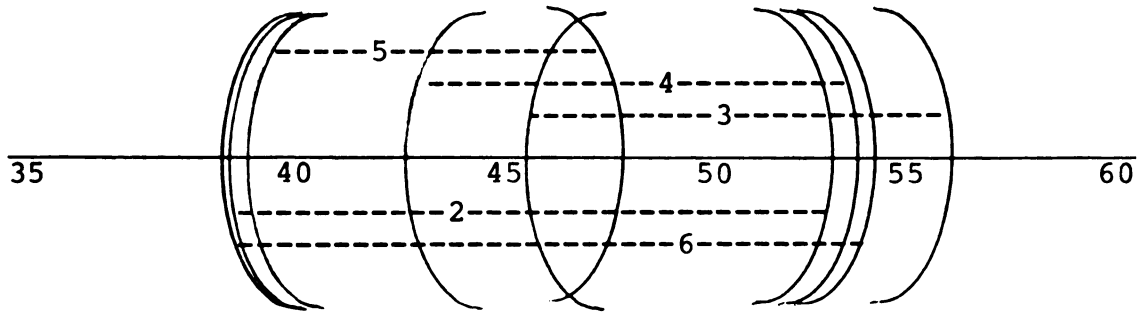
Analysis of variance was used with the F ratio as the test statistic to determine the difference among the sample means. The ratio is determined by dividing the within group sum of squares and the between group sum of squares by their appropriate degrees of freedom.

If there really is a difference among the sample means...the value of F will be substantially larger than 1.0...If there is no significant difference among the sample means, the two estimates (the between and within group sum of squares) will be approximately equal and the value of F will be close to 1.0 (Herzow and Hooper, 1976, p. 393).

The F value was .0775 with a significance level of .7808. It was not statistically significant at the .05 level of significance.

Analysis of covariance was used with the Stanford Achievement Intermediate II vocabulary test as the covariant. The F value was .160 with a significance level of .689. It was not statistically significant at the .05 level of significance.

The ninety-five percent confidence intervals for the means of Groups 2, 3, 4, 5, and 6 (those groups who had been administered the vocabulary section of the Gates-MacGinite Reading Test in the ninth grade) are presented in Figure 2.



	<u>N</u>	<u>Mean</u>	<u>95% Confidence Interval</u>
*Group 2	43	45.67	38.01 to 53.34
*Group 3	87	50.67	45.40 to 55.94
*Group 4	87	48.62	43.14 to 54.10
*Group 5	87	43.34	38.22 to 48.47
Group 6	52	46.21	38.09 to 54.33

* = CAP

There was overlap of the confidence intervals of each group with every other group. The group means of Groups 3, 4, and 5 lie outside of the overlap. If they were accepted as the true mean of the population, the null hypothesis would have to be rejected. The confidence interval provides a range within which the "true" population mean might be with ninety-five percent confidence.

Grade Equivalents

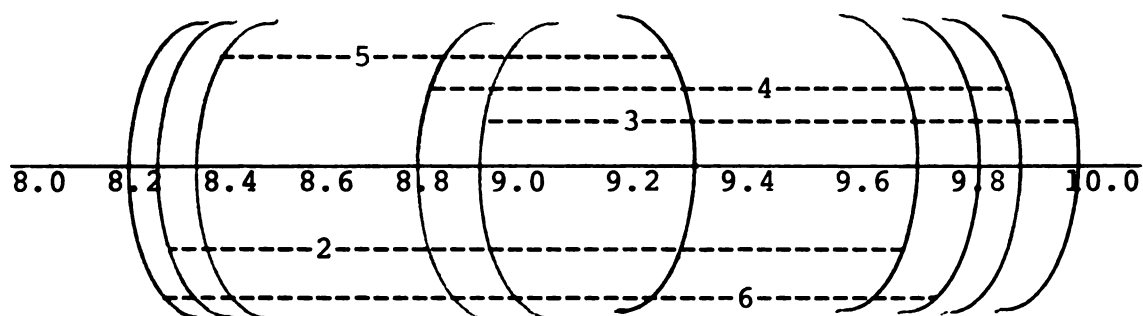
The t test was used to statistically test for the difference between the group mean vocabulary scores of CAP and non-CAP groups. The t value was -20.488. The two-tailed probability level of .000 was statistically significant at the .05 level of significance.

	Posttest--Gates (9) Vocabulary (GE)					sig. at .05
	N	Mean	SD	t-value	t-probability	
Groups 2,3,4,5	304	9.17	2.40	-20.488	.0000	yes
Group 6	52	9.00	2.87			

Analysis of variance was used. The F value was .2081 with a significance level of .6485. It was not statistically significant at the .05 level of significance.

Analysis of covariance was used with the Stanford Achievement Intermediate II vocabulary test as the covariant. The F value was .488 with a significance level of .485. It was not statistically significant at the .05 level of significance.

The ninety-five percent confidence intervals for the means of Groups 2, 3, 4, 5, and 6 for the Gates vocabulary administered in the ninth grade are presented in Figure 3.



	<u>N</u>	<u>Mean</u>	<u>95% Confidence Interval</u>
*Group 2	43	8.97	8.25 to 9.69
*Group 3	87	9.46	8.96 to 9.97
*Group 4	87	9.33	8.80 to 9.87
*Group 5	87	8.80	8.30 to 9.29
Group 6	52	9.00	8.20 to 9.80

* = CAP

There was overlap of the confidence intervals of each group with every other group. The groups means of Groups 3, 4, and 5 lie outside of the overlap. If they were accepted as the true mean of the population, the null hypothesis would have to be rejected. The confidence interval provides a range within which the "true" population mean might lie with ninety-five percent confidence.

The t tests for the group mean scores recorded by both percentiles and grade equivalents produced t-values with probabilities which were statistically significant at the .05 level of significance.

	<u>t-value</u>	<u>t-probability</u>	<u>sig. at .05</u>
Percentile	-10.240	.0000	yes
Grade Equivalent	-20.488	.0000	yes

The analysis of variance for the data in percentiles and grade equivalents produced F values with significance levels which were not statistically significant at the .05 level of significance.

	<u>F value</u>	<u>sig. of F</u>	<u>sig. at .05</u>
Percentile	.0775	.7808	no
Grade Equivalent	.488	.485	no

The analysis of covariance for the data in percentiles and grade equivalents produced F values with significance levels which were not statistically significant at the .05 level of significance.

	<u>F value</u>	<u>sig. of F</u>	<u>sig. at .05</u>
Percentile	.160	.689	no
Grade Equivalent	.488	.485	no

Figure 3 provided information about the ninety-five percent confidence interval for the means of Groups 2, 3, 4, 5, and 6. There was overlap of the confidence intervals of each group with every other group.

The analysis of variance, analysis of covariance, and the confidence interval ranges provide stronger statistical evidence that the null hypothesis 1.8 cannot be rejected than does the t test statistic for rejection. The null hypothesis 1.8 cannot be rejected.

Null Hypothesis 1.9

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the reading comprehension section of the Gates-MacGinitie Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group mean (Groups 2,3,4,5)

M_2 = non-CAP group mean (Group 6)

Significance level: .05

Alternate Hypothesis 1.9

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP

students on the reading comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = CAP group mean (Groups 2,3,4,5)

M_2 = non-CAP group mean (Group 6)

Significance level: .05

Scores were reported in percentiles and grade equivalents.

Percentiles

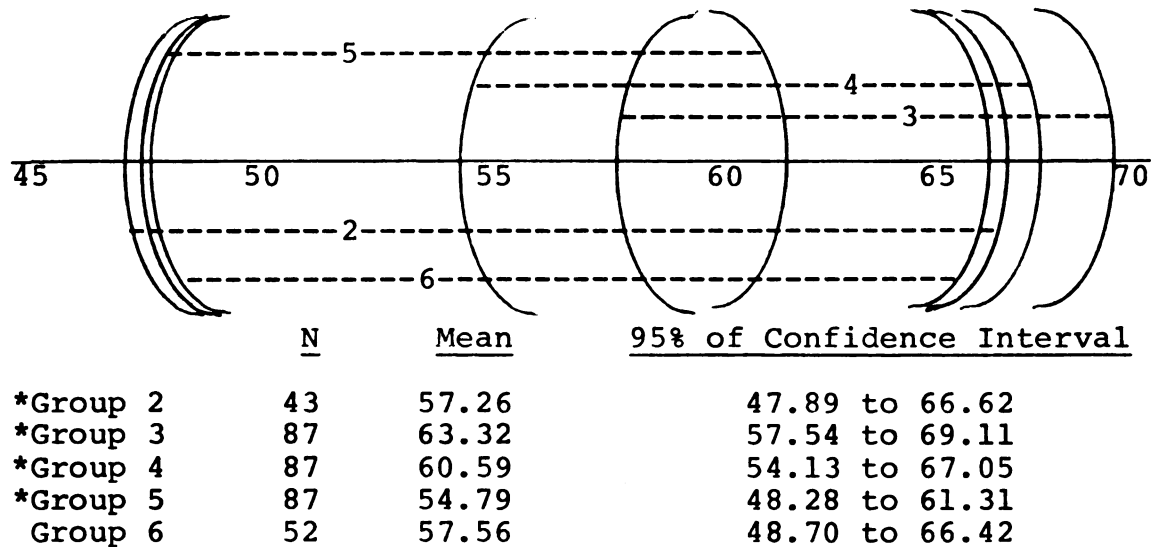
The t test was used to statistically test the difference between the group mean comprehension scores of CAP and non-CAP groups. The t-value was -11.44. The two-tailed probability level of .000 was statistically significant at the .05 level of significance.

	N	Posttest--Gates (9) Comprehension (PR)				sig. at .05
		Mean	SD	t-value	t-probability	
Groups 2,3,4,5	304	59.24	29.58	-11.022	.000	yes
Group 6	52	57.56	31.83			

Analysis of variance was used with the F ratio as the test statistic to determine the difference between the sample means. The F value was .1405 with a significance of .7081. It was not statistically significant at the .05 level of significance.

Analysis of covariance was used with the Stanford Achievement Intermediate II comprehension test as the co-variant. The F value was 1.197 with a significance level of .275. It was not statistically significant at the .05 level of significance.

The ninety-five percent confidence intervals for the means of Group 2, 3, 4, 5, and 6 for the Gates comprehension test administered in the ninth grade are presented below.



* = CAP

There was overlap of the confidence intervals of each group with every other group. The group mean of Groups 2, 3, and 5 lie outside of the overlap. The confidence interval provides a range within which the "true" population mean might lie with ninety-five percent confidence.

Grade Equivalents

The t test was used to statistically test the difference between the group mean comprehension scores of CAP

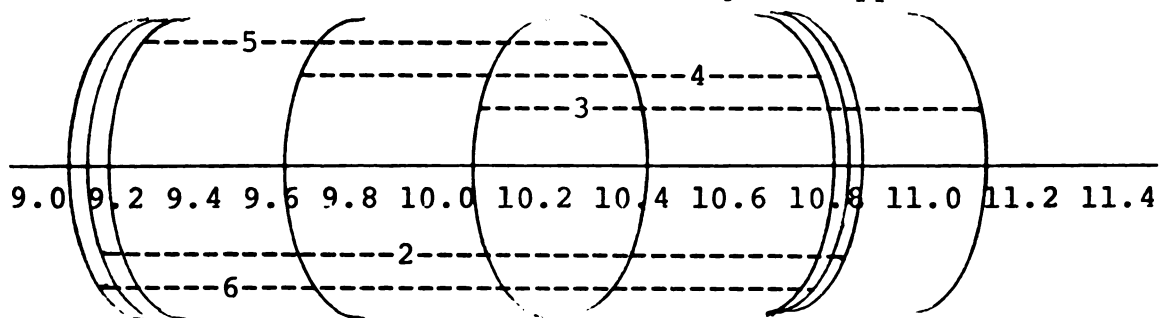
and the non-CAP groups. The t-value was -20.925. The two-tailed probability level of .000 was statistically significant at the .05 level of significance.

	Posttest--Gates (9) Comprehension (GE)					sig. at .05
	N	Mean	SD	t-value	t-probability	
Groups 2,3,4,5	304	10.19	2.76	-20.100	.000	yes
Group 6	52	9.98	3.13			

Analysis of variance was used with the F ratio as the test statistic to determine the difference among the sample means. The F-value was .2583 with a significance of .6116. It was not statistically significant at the .05 level of significance.

Analysis of covariance was used with the Stanford Achievement Intermediate II comprehension test as the covariant. The F value was 1.434 with a significance of .232. It was not statistically significant at the .05 level of significance.

The ninety-five percent confidence intervals for the means of Groups 2, 3, 4, 5, and 6 for the Gates comprehension test administered in the ninth grade appear below.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Interval</u>
*Group 2	43	10.02	9.18 to 10.87
*Group 3	87	10.61	10.07 to 11.15
*Group 4	87	10.22	9.62 to 10.82
*Group 5	87	9.82	9.20 to 10.43
Group 6	52	9.98	9.10 to 10.85

* = CAP

There was overlap of the confidence interval of each group with every other group. The group mean of Groups 2, 3, 5, and 6 lie outside of the overlap. The confidence interval provides a range within which the "true" population mean might lie with ninety-five percent confidence.

The t tests for the group mean scores recorded by both percentiles and grade equivalents produced t-values with probabilities which were statistically significant at the .05 level of significance.

	<u>t-value</u>	<u>t-probability</u>	<u>sig. at .05</u>
Percentile	-11.022	.000	yes
Grade Equivalent	-20.100	.000	yes

The analysis of variance for the data in percentiles and grade equivalents produced F values with significance levels which were not statistically significant at the .05 level of significance.

	<u>F-value</u>	<u>sig. of F</u>	<u>sig. at .05</u>
Percentile	.1405	.7081	no
Grade Equivalent	.2583	.6116	no

The analysis of covariance for the data in percentiles and grade equivalents produced F values with significance levels which were not statistically significant at the .05 level of significance.

	<u>F value</u>	<u>sig. of F</u>	<u>sig. at .05</u>
Percentiles	1.197	.275	no
Grade Equivalents	1.434	.232	no

Figure 4 provided information about the ninety-five percent confidence interval for the means of Groups 2, 3, 4, 5, and 6. There was overlap of the confidence interval of each group with every other group.

The analysis of variance, analysis of covariance, and the confidence interval ranges provide stronger statistical evidence that the null hypothesis 1.9 cannot be rejected than does the t-test statistic for rejection. The null hypothesis 1.9 cannot be rejected.

Null Hypothesis 1.10

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the spelling section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: $M_1 = \text{CAP group mean (Groups 2,3,4,5)}$

$M_2 = \text{non-CAP group mean (Groups 1,6)}$

Significance level: .05

Alternate Hypothesis 1.10

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the spelling section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = CAP group mean (Groups 2,3,4,5)

M_2 = non-CAP group mean (Groups 1,6)

Significance level: .05

Scores were reported in percentiles.

The t test was used to statistically test the difference between the group mean spelling scores of CAP and non-CAP groups. The t-value was 1.643. The two-tailed probability level of .101 was not statistically significant at the .05 level of significance.

	Posttest--DAT (8) Spelling (PR)					sig. at .05
	N	Mean	SD	t-value	t-probability	
Groups 2,3,4,5	303	46.96	28.28	1.643	.101	no
Group 6	139	50.33	26.18			

Analysis of variance was used with the F ratio as the test statistic to determine the difference among the sample means. The F-value was 1.4196 with a significance of .2341. It was not statistically significant at the .05 level of significance.

Analysis of covariance was used with the Stanford Achievement Intermediate II spelling test as the covariant. The F-value was .3534 with a significance of .061. It was not statistically significant at the .05 level of significance.

The null hypothesis 1.10 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 1.11

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the language usage section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group mean (Groups 2,3,4,5)

M_2 = non-CAP group mean (Groups 1,6)

Significance level: .05

Alternate Hypothesis 1.11

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the language usage section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = CAP group mean (Groups 2,3,4,5)

M_2 = non-CAP group mean (Groups 1,6)

Significance level: .05

Scores were reported in percentiles.

The t test was used to statistically test the difference between the group mean language usage scores of CAP and non-CAP groups. The t-value was -2.952. The two-tailed probability level of .003 was statistically significant at the .05 level of significance.

	Posttest--DAT (8) Language Usage (PR)				sig. at .05
	N	Mean	SD	t-value	
Groups 2,3,4,5	303	59.65	23.92	-2.952	yes
Groups 1,6	139	50.54	23.84		

The t-value which occurred with different combinations of CAP/non-CAP groupings also had significant levels of probability.

	Posttest--DAT (8) Language Usage (PR)			
	N	t-value	t-probability	sig. at .05
*Group 3	87	-4.395	.000	yes
Group 1	87			
*Groups 2,3	174	-3.026	.003	yes
Groups 1,6	139			
*Groups 2,3,4,5	303	-2.952	.003	yes
Groups 1,6	139			

* = CAP

Analysis of variance was used with the F-ratio as the test statistic to determine the difference among sample means. The F-value was 13.8419 with a significance of

.0002. It was statistically significant at the .05 level of significance.

Analysis of covariance was used with the Stanford Achievement Intermediate II Language Test as the covariant. The F-value was 7.493 with a significance of .006. It was statistically significant at the .05 level of significance.

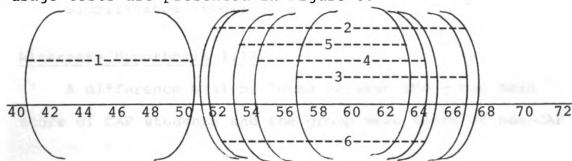
Analysis of covariance used with the independent variables being Groups 1 and Groups 3, 4, and 5 and with the Stanford Achievement Intermediate II Language Test as the covariant showed a significant F. The F value was 17.041 with a significance level of .001. It was statistically significant at the .05 level of significance.

ANCOVA-DAT Language Usage with Stanford Lang.

	N	F-value	sig. of F	Sig. at .05
*Groups 3,4,5 Group 1	261 87	17.041	.001	yes
*Groups 2,3,4,5 Groups 1,6	303 139	7.493	.006	yes

* = CAP

The ninety-five percent confidence intervals for the means of Groups 1, 2, 3, 4, 5, and 6 for the DAT language usage tests are presented in Figure 6.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Interval</u>
Group 1	87	45.86	40.94 to 50.79
*Group 2	43	59.00	51.38 to 66.62
*Group 3	87	61.68	56.93 to 66.43
*Group 4	87	59.40	54.28 to 64.52
*Group 5	87	58.16	52.75 to 63.58
Group 6	52	58.37	51.91 to 64.82

* = CAP

The range of the confidence interval for Group 1 (non-CAP) lies outside of the confidence intervals of each of the other groups. The confidence interval provides a range within which the "true" population mean might lie with ninety-five percent confidence.

The null hypothesis 1.11 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 1.12

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the verbal reasoning section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group mean (Groups 2,3,4,5)

M_2 = non-CAP group mean (Groups 1,6)

Significance level: .05

Alternate Hypothesis 1.12

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP

students on the verbal reasoning section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = CAP group mean (Groups 2,3,4,5)

M_2 = non-CAP group mean (Groups 1,6)

Significance level: .05

Scores were reported in percentiles.

The t test was used to statistically test the difference between the group mean verbal reasoning scores of CAP and non-CAP groups. The t-value was $-.837$. The two-tailed probability level of $.403$ was not statistically significant at the $.05$ level of significance.

	Posttest--DAT (8) Verbal Reasoning (PR)					sig. at .05
	N	Mean	SD	t-value	t-probability	
Groups 2,3,4,5	303	62.25	25.75	$-.837$	$.403$	no
Groups 1,6	139	59.51	26.67			

Analysis of variance was used with the F ratio as the test statistic to determine the difference among the sample means. The F value was 1.0577 with a significance of $.1793$. It was not statistically significant at the $.05$ level of significance.

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F value was 1.025 with a significance level of $.312$. It was not statistically significant at the $.05$ level of significance.

The null hypothesis 1.12 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 1.13

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP students in the English usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group mean (Groups 2,3,4,5)

M_2 = non-CAP group mean (Groups 1,6)

Significance level: .05

Alternate Hypothesis 1.13

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the English usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = CAP group mean (Groups 2,3,4,5)

M_2 = non-CAP group mean (Groups 1,6)

Significance level: .05 .

Scores were reported in percentiles.

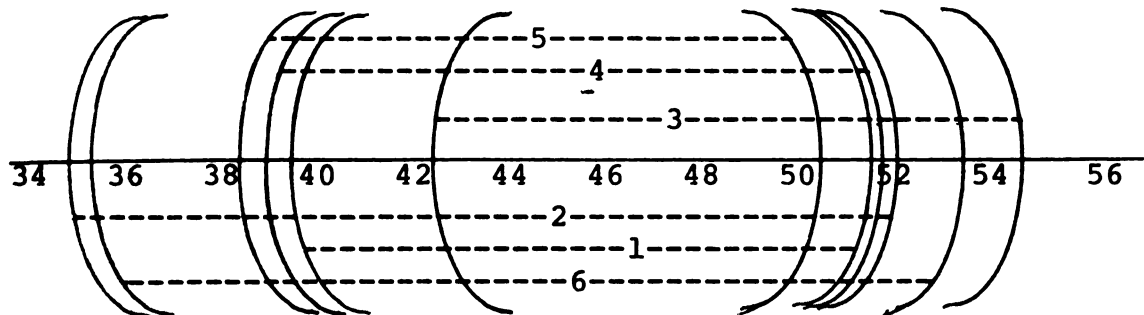
The t test was used to statistically test the differences between the group mean English usage scores of CAP and non-CAP groups. The t-value was -.201. The two-tailed probability level of .841 was not statistically significant at the .05 level of significance.

	Posttest--NEDT English Usage (PR)					sig. at .05
	N	Mean	SD	t-value	t-probability	
Groups 2,3,4,5	304	45.60	28.76	-.085	.933	no
Groups 1,6	139	45.18	29.89			

Analysis of variance was used with the F ratio as the test statistic to determine the difference among the sample means. The F-value was .0200 with a significance of .8875. It was not statistically significant at the .05 level of significance.

Analysis of covariance was used with the Stanford Achievement Intermediate II Language Test as the covariant. The F-value was 5.174 with a significance level of .023. It was statistically significant at the .05 level of significance.

The ninety-five percent confidence intervals for the means of Groups 1, 2, 3, 4, 5, and 6 are presented in Figure 7 for NEDT English usage.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Interval</u>
Group 1	87	45.55	39.46 to 51.64
*Group 2	87	43.35	35.00 to 51.90
*Group 3	87	48.24	42.09 to 54.39
*Group 4	87	45.47	39.08 to 51.87
*Group 5	87	44.21	28.14 to 50.28
Group 6	52	44.56	35.58 to 53.53

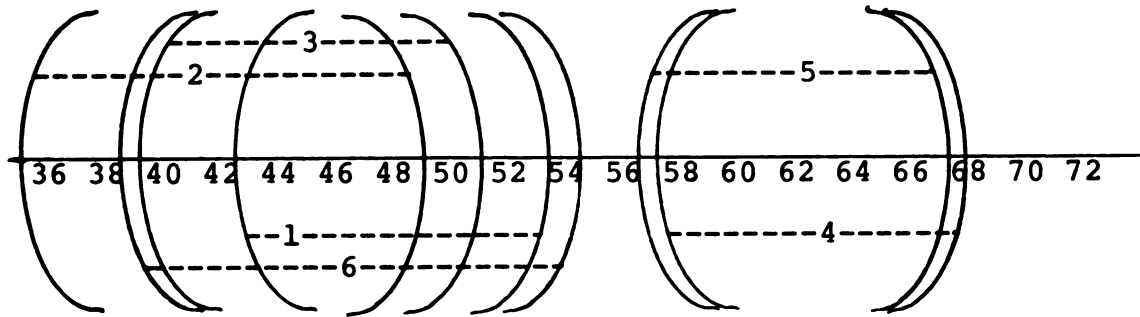
* = CAP

There was overlap of the confidence intervals of each group with every other group. The group mean of each group lies within the overlap. The confidence interval provides a range within which the "true" population mean might lie with ninety-five percent confidence.

The t test was used to statistically test the difference between the group mean Stanford Language scores of the CAP and non-CAP students (the covariant used in the analysis of covariance above). The t-value was -2.050. The two-tailed probability level of .041 was not statistically significant at the .05 level of significance.

Analysis of variance was used to determine the difference among the sample means on the Stanford Language Test. The F-value was 6.3980 with a significance level of .0118. It was statistically significant at the .05 level of significance.

The ninety-five percent confidence intervals for the means of Groups 1, 2, 3, 4, 5, and 6 are presented in Figure 8 for Stanford Language.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Interval</u>
Group 1	87	48.22	42.55 to 53.89
*Group 2	87	42.51	35.59 to 49.43
*Group 3	87	45.21	39.45 to 50.97
*Group 4	87	62.41	57.13 to 67.70
*Group 5	87	61.78	56.36 to 67.20
Group 6	52	46.50	38.96 to 54.04

* = CAP

The range of the confidence intervals for Groups 4 and 5 lie outside the confidence intervals of the other groups. The confidence interval and the group means are very similar for Groups 4 and 5.

The .05 level of significance was reached in the analysis of covariance with the Stanford Achievement Intermediate II Language Test as the covariant. There appears to be a difference. The null hypothesis 1.13 cannot be accepted.

Null Hypothesis 1.14

No difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the word usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = CAP group mean (Groups 2,3,4,5)

M_2 = non-CAP group mean (Groups 1,6)

Significance Level: .05

Alternate Hypothesis 1.14

A difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the word usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = CAP group mean (Groups 2,3,4,5)

M_2 = non-CAP group mean (Groups 1,6)

Significance Level: .05

Scores were reported in percentiles.

The t test was used to statistically test the difference between the group mean word usage scores of CAP and non-CAP groups. The t-value was -.792. The two-tailed probability level of .442 was not statistically significant at the .05 level of significance.

<u>Posttest--NEDT Word Usage (PR)</u>						sig. at .05
	N	Mean	SD	t-value	t-probability	
Groups 2,3,4,5	304	50.38	27.37	-.769	.442	no
Groups 1,6	139	48.15	28.42			

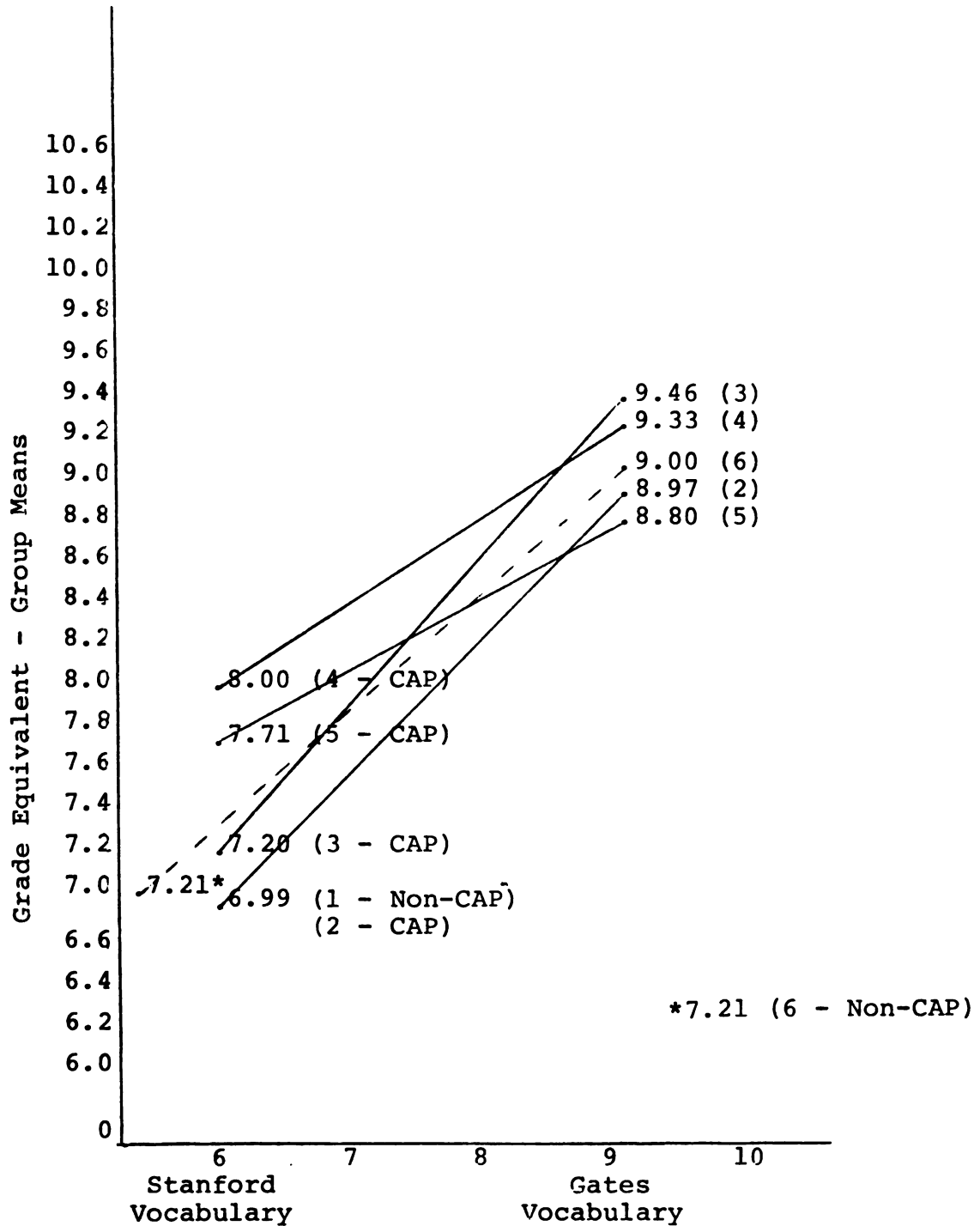
Analysis of variance was used with the F-ratio as the test statistic to determine the difference among the sample

GRAPH 4:1.5

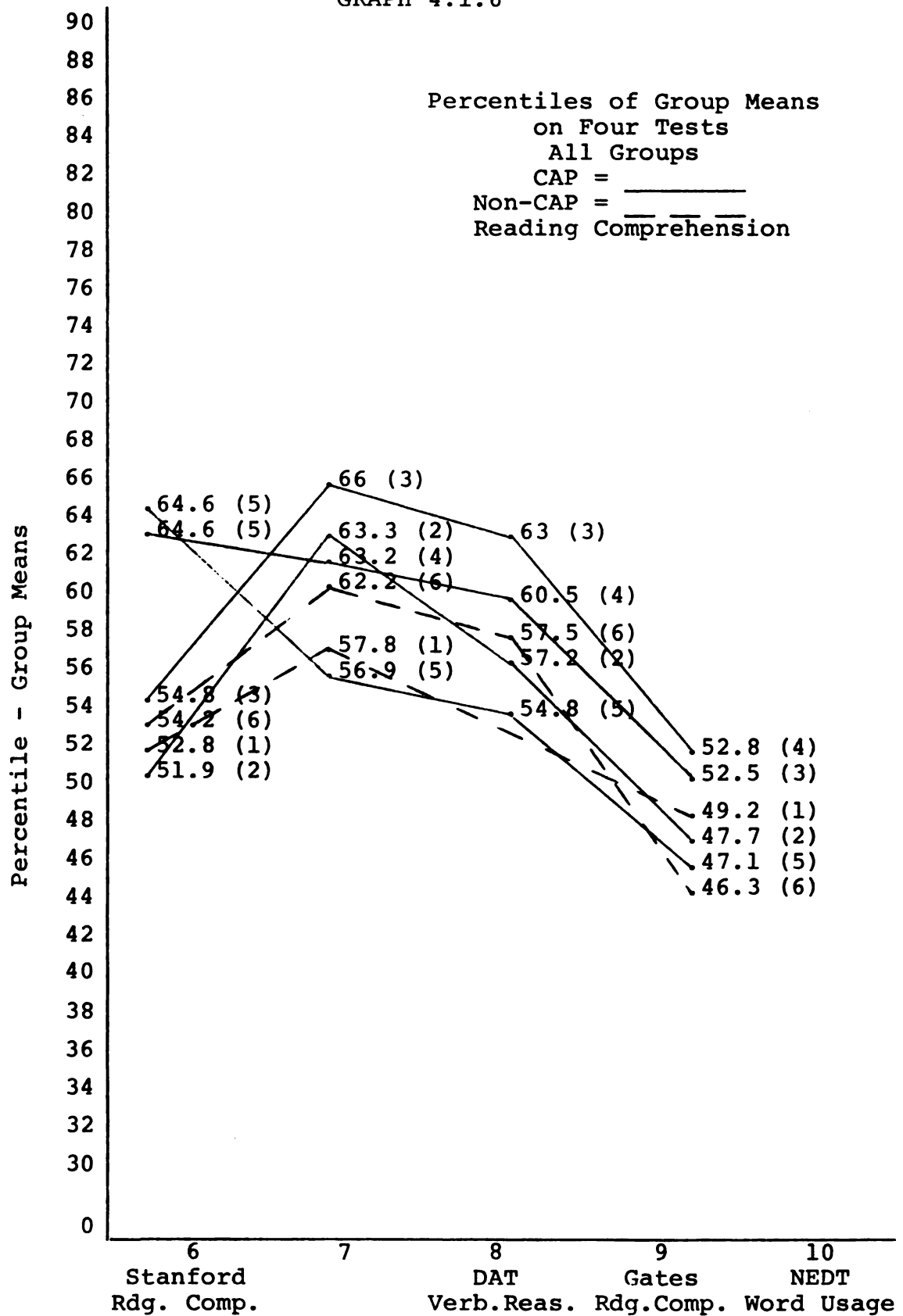
Grade Equivalents of
Group Means on
2 tests
All Groups

CAP = _____

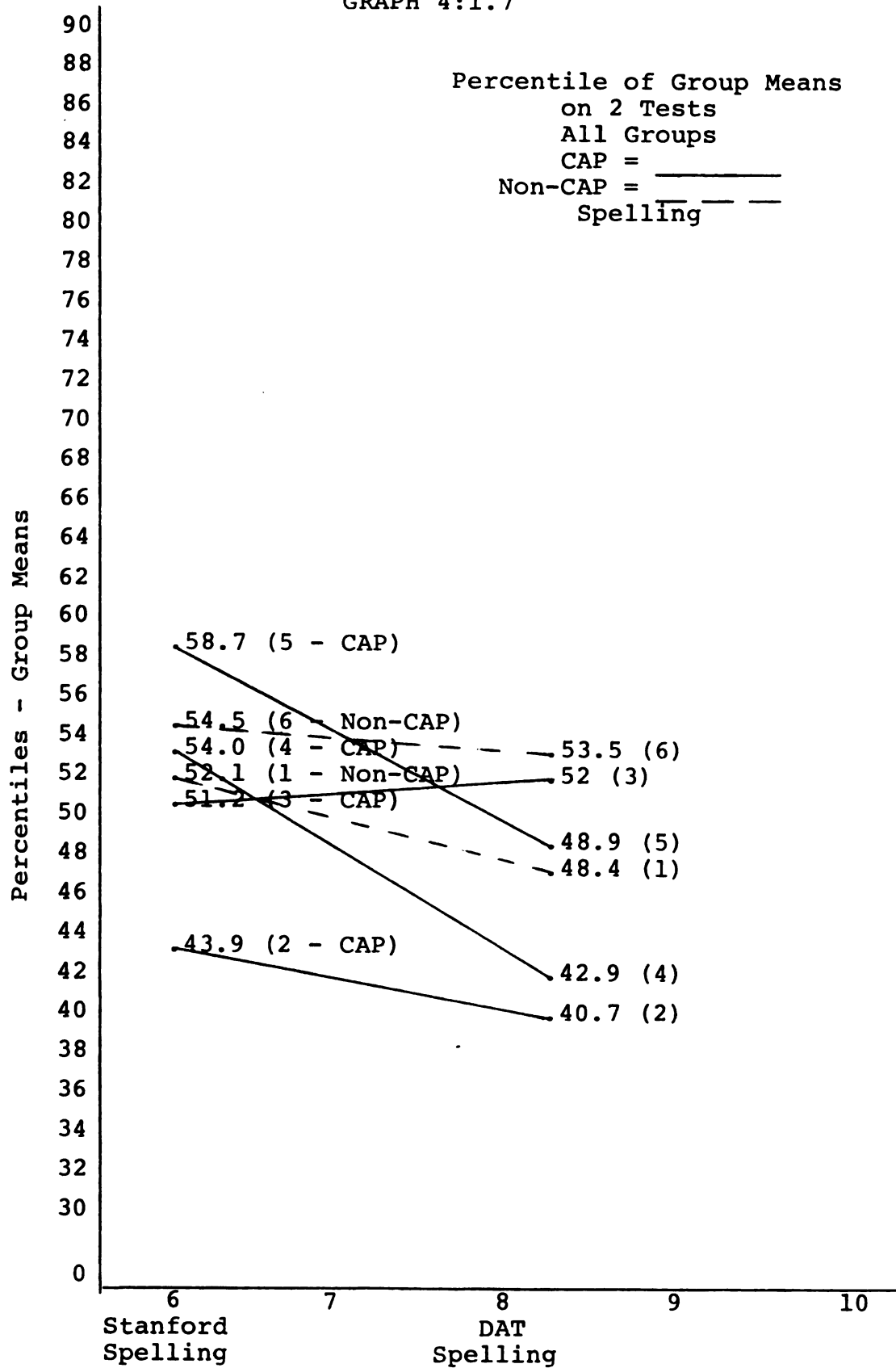
Non-CAP = _____



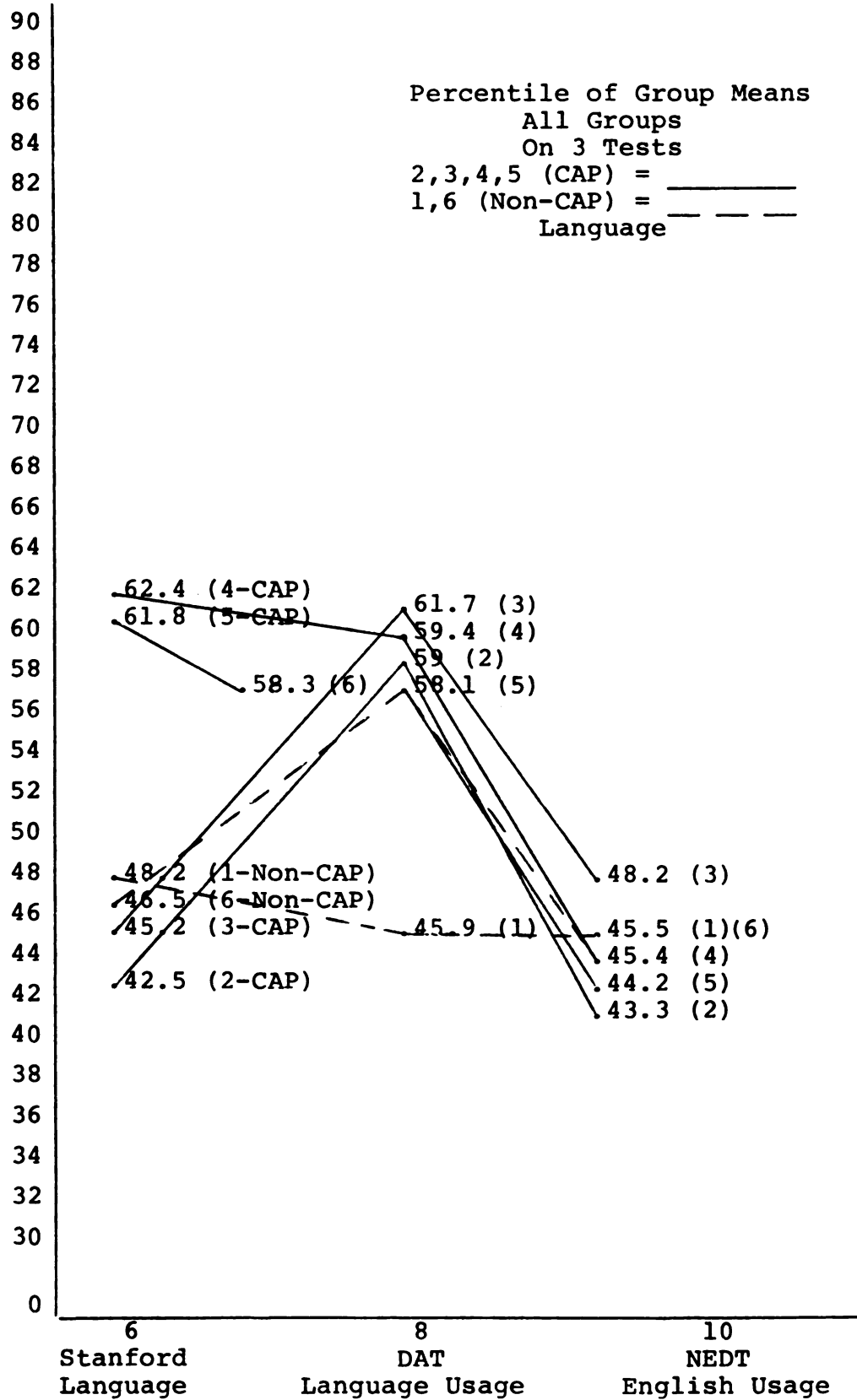
GRAPH 4:1.6



GRAPH 4:1.7



GRAPH 4:1.8



means. The F-value was .6203 with a significance of .4314. It was not statistically significant at the .05 level of significance.

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F-value was 3.041 with a significance level of .082. It was not statistically significant at the .05 level.

The null hypothesis 1.14 cannot be rejected. The .05 level of significance was not reached.

Research Question 2: Hypothesis and Analysis

Research Question 2: is the effect of CAP upon student groups consistent over time, as measured by standardized tests?

Groups 2 and 3 are the groups referred to in Hypotheses 2.1 to 2.9. Groups 3, 4, and 5 were analyzed in Hypotheses 2.10 to 2.18. Groups 4 and 5 were analyzed in Hypotheses 2.19 to 2.25.

The analysis of covariance was used to determine difference in the effect of CAP during different years of the program. The only group which had had students randomly assigned to it was Group 2. Group 2 was the first group of students to have CAP. They had the program for one year. Groups 3, 4, and 5 each had CAP for two years. All students in Groups 3, 4, and 5 were in CAP.

Null Hypothesis 2.1

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05.

Alternate Hypothesis 2.1

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Scores were reported in two forms, percentiles and grade equivalents.

Percentiles

Analysis of covariance was used with the Stanford Achievement Intermediate II vocabulary Test as the

covariant. The F-value was .507 with a significance of .478. It was not statistically significant at the .05 level of significance.

Grade Equivalents

Analysis of covariance was used with the Stanford Achievement Intermediate II vocabulary test as the covariant. The F-value was .644 with a significance of .424. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.1 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.2

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Alternate Hypothesis 2.2

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the comprehension

section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Scores were reported in two forms, percentiles and grade equivalents.

Percentiles

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F-value was 1.218 with a significance level of .272. It was not statistically significant at the .05 level of significance.

Grade Equivalents

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F-value was 1.824 with a significance level of .179. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.2 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.3

No difference will be found between the group mean score of students in one CAP group and the group mean score

of students in a second CAP group on the spelling section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Alternate Hypothesis 2.3

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the spelling section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II spelling test as the covariant. The F-value was 3.107 with a significance level of .080. It was not significant at the .05 level of significance.

The null hypothesis 2.3 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.4

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the language usage section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = asecond CAP group mean (Group 3)

Significance Level: .05

Alternate Hypothesis 2.4

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the language usage section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II Language test as the covariant. The F-value was .113 with a significance level of .738. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.4 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.5

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the verbal reasoning section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Alternate Hypothesis 2.5

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the verbal reasoning section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F-value was .037 with a significance

level of .847. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.5 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.6

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the English usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Alternate Hypothesis 2.6

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the English usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford

Achievement Intermediate II language section as the covariant. The F-value was .564 with a significance level of .454. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.6 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.7

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the word usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Alternate Hypothesis 2.7

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the word usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II reading comprehension section of the National Educational Development Test as the covariant. The F-value was .194 with a significance level of .660. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.7 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.8

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Alternate Hypothesis 2.8

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Scores were reported in grade equivalents.

Analysis of covariance was used with the Gates-MacGinitie Reading Test, vocabulary section, administered in the seventh grade, as the covariant. The F-value was .245 with a significance level of .622. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.8 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.9

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the comprehension section of the Gates-MacGinitie Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

Significance Level: .05

Alternate Hypothesis 2.9

A difference will be found between the group mean score of students in one CAP group and the group mean score of

students in a second CAP group on the comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 2)

M_2 = second CAP group mean (Group 3)

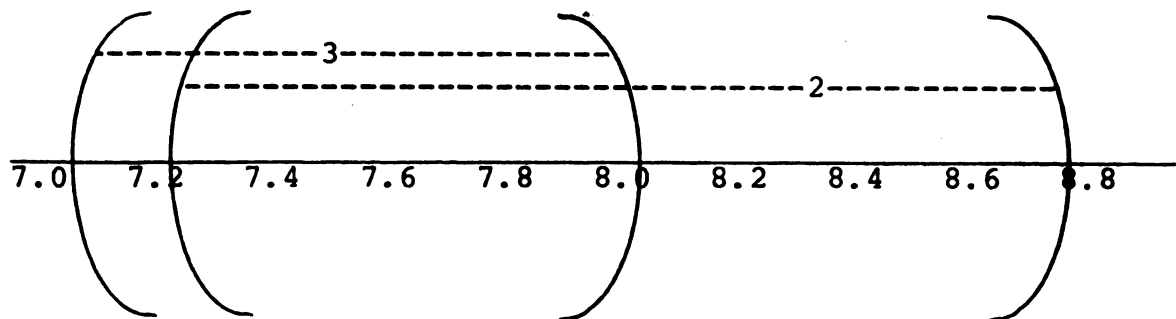
Significance Level: .05

Scores were reported in grade equivalents.

Analysis of covariance was used with the Gates-MacGinite Reading Test, comprehension section, administered in the seventh grade, as the covariant. The F-value was 6.289 with a significance level of .013. It was statistically significant at the .05 level of significance.

The mean of Group 2 on the Gates (7) comprehension section was 8.00. The mean of group 3 on the comprehension section was 7.6

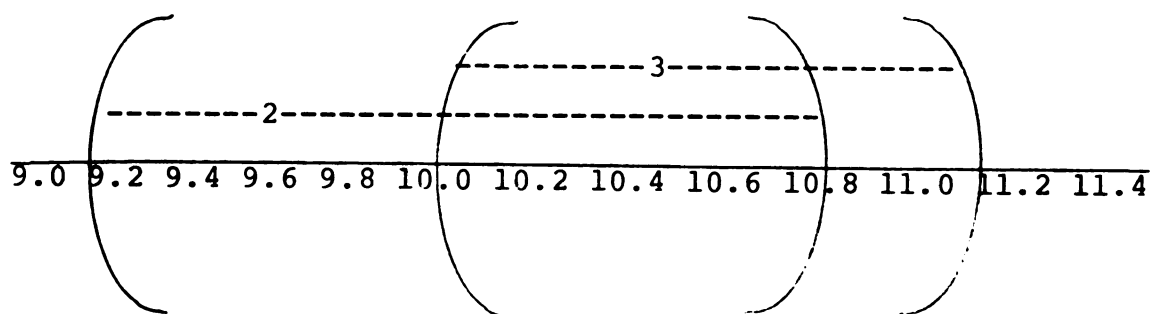
The ninety-five percent confidence intervals for the mean of Groups 2 and 3 for the comprehension section of the Gates-MacGinite Reading Test administered in the seventh grade are shown in Figure 9.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Interval</u>
Group 2	43	8.00	7.23 to 8.78
Group 3	87	7.60	7.07 to 8.13

The mean of Group 2 (8.00) is near the upper end of the confidence interval for Group 3 (8.13).

The ninety-five percent confidence interval for the means of Groups 2 and 3 on the comprehension section of the Gates-MacGinitie Reading Test administered in the ninth grade are shown in Figure 10.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Interval</u>
Group 2	43	10.02	9.18 to 10.87
Group 3	87	10.61	10.01 to 11.15

The overlap of the ranges had changed.

The null hypothesis 2.9 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 2.10

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section

of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

Significance Level: .05

Alternate Hypothesis 2.10

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_1 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

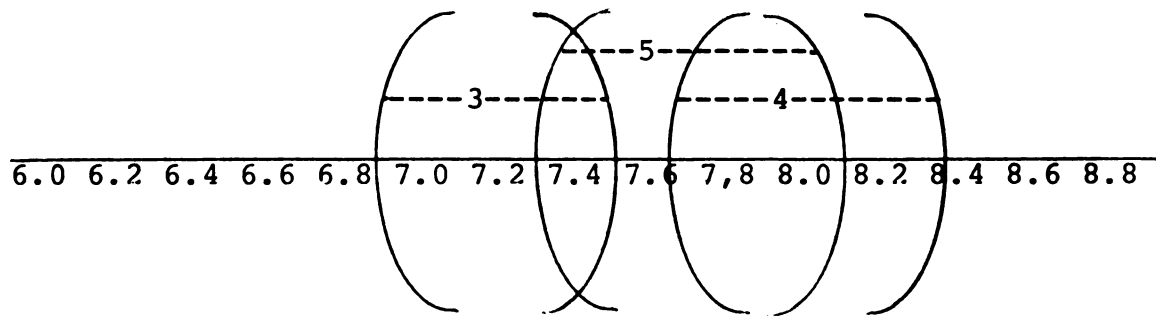
Significance Level: .05

Scores were reported in grade equivalents.

Analysis of covariance was used with the Stanford Achievement Intermediate II vocabulary section as the co-variant. The F-value was 21.465 with a significance level of .001. It was statistically significant at the .05 level of significance.

The mean of Group 3 on the vocabulary section of the Stanford was 7.20. The mean for Group 4 was 7.99. The mean for Group 5 was 7.72

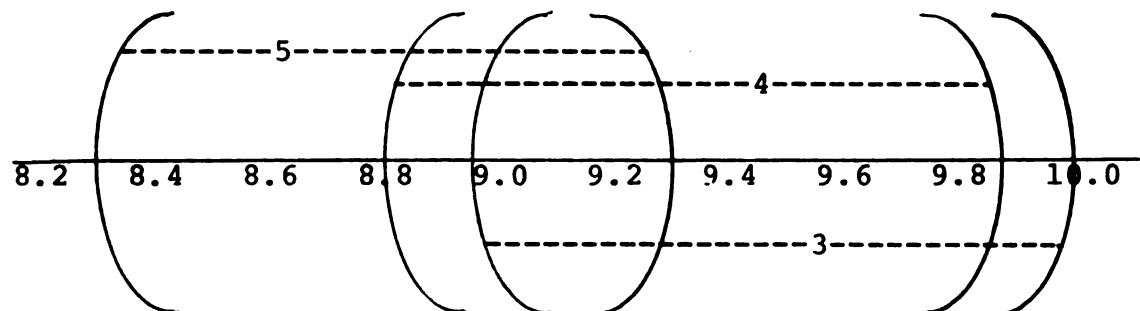
The ninety-five percent confidence intervals for the means of Groups 3, 4, and 5 for the vocabulary section of the Stanford are shown in Figure 11.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 3	87	7.21	6.90 to 7.52
Group 4	87	8.00	7.64 to 8.36
Group 5	87	7.72	7.32 to 8.11

The mean of Group 3 was outside the lower limits of Groups 4 and 5. The upper limit of the range of Group 3 was outside the lower limit of the range of Group 4.

The ninety-five percent intervals for the same groups on the vocabulary section of the Gates-MacGinitie, ninth grade, are shown in Figure 12.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Interval</u>
Group 3	87	9.46	8.96 to 9.97
Group 4	87	9.33	8.80 to 9.87
Group 5	87	8.80	8.30 to 9.29

There was a shift in position of the three groups in relation to each other.

The null hypothesis 2.10 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 2.11

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group in the comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

Significance level: .05

Alternate Hypothesis 2.11

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

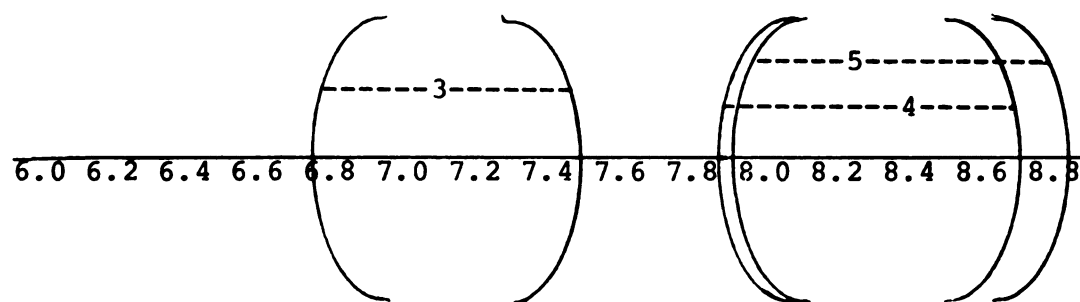
Significance Level: .05

Scores were reported in grade equivalents.

Analysis of covariance was used with the Stanford Achievement Intermediate II comprehension section as the co-variant. The F-level was 22.149 with a significance level of .001. It was statistically significant at the .05 level of significance.

The mean of Group 3 on the Stanford was 7.14. The mean of Group 4 was 8.30. The mean of Group 5 was 8.35.

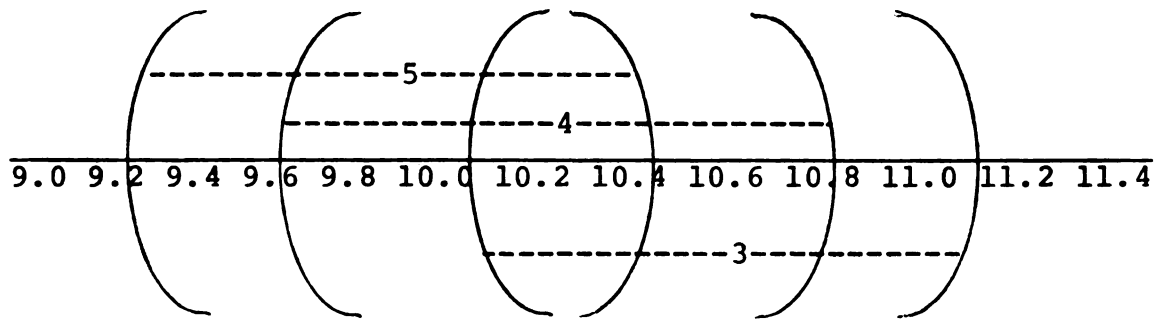
The ninety-five percent confidence intervals for the means of Groups 3, 4, and 5 for the comprehension section of the Stanford are shown in Figure 13.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 3	87	7.14	6.78 to 7.50
Group 4	87	8.30	7.89 to 8.71
Group 5	87	8.35	7.90 to 8.80

There was no overlap of the confidence interval of Group 3 with those of Groups 4 and 5.

The ninety-five percent confidence intervals of Groups 3, 4, and 5 on the Gates-MacGinite comprehension section, ninth grade, are shown in Figure 14.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 3	87	10.61	10.07 to 11.15
Group 4	87	10.22	9.62 to 10.82
Group 5	87	9.82	9.20 to 10.43

There appears to have been a shift in relative positions of the three groups, indicated by the range of the confidence intervals.

The null hypothesis 2.11 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 2.12

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the spelling section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

Significance Level: .05

Alternate Hypothesis 2.12

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the spelling section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

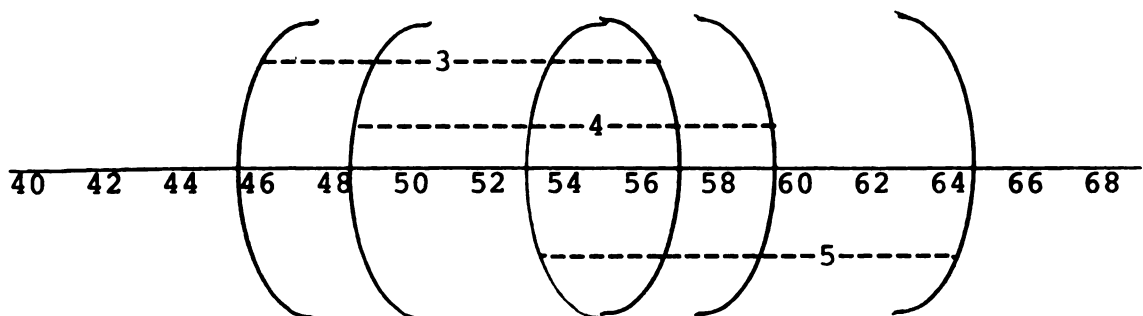
Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II spelling test as the covariant. The F-value was 22.149 with a significance level of .001. It was statistically significant at the .05 level of significance.

The mean of Group 3 for the Stanford spelling test was 51.21. The mean of Group 4 was 54.09. The mean of Group 5 was 58.71.

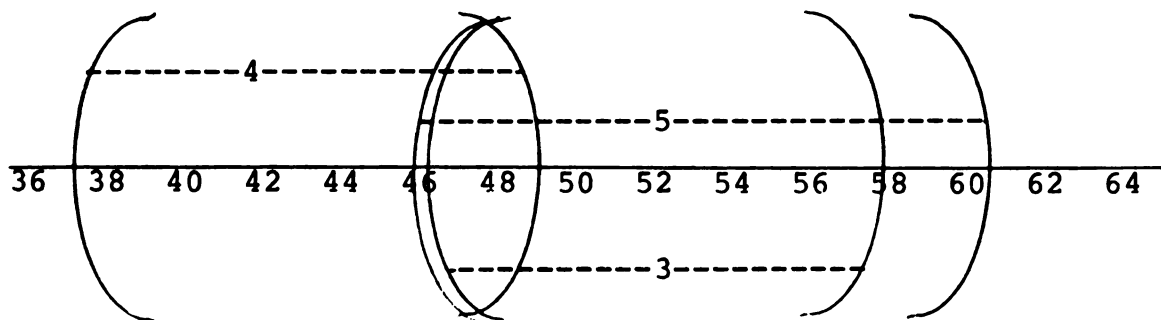
The ninety-five percent confidence intervals for the means of Groups 3, 4, and 5 for the spelling section of the Stanford are shown in Figure 15.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 3	87	51.21	45.59 to 56.82
Group 4	87	54.09	48.69 to 59.50
Group 5	87	58.71	52.83 to 64.60

The mean of Group 4 lies within the overlap of the intervals of the three groups. The mean of Group 3 lies below it and that of Group 5 lies above it.

The ninety-five percent confidence interval of Groups 3, 4, and 5 on the spelling section of the Differential Aptitude Test is shown in Figure 16.



	<u>N</u>	<u>Mean</u>	<u>95% Confidence Interval</u>
Group 3	87	52.12	46.28 to 57.97
Group 4	87	42.95	37.16 to 48.75
Group 5	87	48.92	46.60 to 60.40

There has been an apparent shift in position, but it is most evident between the Groups 4 and 5.

The null hypothesis 2.12 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 2.13

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the language usage

section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

Significance Level: .05

Alternate Hypothesis 2.13

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the language usage section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

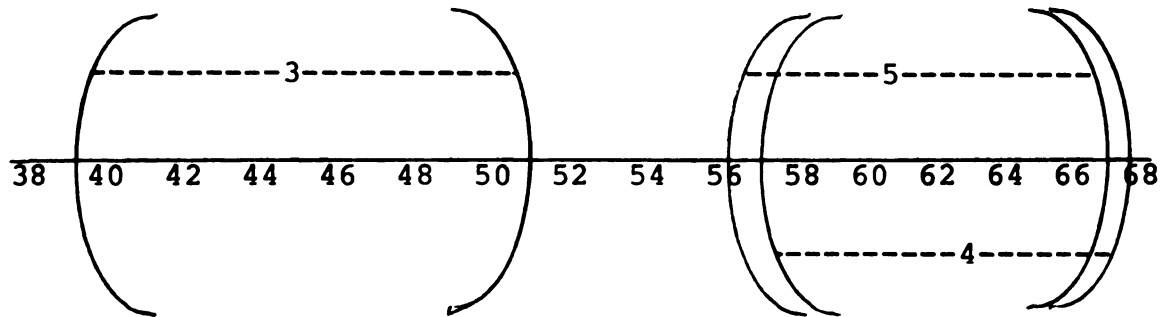
Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II Language Test as the covariant. The F-value was 33.766 with a significance level of .001. It was statistically significant at the .05 level of significance.

The mean of Group 3 for the Stanford language test was 45.20. The mean of Group 4 was 62.41. The mean of Group 5 was 61.78.

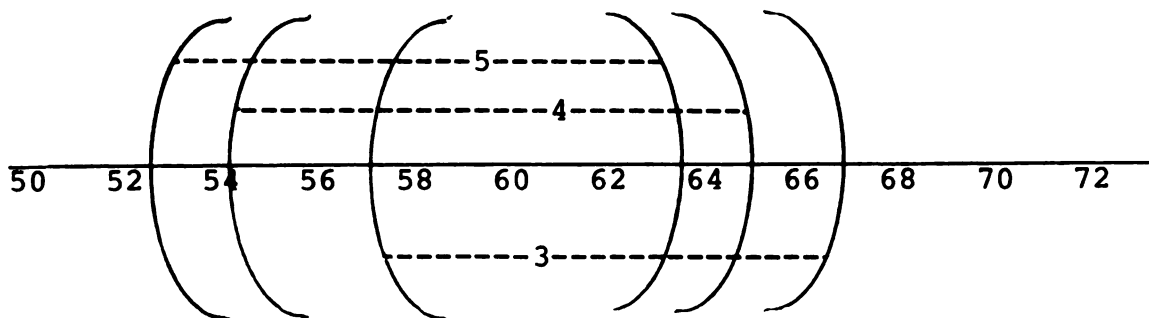
The ninety-five percent confidence intervals for the means of Groups 3, 4, and 5 for the language section of the Stanford are shown in Figure 17.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 3	87	45.20	39.45 to 50.97
Group 4	87	62.41	57.13 to 67.70
Group 5	87	61.78	56.36 to 67.20

The upper limit of the range of the confidence interval for Group 3 was below the lower limit of the range for Groups 4 and 5. There is no overlap between Group 3 and Groups 4 and 5.

The ninety-five percent confidence intervals for the means of Groups 3, 4, and 5 for the language section of the DAT are shown in Figure 18.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 3	87	61.68	56.93 to 66.43
Group 4	87	59.40	54.28 to 64.52
Group 5	87	58.12	52.75 to 63.58

The shift in relative position of the range of confidence intervals is shown. In the covariant the ranges did not have an area of overlap (between Group 3 and Groups 4 and 5). In the posttest, the range of confidence intervals indicates quite an overlap and all of the group sample means are within the overlap.

The null hypothesis 2.13 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 2.14

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the verbal reasoning section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

Significance Level: .05

Alternate Hypothesis 2.14

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the verbal reasoning

section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

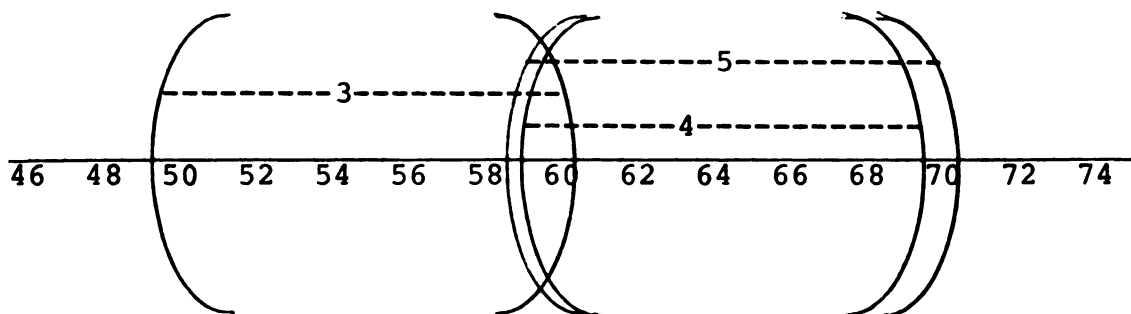
Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F-value was 30.88 with a significance level of .001. It was statistically significant at the .05 level of significance.

The mean of Group 3 for the Stanford reading comprehension test was 54.87. The mean of Group 4 was 64.43. The mean of Group 5 was 64.60.

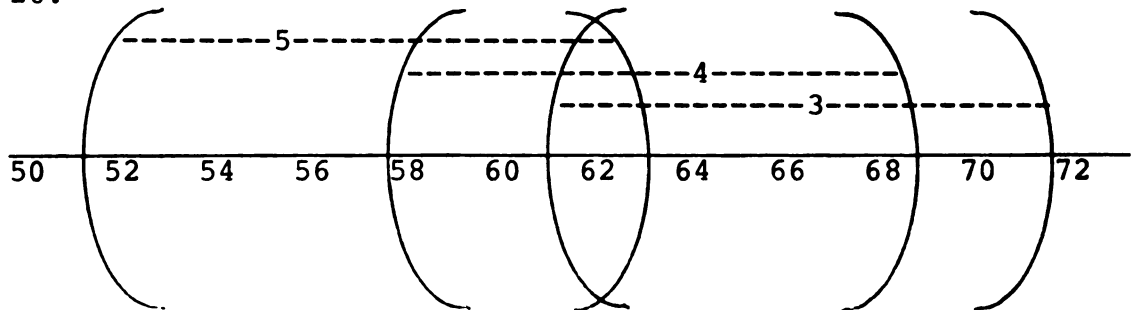
The ninety-five percent confidence interval for the means of Groups 3, 4, and 5 for the reading comprehension section of the Stanford are shown in Figure 19.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 3	87	54.87	49.62 to 60.12
Group 4	87	64.43	59.32 to 69.53
Group 5	87	64.60	59.15 to 70.09

There was an overlap of the confidence intervals of the three groups.

The ninety-five percent confidence interval for the means of Groups 3, 4, and 5 for the verbal reasoning section of the Differential Aptitude Test are shown in Figure 20.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 3	87	66.00	60.97 to 71.03
Group 4	87	63.25	57.89 to 68.62
Group 5	87	56.92	51.02 to 62.82

Group 3 had shifted position in relation to Groups 4 and 5. In the figure showing means in intervals on the Stanford Reading Comprehension Test, the range of Group 3 was the lowest range. In Figure 20, it has the highest range.

The null hypothesis 2.14 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 2.15

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the English usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

Significance Level: .05

Alternate Hypothesis 2.15

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the English usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

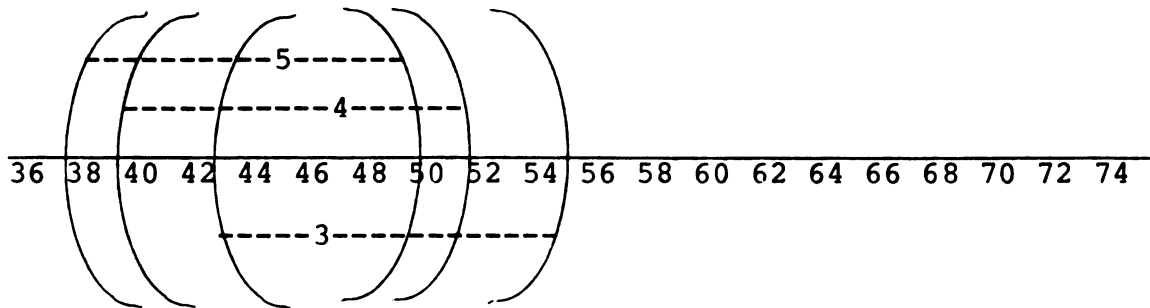
Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II Language Test as the covariant. The F-value was 40.109 with a significance level of .001. It was statistically significant at the .05 level of significance.

See Figure 17 for information on the ninety-five percent confidence intervals for the means of Groups 3, 4, and 5 of the Stanford Language test.

The ninety-five percent confidence interval for the means of Groups 3, 4, and 5 for the English usage section of the NEDT, administered in the tenth grade, is shown in Figure 21.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 3	87	48.24	42.09 to 54.39
Group 4	87	45.47	39.08 to 51.87
Group 5	87	44.21	38.14 to 50.28

The relative position of confidence intervals has changed. In Figure 17 there was no overlap between the confidence interval of Group 3 and those of Groups 4 and 5.

The null hypothesis 2.15 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 2.16

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the word usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

Significance Level: .05

Alternate Hypothesis 2.16

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the word usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F-value was 23.129 with a significance level of .001. It was statistically significant at the .05 level of significance.

See Figure 19 for information on the ninety-five percent confidence intervals for the means of Groups 3, 4, and 5 of the Stanford Achievement Intermediate II Reading Comprehension test.

The null hypothesis 2.16 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 2.17

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

Significance Level: .05

Alternate Hypothesis 2.17

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

Significance Level: .05

Scores were reported in grade equivalents.

Analysis of covariance was used with the Gates-MacGinite REading Test, vocabulary section, administered in the seventh grade, as the covariant. The F-value was 1.63 with a significance level of .687. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.17 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.18

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

Significance Level: .05

Alternate Hypothesis 2.18

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 3)

M_2 = second CAP group mean (Groups 4,5)

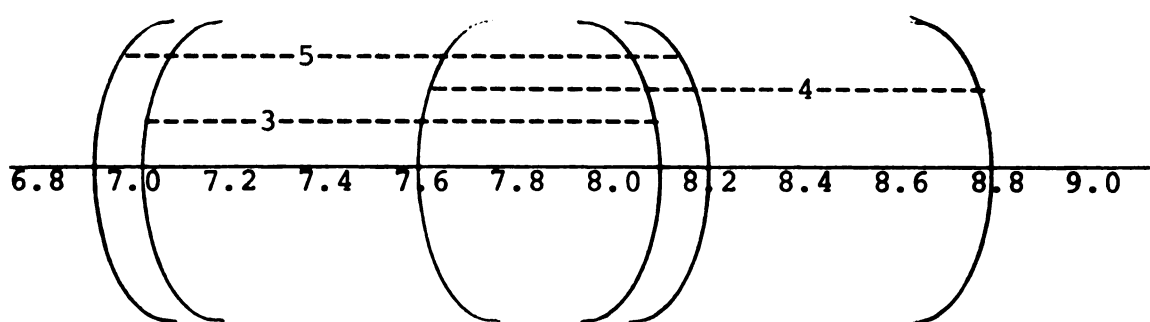
Significance Level: .05

Scores were reported in grade equivalents.

Analysis of covariance was used with the Gates-MacGinite Reading Test, comprehension section, administered in the seventh grade, as the covariant. The F-value was

11.781 with a significance level of .001. It was statistically significant at the .05 level of significance.

The ninety-five percent confidence intervals for the means of Groups 3, 4, and 5 for the comprehension section of the Gates-MacGinite Reading test, administered in the seventh grade, are shown in Figure 22.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 3	87	7.60	7.07 to 8.13
Group 4	87	8.21	7.60 to 8.82
Group 5	87	7.56	6.94 to 8.19

None of the group means is in the overlap area. The overlap is below the group mean of Group 4 and above the group mean of Groups 3 and 5.

The null hypothesis 2.18 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 2.19

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Alternate Hypothesis 2.19

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Scores were reported in grade equivalents.

Analysis of covariance was used with the Stanford Achievement Intermediate II Vocabulary Test as the covariant. The F-value was 1.034 with a significance level of .311. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.19 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.20

No difference will be found between the group mean score of students in one CAP group and the group mean score

of students in a second CAP group on the reading comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Alternate Hypothesis 2.20

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the reading comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Scores were reported in grade equivalents.

Analysis of covariance was used with the Stanford Achievement Intermediate II Comprehension Test as the covariant. The F-value was 2.513 with a significance level of .115. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.20 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.21

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the spelling section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Alternate Hypothesis 2.21

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the spelling section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Score were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II Spelling Test as the covariant. The F-value was .676 with a significance level of .412. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.21 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.22

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the language usage section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Alternate Hypothesis 2.22

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the language usage section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Scores were reported in percentiles

Analysis of covariance was used with the Stanford Achievement Intermediate II Language Test as the covariant. The F-value was .058 with a significance level of .810. It

was not statistically significant at the .05 level of significance.

The null hypothesis 2.22 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.23

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the verbal reasoning section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Alternate Hypothesis 2.23

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the verbal reasoning section of the Differential Aptitude Test administered in the eighth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F-value was 5.541 with a significance level of .020. It was statistically significant at the .05 level of significance.

The ninety-five percent confidence intervals for the means of Groups 4 and 5 for the reading comprehension section of the Stanford test were shown in Figure 19. The intervals are very similar. The means of both groups are very similar:

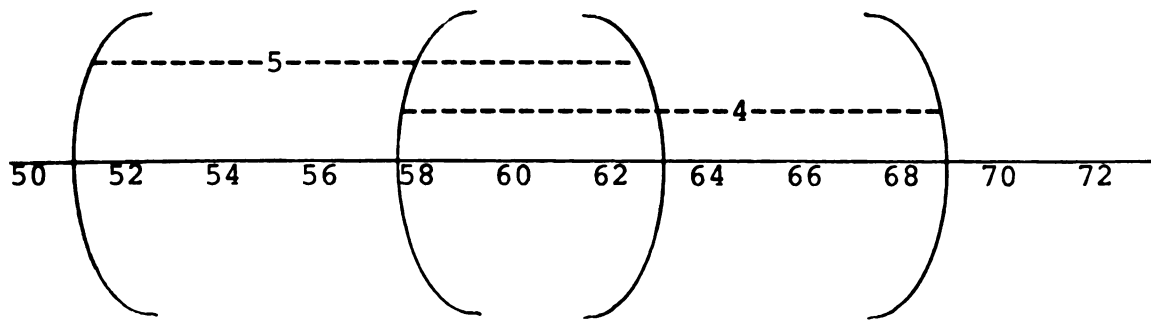
Group 4 mean = 64.43

Group 5 mean = 64.60

The t-test was used to statistically test the difference in the group mean Stanford Achievement Intermediate II Reading Comprehension scores between Group 4 and Group 5. The t-value was -.050. The two-tailed probability level of .960 was not statistically significant at the .05 level of significance.

The t-test was also used to statistically test the difference in group mean Differential Aptitude Test verbal reasoning scores between Group 4 and Group 5. The t-value was 1.605. The two-tailed probability level of .109 was not statistically significant at the .05 level of significance.

The ninety-five percent confidence intervals for the means of Groups 4 and 5 for the verbal reasoning section of the Differential Aptitude Test are shown in Figure 23.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 4	87	63.25	57.89 to 68.62
Group 5	87	56.92	51.02 to 62.82

The area of overlap of the confidence intervals in Figure 23 was reduced from the area of overlap shown in Figure 19. The groups had shifted from being similar to being not so similar.

The null hypothesis 2.23 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 2.24

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the English usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Alternate Hypothesis 2.24

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the English usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II Language Test as the covariant. The F-value was .057 with a significance level of .811. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.24 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.25

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the word usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 4)
 M_2 = second CAP group mean (Group 5)
 Significance Level: .05

Alternate Hypothesis 2.25

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the word usage section of the National Educational Development Test administered in the tenth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 4)
 M_2 = second CAP group mean (Group 5)

Significance Level: .05

Scores were reported in percentiles.

Analysis of covariance was used with the Stanford Achievement Intermediate II Reading Comprehension Test as the covariant. The F-value was 5.777 with a significance level of .017. It was statistically significant at the .05 level of significance.

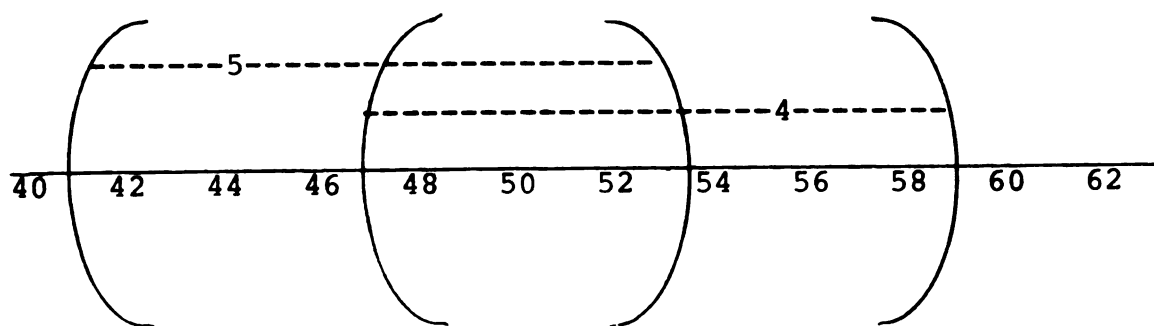
The information about confidence intervals for Groups 4 and 5 was provided in Figure 19. The ranges were very similar. The means were very similar:

Group 4 mean = 64.43

Group 5 mean = 64.60

The ninety-five percent confidence intervals for the means of Groups 4 and 5 for the word usage section of the

National Educational Development Test are shown in Figure 24.



	<u>N</u>	<u>Mean</u>	<u>95% of Confidence Intervals</u>
Group 4	87	52.82	47.03 to 58.60
Group 5	87	47.10	40.94 to 53.27

The mean of Group 4 was just below the top of the confidence interval for Group 5. The mean of Group 5 was just above the bottom of the confidence interval for Group 4.

The null hypothesis 2.25 cannot be accepted. The .05 level of significance was reached.

Null Hypothesis 2.26

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Alternate Hypothesis 2.26

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 4)

M_2 = second CAP group mean (Group 5)

Significance Level: .05

Scores were reported in grade equivalents.

Analysis of covariance was used with the Gates-MacGinite Reading Test, vocabulary section, administered in the seventh grade, as the covariant. The F-value was 1.934 with a significance level of .166. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.26 cannot be rejected. The .05 level of significance was not reached.

Null Hypothesis 2.27

No difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the reading comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 = M_2$

Legend: M_1 = one CAP group mean (Group 4)
 M_2 = second CAP group mean (Group 5)
 Significance Level: .05

Alternate Hypothesis 2.27

A difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the reading comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

Symbolically: $H_0 : M_1 \neq M_2$

Legend: M_1 = one CAP group mean (Group 4)
 M_2 = second CAP group mean (Group 5)

Significance Level: .05

Scores were reported in grade equivalents.

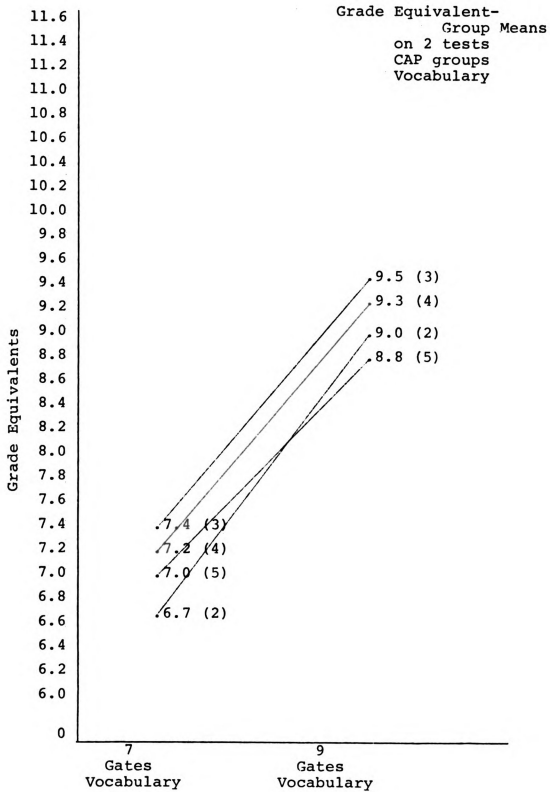
Analysis of covariance was used with the Gates-MacGinite Reading Test Comprehension Section, administered in the seventh grade, as the covariant. The F-value was .144 with a significance level of .705. It was not statistically significant at the .05 level of significance.

The null hypothesis 2.27 cannot be rejected. The .05 level of significance was not reached.

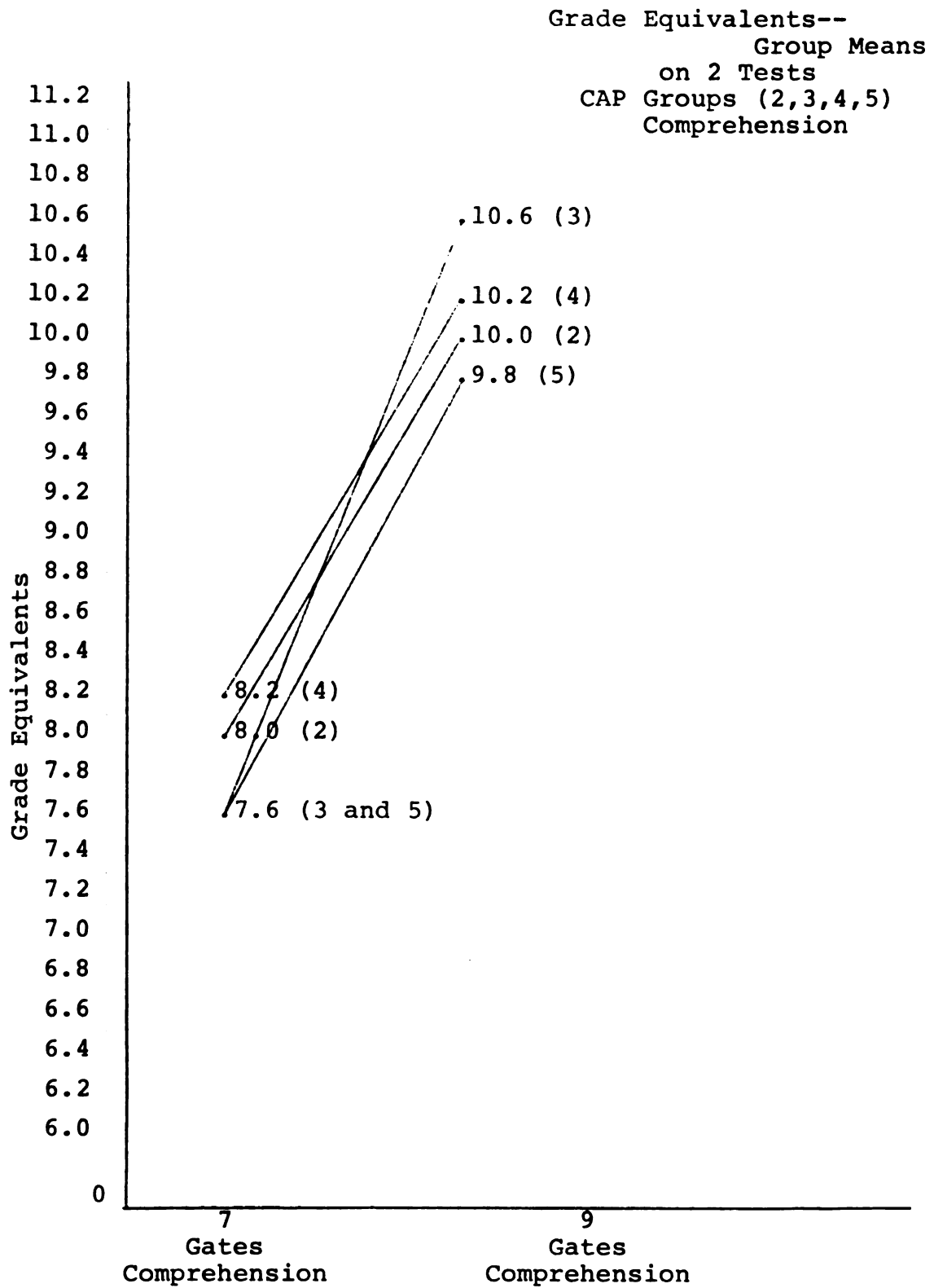
Research Question 3: Hypothesis and Analysis

Research Question 3: is there a difference in the effect of CAP for students of varying reading abilities?

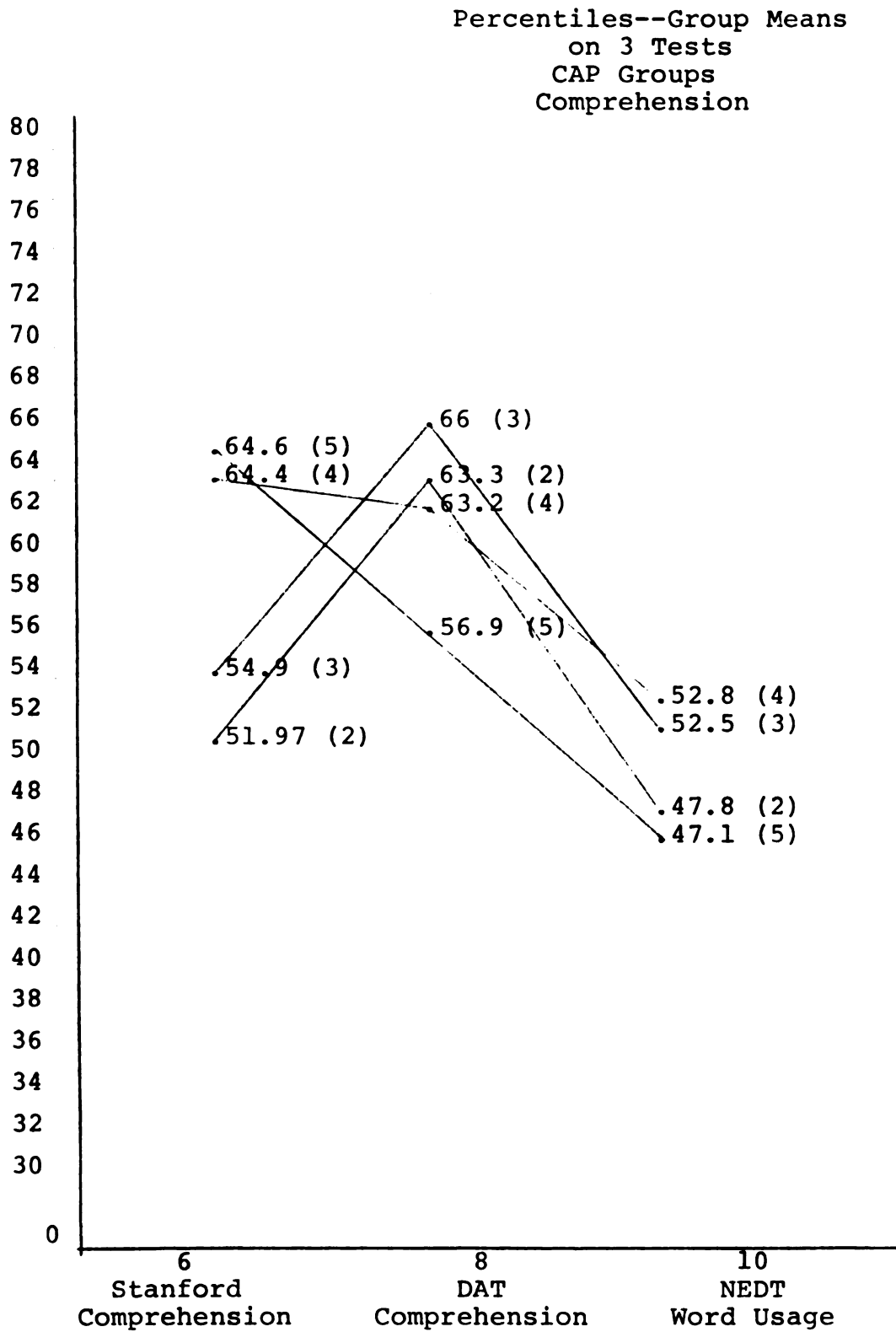
GRAPH 4:2.1



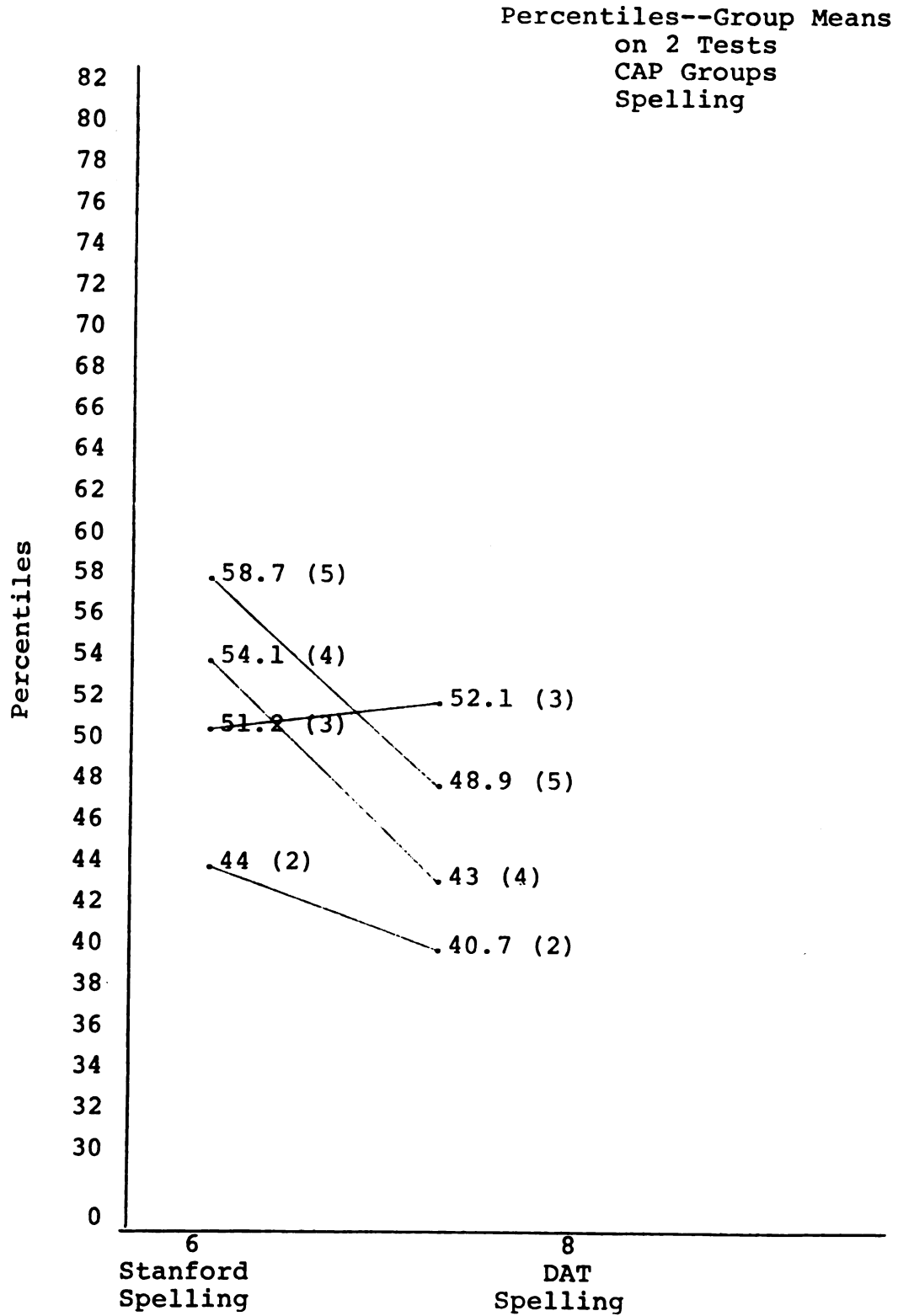
GRAPH 4:2.2



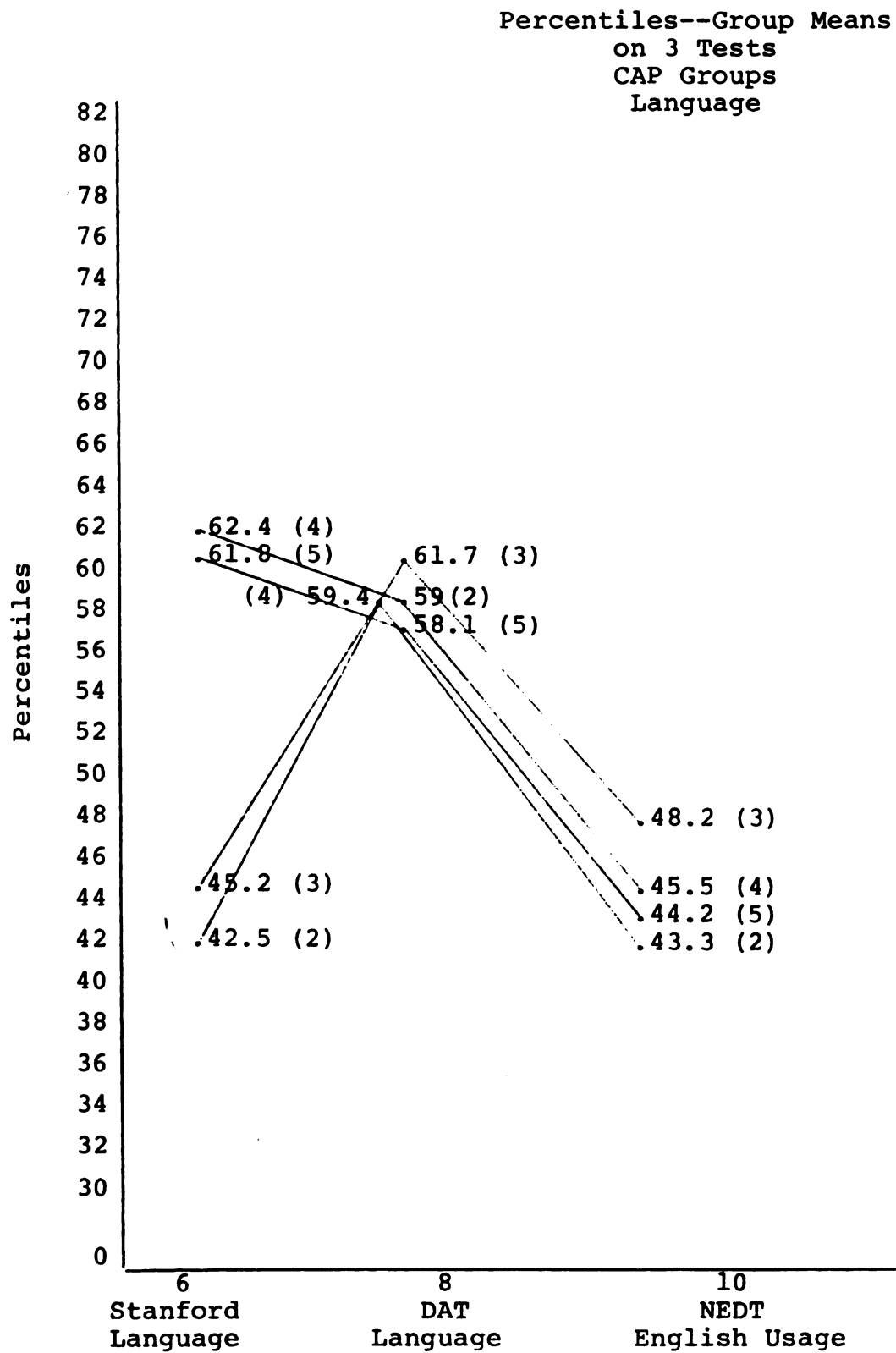
GRAPH 4:2.3



GRAPH 4:2.4



GRAPH 4:2.5



Null Hypothesis 3.1

No difference will be found in the effect of CAP on students of varying reading abilities. No aptitude treatment interaction will be evident in the regression line drawn for each CAP group.

Alternate Hypothesis 3.1

A difference will be found in the effect of CAP on students of varying reading abilities. An aptitude treatment interaction will be evident in the regression line drawn for each CAP group.

The Gates-MacGinite Reading Test, vocabulary section, administered in the ninth grade, is the dependent variable. The Gates-MacGinite Reading Test, vocabulary section, administered in the seventh grade, is the independent variable.

The regression line for each CAP group is shown in Graph 4:3.1. The linear trend is positive. The indication is that large values on the independent variable (Gates 7) are accompanied by large values on the dependent variable (gates 9). If a student received a high score on the Gates (7) vocabulary test, s/he would probably receive a high score on the Gates (9) vocabulary test. Likewise, a low score on the Gates (7) vocabulary test would probably be accompanied by a low score on the Gates (9) vocabulary test.

The null hypothesis 3.1 cannot be rejected. The trend is linear and positive.

Null Hypothesis 3.2

No difference will be found in the effect of CAP on students of varying reading abilities. No aptitude treatment interaction will be evident in the regression lines drawn for each CAP group.

Alternate Hypothesis 3.2

A difference will be found in the effect of CAP on students of varying reading abilities. An aptitude treatment interaction will be evident in the regression line drawn for each CAP group.

The Gates-MacGinite Reading Test, comprehension section, administered in the ninth grade, is the dependent variable. The Gates-MacGinite Reading Test, comprehension section, administered in the seventh grade, is the independent variable.

The regression line for each CAP group is shown in Graph 4:3.1. The linear trend is positive. The indication is that large values on the independent variable (Gates 7) are accompanied by large values on the dependent variables (Gates 9).

The null hypothesis 3.2 cannot be rejected. The trend is linear and positive.

The following chart provides the information for establishment of the regression lines.

Groups 2,3,4,5

Groups	$\bar{y}$	$\bar{x}$	Gates (9) Vocabulary Dependent	Gates (7) Vocabulary Independent	^a (y intercept)	b (slope)	r	r^2	coefficient of determination	Significance of r
2	9.0	6.7			3.7	.79	.71	.50	.50	.00001
3	9.5	7.4			2.7	.91	.76	.57	.57	.00001
4	9.3	7.2			4.1	.73	.62	.38	.38	.00001
5	8.8	7.0			3.3	.79	.70	.49	.49	.00001

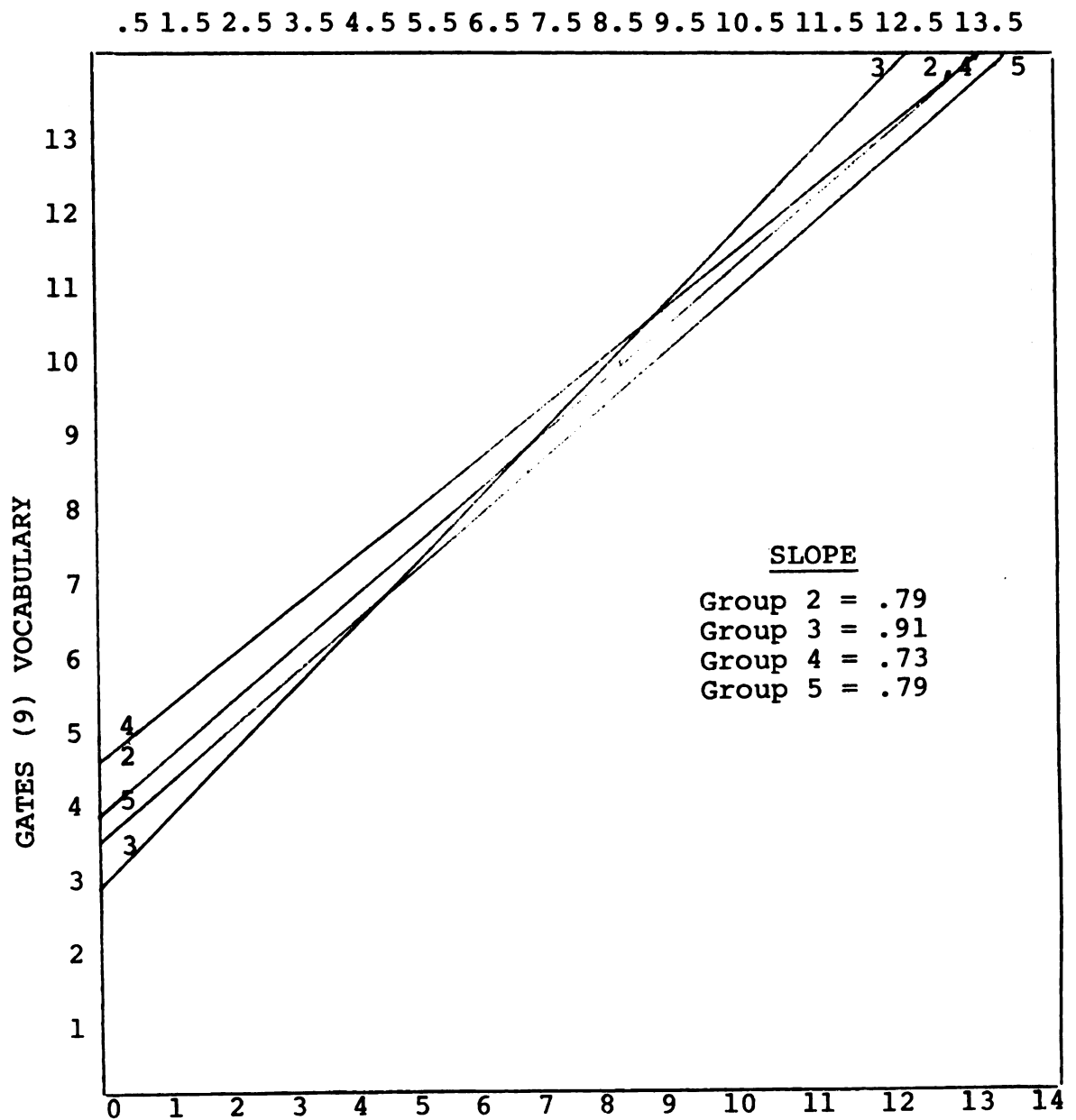
Groups	$\bar{y}$	$\bar{x}$	Gates (9) Vocabulary Dependent	Gates (7) Vocabulary Independent	^a (y intercept)	b (slope)	r	r^2	coefficient of determination	Significance of r
2	10.0	8.0			3.2	.85	.78	.60	.60	.00001
3	10.0	7.6			5.5	.67	.66	.43	.43	.00001
4	10.2	8.2			4.2	.74	.75	.57	.57	.00001
5	9.8	7.6			3.6	.83	.85	.72	.72	.00001

GRAPH 4:3.1

CAP Groups 2,3,4,5

Regression Lines
Gates (1)--Gates (7)
Vocabulary

GATES (7) VOCABULARY

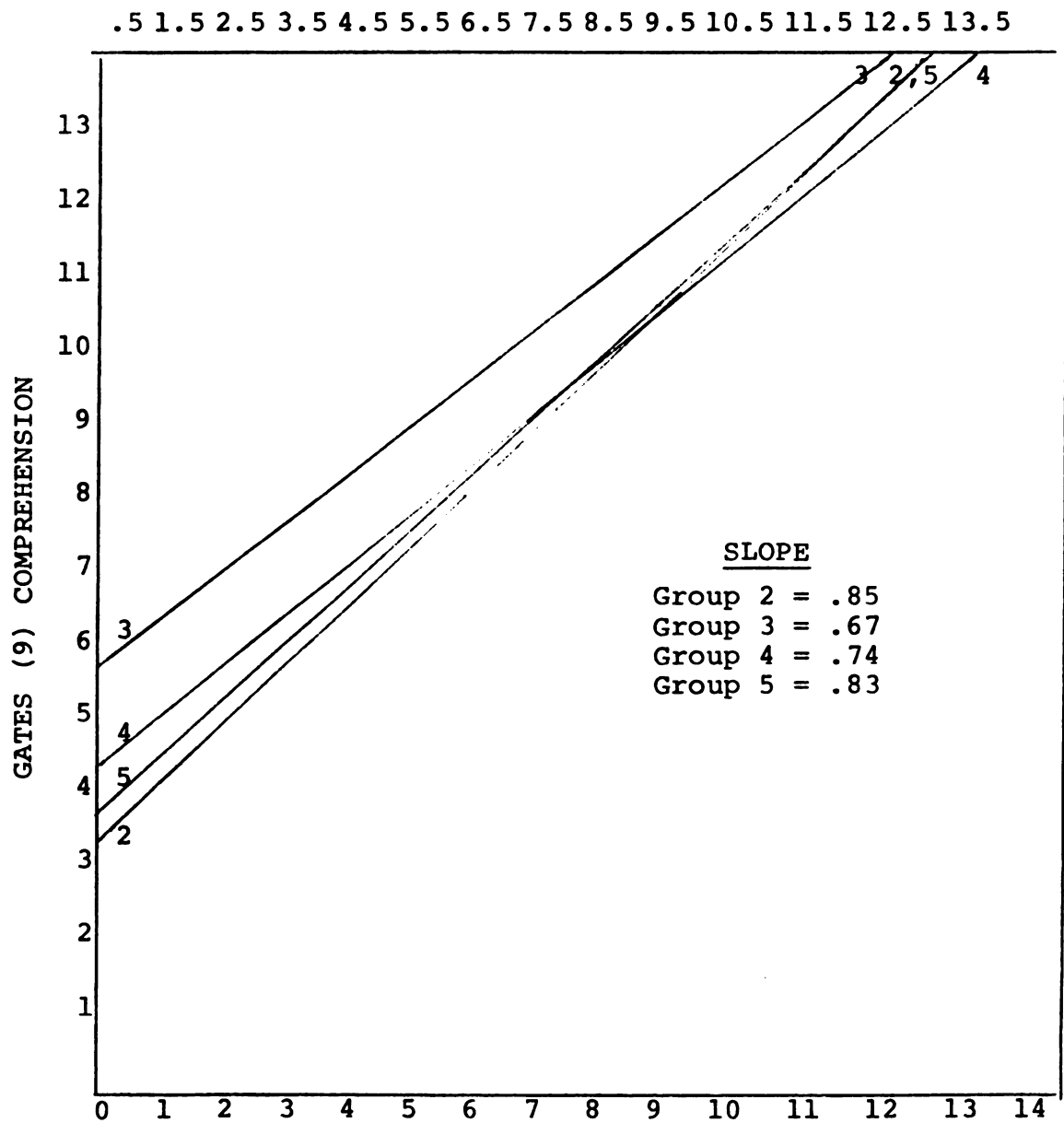


GRAPH 4:3.2

CAP Groups 2,3,4,5

Regression Lines
Gates (9)--Gates (7)
Comprehension

GATES (7) COMPREHENSION



Hypothesis	* Groups = CAP	N	t-value	t- Probability	Significant at .05	ANOVA F-VALUE	ANOVA Significance of F	Significant at .05	ANOVA F-VALUE	ANOVA Significance of F	Significant at .05	Null Hypothesis Rejected
1.1 (%)	*2 6	43 52	-.102	.919	NO							NO
1.1 (GE)	*2 6	43 52	-.052	.959	NO				.801	.373	NO	NO
1.2 (%)	*2 6	43 52	-.049	.961	NO							NO
1.2 (GE)	*2 6	43 52	.083	.934	NO				.316	.575	NO	NO
1.3 (%)	*2 6	43 52	-2.263	.024	YES				1.879	.174	NO	NO
1.4 (%)	*2 6	43 52	.130	.897	NO				1.111	.295	NO	NO
1.5 (%)	*2 6	43 52	.197	.844	NO				.570	.452	NO	NO
1.6 (%)	*2 6	43 52	-.201	.841	NO				.284	.595	NO	NO
1.7 (%)	#2 6	43 52	.249	.803	NO				.934	.336	NO	NO

Hypothesis	* Groups = CAP	N	t-value	t- Probability	Significant at .05	ANOVA F-VALUE	ANOVA Significance of F	Significant at .05	ANCOVA F-VALUE	ANCOVA Significance of F	Significant at .05	Null Hypothesis Rejected
1.8 (%)	*2,3,4,5 6	304 52	-10.240	.000	yes	.0775	.7808	no	.160	.689	no	no
1.8 (GE)	*2,3,4,5 6	304 52	-20.488	.000	yes	.2081	.6485	no	.488	.485	no	no
1.9 (%)	*2,3,4,5 6	304 52	-11.022	.000	yes	.1405	.7081	no	1.197	.275	no	no
1.9 (GE)	*2,3,4,5 6	304 52	-20.100	.000	yes	.2583	.6116	no	1.434	.232	no	no
1.10 (%)	*2,3,4,5 1,6	303 139	1.643	.101	no	1.4196	.2341	no	.3534	.061	no	no
1.11 (%)	*2,3,4,5 1,6	303 139	-2.952	.003	yes	13.8419	.0002	yes	17.041	.001	yes	yes
1.12 (%)	*2,3,4,5 1,6	303 139	-.837	.403	no	1.0577	.1793	no	1.025	.312	no	no
1.13 (%)	*2,3,4,5 1,6	304 139	-.085	.933	no	.0200	.8875	no	5.174	.023	yes	yes
1.14 (%)	*2,3,4,5 1,6	304 139	-.769	.442	no	.6203	.4314	no	3.041	.082	no	no

Hypothesis	Group * = CAP	N	ANCOVA F-VALUE	ANCOVA Significance of F	Significant at .05	Null Hypothesis Rejected
2.1 (%)	*2 *3	43 87	.507	.478	no	no
2.1 (GE)	*2 *3	43 87	.644	.424	no	no
2.2 (%)	*2 *3	43 87	1.218	.272	no	no
2.2 (GE)	*2 *3	43 87	1.824	.179	no	no
2.3 (%)	*2 *3	43 87	3.107	.080	no	no
2.4 (%)	*2 *3	43 87	.113	.738	no	no
2.5 (%)	*2 *3	43 87	.037	.847	no	no
2.6 (%)	*2 *3	43 87	.564	.454	no	no
2.7 (%)	*2 *3	43 87	.194	.660	no	no
2.8 (GE)	*2 *3	43 87	.245	.622	no	no
2.9 (GE)	*2 *3	43 87	6.289	.013	yes	yes*

Hypothesis	Group * = CAP	N	ANCOVA F-VALUE	ANCOVA Significance of F	Significant at .05	Null Hypothesis Rejected
2.10 (GE)	*3 *(4,5)	87 174	21.465	.001	yes	yes*
2.11 (GE)	*3 *(4,5)	87 174	22.149	.001	yes	yes*
2.12 (%)	*3 *(4,5)	87 174	22.149	.001	yes	yes*
2.13 (%)	*3 *(4,5)	87 174	33.766	.001	yes	yes*
2.14 (%)	*3 *(4,5)	87 174	30.88	.001	yes	yes*
2.15 (%)	*3 *(4,5)	87 174	40.109	.001	yes	yes*
2.16 (%)	*3 *(4,5)	87 174	23.129	.001	yes	yes*
2.17 (GE)	*3 *(4,5)	87 174	1.63	.687	no	no
2.18 (GE)	*3 *(4,5)	87 174	11.781	.001	yes	yes*

Summary

The purpose of this study was to do a summative evaluation of a junior high school communication arts program to determine the effect of the program on the reading and associated abilities of the students who participated in the program. Three research questions were explored in this study.

Research Question 1 and Research Findings

Is there a difference, as measured by standardized tests, between the reading and associated abilities of students who participated in the Communication Arts Pilot (CAP) program and students who did not?

The six groups were analyzed in two pairings. Group 2 and Group 6, the randomly assigned treatment and control groups, were analyzed together. Hypotheses 1.1 to 1.7 pertain to Group 2 and 6.

The second pairing was of Groups 2, 3, 4, and 5 and Groups 1 and 6, the non-equivalent treatment and control groups. Hypotheses 1.8 to 1.14 show these findings.

Chart 1 provides information for Groups 2 and 6; Hypotheses 1.1 to 1.7 and Null Hypotheses were rejected. No null hypothesis was rejected.

Chart 2 provides information about Groups 2, 3, 4, and 5 and Groups 1 and 6, Hypotheses 1.8 to 1.14, and Null Hypotheses rejected. Null Hypotheses 1.11 and 1.13 were rejected.

Research Question 2 and Research Findings

Is the effect of CAP upon different student groups consistent over time, as measured by standardized tests?

The six groups were analyzed in three pairings. Only CAP groups were analyzed. The first comparison was between Group 2 and Group 3. Group 2 was the first CAP group and had had CAP for one year. Not all students in the class had participated. Group 3 was the first CAP group with all students in the class (grade level) participating. Group 3 participated in CAP for two years. Hypotheses 2.1 to 2.9 pertain to Groups 2 and 3.

The second comparison was between Group 3 and Groups 4 and 5. All three groups had CAP for two years. Hypotheses 2.10 to 2.18 pertain to Group 3 and Groups 4 and 5.

The third comparison was between Group 4 and Group 5. Hypotheses 2.19 to 2.27 pertain to these two groups.

Chart 3 provides information for Groups 2 and 3, Hypotheses 2.1 to 2.9, and Null Hypotheses rejected. Null Hypothesis 2.9 was rejected.

Chart 4 provides information for Group 3 and Groups 4 and 5, Hypotheses 2.10 to 2.18, and Null Hypotheses rejected. All Null Hypotheses in this section except 2.17 were rejected.

Chart 5 provides information for Groups 4 and 5, Hypotheses 2.19 to 2.27, and Null Hypotheses rejected. Null Hypotheses 2.23 and 2.25 were rejected.

Research Question 3 and Research Findings

Is there a difference in the effect of CAP for students of varying abilities?

CAP Groups 2, 3, 4, and 5 were analyzed. Regression lines were shown. Scattergrams had been developed using the Gates-MacGinite Reading Test, vocabulary and comprehension sections. The tests administered in the seventh grade were used as the independent variable. The tests administered in the ninth grade were used as the dependent variable.

The vocabulary tests were dealt with in Null Hypothesis 3.1. It was not rejected. The trend was linear and positive.

The comprehension tests were dealt with in Null Hypothesis 3.2. It was not rejected. The trend was linear and positive.

Chapter V includes a summary, findings, conclusions, implications of the study for Traverse City Junior High School, recommendations, and recommendations for further study.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

Introduction

The national concern about the problem of reading ability of students had grown since Flesch's 1955 publication in Why Johnny Can't Read, to the point where Richard M. Petre wrote in the December, 1972, volume of Journal of Reading:

Never before have the mass media reported so many conflicting stories about how "well" or how "poorly" Americans read. According to these reports, Americans appear to believe that they can hardly read, that large numbers of people are illiterate, that reading failures have reached epidemic proportions, and that school systems neglect reading to such an extent that a national crisis exists (p. 200).

Programs were developed to address this national concern. This study was a summative evaluation of a junior high school communication arts program that shifted its program format from the "traditional" to one emphasizing the teaching of reading and writing skills.

The purpose of this study was to attempt to answer three questions.

1. Was there a difference between the effect of a new program, Communication Arts Pilot (CAP), and the program it replaced, as measured by selected standardized tests. Four

reading and associated ability areas were analyzed: reading comprehension, vocabulary, spelling, and language.

2. Was the effect of the CAP program the same for all groups who participated in the program, as measured by standardized tests?

3. Did the program have the same effect on students of all reading abilities?

Conclusions

The conclusions reported in this study are numerous because of the number of hypotheses analyzed. The conclusions are presented with the appropriate research questions.

Research Question 1

Is there a difference, as measured by standardized tests, between the reading and associated abilities of students who participated in CAP and students who did not?

Group 2 and Group 6 Comparisons. The students had been randomly assigned to either CAP or non-CAP. The true experimental design, experiment and control group, was used.

1. No statistically significant difference was found between the reading abilities (comprehension and vocabulary), as measured by standardized tests, of students who participated in CAP and students who did not.

Conclusion: There was no statistically significant difference, as measured by standardized tests, between the effect of CAP on the reading ability of students and the effect of the program it replaced.

2. No statistically significant difference was found between the associated abilities (spelling, language), as measured by standardized tests, of students who participated in CAP and students who did not.

Conclusion: There was no statistically significant difference found, as measured by standardized tests, between the effect of CAP on the associated abilities (spelling, language) of students and the effect of the program it replaced.

Groups 2, 3, 4, 5 and Group 1, 6 Comparison. Students had not been randomly assigned to the groups (except Groups 2 and 6 as previously discussed). The true experimental design, non-equivalent control group, was used.

1. No statistically significant difference was found between the reading abilities (vocabulary, comprehension), as measured by standardized tests, of students who participated in CAP and students who did not.

Conclusion: There was no statistically significant difference, as measured by standardized tests, between the effect of CAP on the reading abilities (vocabulary and comprehension) of students and the effect of the program it replaced.

2. No statistically significant difference was found between the spelling abilities, as measured by standardized tests, between the CAP and non-CAP students. A statistically significant difference was found between the language ability, as measured by standardized tests, of students who participated in CAP and students who did not. Two null hypotheses, 1.11 and 1.13, were rejected. Both involved language scores.

Null Hypothesis 1.11: no difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the language usage section of the Differential Aptitude Test administered in the eighth grade. The level of significance was .05.

The t-value was -2.952 with a t-probability of .003. It was statistically significant at the .05 level of significance.

The analysis of covariance F-value was 17.041 with a significance of F of .001. It was statistically significant at the .05 level of significance.

Null Hypothesis 1.13: no difference will be found between the group mean score of CAP students and the group mean score of non-CAP students on the English usage section of the National Educational Development Test administered in the tenth grade. The level of significance was .05.

The t-value was -.085 with a t-probability of .933. It was not statistically significant at the .05 level of significance.

The analysis of covariance F-value was 5.174 with a significance of F of .023. It was statistically significant at the .05 level of significance.

The t test for Hypothesis 1.11 indicated a statistically significant difference between scores at the eighth grade level on the Differential Aptitude Test. The t-test for Hypothesis 1.13 indicated that there was no

statistically significant difference at the tenth grade levels on the National Educational Development Test.

The covariant in both cases had been the language section of the Stanford Achievement Intermediate II Test administered in the sixth grade. The t-value for that test was -2.050 with a two-tailed t-probability of .041. It was not statistically significant at the .05 level of significance.

Students in the CAP program and the non-CAP program began at a point where there was no statistically significant difference between them (sixth grade). The students reached the tenth grade, after CAP or the other program, and there was no statistically significant difference between them.

Conclusion: There was no statistically significant difference found, as measured by standardized tests, between the effect of CAP on the associated abilities (spelling, language) of students and the effect of the program it replaced.

Research Question 2

Is the effect of CAP upon different student groups consistent over time, as measured by standardized tests?

Group 2 and Group 3 Comparison. Group 2 was the first group of students to participate in CAP. They had the program for one year. Group 3 was the second group to participate in CAP. They had the program for two years.

1. Statistically significant differences

a. A statistically significant difference was found between the reading abilities (comprehension), as measured by standardized tests, of students in Group 2 and Group 3. The null hypothesis 2.9 was rejected.

Null Hypothesis 2.9: no difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the comprehension section of the Gates-MacGinite Reading Test administered in the ninth grade.

The analysis of covariance F-value was 6.289 with a significance level of .013. It was statistically significant at the .05 level of significance. The covariant was the comprehension section of the Gates-MacGinite Reading Test administered in the seventh grade. The difference between the two groups is shown in Graph 4:2.2.

	<u>Gates (7) mean</u>	<u>Gates (9) mean</u>
Group 2	8.0	10.0
Group 3	7.6	10.6

Conclusion: There was a statistically significant difference found, as measured by standardized tests, between the effect of CAP on the reading abilities (comprehension) of students in Group 2 and Group 3. CAP was a more effective program in the area of reading comprehension for Group 3 than for Group 2.

b. There was no statistically significant difference found in the vocabulary reading abilities between Group 2 and Group 3.

Conclusion: There was no statistically significant difference between the effect of CAP on the vocabulary reading abilities of students in Group 2 and Group 3.

2. No statistically significant difference was found between the associated abilities (spelling, language), as measured by standardized tests, of students who were in CAP Group 2 and students in CAP Group 3.

Conclusion: There was no statistically significant difference between the effect of CAP on the spelling and language abilities of students in Group 2 and Group 3.

Summary. There was a difference in the effects of CAP for Groups 2 and 3 in one area, reading comprehension. It was more effective for Group 3. It was equally effective for both groups in the areas of vocabulary, spelling, and language.

<u>Reading Comprehension</u>	<u>Effect of CAP</u>		<u>Language Usage</u>
	<u>Vocabu- lary</u>	<u>Spelling</u>	
2 < 3	3 = 2	2 = 3	2 = 3

= similar effect
< less effect

Group 3 and Groups 4,5 Comparison. Groups 3, 4, and 5 each participated in CAP for two years. Group 3 was the second group to participate in the program. Group 4 was the third group. Group 5 was the fourth group.

There was a statistically significant difference between the reading and associated abilities, as measured by standardized tests, of the effect of CAP on the students in Group 3 and Groups 4 and 5. All Null Hypotheses from 2.10 to 2.18 were rejected except Null Hypothesis 2.17.

1. Statistically significant differences

a. A statistically significant difference was found for one of the two null hypotheses where vocabulary ability was under consideration. Null Hypothesis 2.10 was rejected.

Null Hypothesis 2.10: no difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

The analysis of covariance F-value was 21.465 with an F-significance of .001. The covariant was the Stanford Achievement Intermediate II vocabulary test. Groups 4 and 5 had been administered the 1973 edition of the test. Group 3 had been administered the 1964 edition of the test. In the 1973 edition, the vocabulary section was dictated to the students.

Null Hypothesis 2.17 also considered vocabulary abilities. It was not rejected.

Null Hypothesis 2.17: no difference will be found between the group mean score of students in one CAP group and the group mean score of students in a second CAP group on the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

The analysis of covariance F value was 1.63 with an F significance of .687. The covariant was the Gates-MacGinite Reading vocabulary section, administered in the seventh grade. All three groups had been administered the same tests. Graph 4:2.1 shows the relative position of the three groups on the two Gates tests.

There is a possibility that the different type of Stanford administered to Groups 4 and 5 might have had an effect

on the analysis of covariance when the Stanford was used as the covariant. That possibility would not occur with the Gates 7 - Gates 9 information. The Gates 7 - Gates 9 information appears more acceptable.

Conclusion: There was no statistically significant difference between the effect of CAP on the vocabulary ability of students in Group 3 and Groups 4,5.

b. Reading comprehension was considered in four of the null hypotheses, 2.11, 2.14, 2.16, and 2.18. All four were not accepted.

Hypothesis	Posttest	F-value	Sig. of F	sig. at .05	Null Rejected
2.11	Gates (9) Comprehension	22.149	.001	yes	yes
2.14	DAT (8) Verbal Reasoning	30.88	.001	yes	yes
2.16	NEDT (10) Word Usage	23.129	.001	yes	yes
2.18	Gates (9)	11.781	.001	yes	yes

The covariant used was Stanford Achievement Intermediate II Reading Comprehension for hypotheses 2.11, 2.14, 2.16. The covariant used for 2.18 was the Gates-MacGinitie Reading Test, comprehension section. The results confirm each other.

Conclusion: There was a statistically significant difference found, as measured by standardized tests, between the effect of CAP on the reading ability (comprehension) of students in Group 3 and

Groups 4,5 combined. CAP was a more effective program in the area of reading comprehension for Group 3 than for Groups 4,5.

2. A statistically significant difference was found between the associated abilities (spelling and language), as measured by standardized tests, of students in Group 3 and Groups 4,5 combined. All null hypotheses dealing with spelling and language were rejected: spelling--hypothesis 2.12, language--hypotheses 2.13 and 2.15.

a. The posttest for hypothesis 2.12 was the spelling test of the Differential Aptitude Test administered in the eighth grade. The Stanford Achievement Intermediate II spelling test was the covariant. The analysis of covariance F-value was 22.149 with an F-significance of .001. It was statistically significant at the .05 level of significance.

	<u>Mean-Covariant</u>	<u>Mean-Post</u>
Group 3	51.20	52.13
Group 4	54.09	42.95
Group 5	58.71	48.92

The format of the spelling section of the 1973 edition of the Stanford Achievement Intermediate II had not been changed. It was correlated with the earlier editions.

Conclusion: There was a statistically significant difference found, as measured by standardized tests, between the effect of CAP on the spelling ability of students in Group 3 and Groups 4,5 combined. CAP was a more effective program, in the area of spelling, for Group 3 than for Groups 4,5 combined.

b. The posttest for hypothesis 2.13 was the language usage test of the Differential Aptitude Test administered in the eighth grade. The posttest for hypothesis 2.15 was the English usage test of the National Educational Development Test administered in the tenth grade. The Stanford Achievement Intermediate II Language Test was the covariant for both analyses.

Hypothesis	Posttest	F-value	Sig. of F	Sig. at .05	Null Rejected
2.13	DAT (8) Language Usage	33.766	.001	yes	yes
2.15	NEDT (10) English Usage	40.109	.001	yes	yes

The format of the language section of the 1973 edition of the Stanford Achievement Intermediate II had not been changed. It was correlated with the earlier editions.

Conclusion: There was a statistically significant difference found, as measured by standardized tests, between the effect of CAP on the language ability of students in Group 3 and Groups 4,5 combined. CAP was a more effective program, in the area of language, for Group 3 than for Groups 4,5 combined.

Summary. There was a difference in the effectiveness of CAP for different CAP groups. CAP was more effective for Group 3 than Groups 4, 5 combined in the areas of reading comprehension, spelling, and language. It was equally effective for all three groups in the area of vocabulary.

Effect of CAP

<u>Reading Comprehension</u>	<u>Vocabulary</u>	<u>Spelling</u>	<u>Language Usage</u>
3 > 4,5	3 = 4,5	3 > 4,5	3 > 4,5

= similar effect
 > greater effect
 < less effect

Group 4 and Group 5 Comparison. Group 4 and Group 5 each participated in CAP for two years. Group 4 was the third group to participate. Group 5 was the fourth group.

1. Statistically significant differences

a. There were conflicting results as to whether there was a difference in the effect of CAP on Groups 4 and 5 on reading ability (comprehension), as measured by standardized tests.

Reading comprehension was under consideration in hypotheses 2.20, 2.23, 2.25, and 2.27. Null Hypotheses 2.23 and 2.25 were rejected; Null Hypotheses 2.20 and 2.27 were not.

The posttest in both hypotheses 2.20 and 2.27 was the Gates-MacGinite Reading Test, comprehension section, administered in the sixth grade. The covariant in hypothesis 2.20 was the Stanford Achievement Intermediate II, Reading Comprehension test, administered in the sixth grade. The covariant in hypothesis 2.27 was the Gates-MacGinite Reading Test, comprehension section, administered in the seventh grade.

Hypothesis	Posttest	F-value	Sig. of F	Sig. at .05	Null Rejected
2.20	Gates (9) Comprehen- sion	2.513	.115	no	no
2.27	Gates (9) Comprehen- sion	.144	.705	no	no

The posttest in hypothesis 2.23 was the verbal reasoning section of the Differential Aptitude Test administered in the eighth grade. The posttest in hypothesis 2.27 was the word usage section of the National Educational Development Test administered in the tenth grade. The covariant in both cases was the reading comprehension section of the Stanford Achievement Intermediate II administered in the sixth grade. The 1973 edition had been used for both Group 4 and Group 5.

Hypothesis	Posttest	F-value	Sig. of F	Sig. at .05	Null Rejected
2.23	DAT (8) Verbal Reasoning	5.541	.020	yes	yes
2.25	NEDT (10) Word Usage	5.777	.017	yes	yes

Graph 4:1.6 shows the relative positions of Groups 4 and 5 on the four reading comprehension tests. Gates (7) is not included in the graph because scores were only reported in percentiles. The path for each group is displayed.

In the sixth grade, the group means were very close. In the eighth grade (next test shown on the graph), the group means of the two groups were separated. The gap remained about the same in the ninth and tenth grades. The two groups started out very similar in the sixth grade and ended up dissimilar in the tenth grade.

Conclusion: There was a statistically significant difference found, as measured by standardized tests, between the effect of CAP on the reading comprehension ability of students in Groups 4 and 5. CAP was a more effective program, in the area of reading comprehension, for Group 4 than for Group 5.

b. No statistically significant difference was found, as measured by standardized tests, between the effect of CAP on the vocabulary ability of students in Groups 4 and 5.

Vocabulary ability was under consideration in hypotheses 2.19 and 2.26. The null hypotheses were not rejected.

The posttest in each hypothesis was the vocabulary section of the Gates-MacGinite Reading Test administered in the ninth grade.

The covariant for Hypothesis 2.19 was the vocabulary section of the Stanford Achievement Intermediate II administered in the sixth grade. The covariant for hypothesis 2.26 was the vocabulary section of the Gates-MacGinite Reading Test administered in the seventh grade.

Hypothesis	F-value	Sig. of F	Sig. at.05	Null Rejected
2.19	1.034	.311	no	no
2.26	1.934	.166	no	no

Conclusion: There was no statistically significant difference found, as measured by standardized tests, between the effect of CAP on the vocabulary ability of students in Groups 4 and 5. CAP was equally effective in the area of vocabulary for Groups 4 and 5.

2. No statistically significant difference was found, as measured by standardized tests, between the effect of CAP on the spelling and language abilities of students in Groups 4 and 5.

Spelling ability was under consideration in hypothesis 2.21. The null hypothesis was not rejected.

Language ability was under consideration in hypotheses 2.22 and 2.24. The null hypotheses were not rejected.

Hypothesis	Covariant	Posttest	F-value	Sig. of F	Sig. at .05	Null Rejected
2.21	Stanford (6) Spelling	DAT (8) Spelling	.676	.412	no	no
2.22	Stanford (6) Language	DAT (8) Language Usage	.058	.810	no	no
2.24	Stanford (6) Language	NEDT (10) English Usage	.057	.811	no	no

Conclusion: There was no statistically significant difference found, as measured by standardized tests, between the effect of CAP on the spelling and language abilities of students in Groups 4 and 5. CAP was equally effective, in the areas of spelling and language, for Groups 4 and 5.

CAP appears to have had a greater effect on Group 4 than Group 5 in the area of reading comprehension. The effect of CAP appears to be similar for Groups 4 and 5 in the areas of vocabulary, spelling, and language arts.

<u>Reading Comprehension</u>	<u>Effect of CAP</u>		<u>Language Usage</u>
	<u>Vocabulary</u>	<u>Spelling</u>	
4 > 5	4 = 5	4 = 5	4 = 5
= similar effect	> greater effect		

Summary for Research Question 2. CAP appears to have had a similar effect on all CAP groups in the area of vocabulary. Test results in the area of vocabulary ability showed no significant difference among groups.

In the area of reading comprehension, CAP appears to have been the most effective with Group 3. The effect appears to have been greater for Group 3 than for Group 2 and Groups 4,5 combined. CAP appears to have had greater effect on Group 4 than Group 5.

In the area of spelling, the effect for Groups 2 and 3 appears similar. The effect for Group 3 appears greater than for Groups 4,5. The effect for Group 4 appears greater than for Group 5.

In the area of language, the arrangement of effect of CAP on the different groups appears similar to the effects described for spelling.

Summary Table:

<u>Reading Comprehension</u>	<u>Effect of CAP</u>		<u>Language Usage</u>
	<u>Vocabulary</u>	<u>Spelling</u>	
2 < 3	2 = 3	2 = 3	2 = 3
3 > 4, 5	3 = 4, 5	3 > 4, 5	3 > 4, 5
4 > 5	4 = 5	4 = 5	4 = 5
= similar effect	> greater effect		< less effect

Research Question 3

Is there a difference in the effect of CAP for students of varying reading abilities?

Regression lines were drawn from scattergrams. The Gates-MacGinite Reading Tests administered in the seventh grade were used as the independent variable. The Gates-MacGinite Reading Tests administered in the ninth grade were used as the dependent variable. Scattergrams were created for each CAP group using vocabulary and comprehension scores. The regression lines were plotted on a graph to show the trend of all four CAP groups.

The regression lines were positive. The trend was linear.

(It appears that the CAP program did not have a differential effect for students of varying abilities. If a student received a low score on the seventh grade Gates test, s/he probably received a low score on the ninth grade Gates test. Similarly, a high score on one probably meant a high score on the other. This is the usual case if an aptitude treatment interaction is not occurring.)

Implications for Traverse City Public Schools

It appears that there was no difference between the effect of CAP on students, as measured by standardized tests and the program it replaced. Four reading and associated ability areas were analyzed--comprehension, vocabulary, spelling, and language. There was no statistically

significant difference found in any of the four areas.

This study dealt only with the effects of the program that could be measured by student performance on standardized tests. That criterion is recognized by the public as a legitimate criterion by which to judge the effectiveness of an educational program. The outcry over apparent test score declines is an example of public acceptance of standardized test scores as evaluation tools.

The CAP program appears to be as successful in the area of student test performance as the program it replaced. If the school district was satisfied with the test performance of the students in the program CAP replaced, then it should be equally satisfied with the performance of the CAP program.

If the performance of students on standardized tests was one of the reasons for replacing the traditional program, then, perhaps, the CAP program should be reviewed. There was no statistically significant difference found between student test performance of CAP and non-CAP students. In this case, CAP would not appear to have had the desired impact. It appears to be similar but not "better." If "better" was the desired outcome, then CAP appears not to have succeeded.

The question of whether CAP has had a similar impact on each of the groups that has participated in the program

was explored in Research Question 2. It appears that CAP was most effective with Group 3. There appears to have been no statistically significant difference in the effect of CAP on Groups 2 and 3 except in the area of reading comprehension. There the effect appears to have been greater for Group 3 than Group 2. In all other areas, they were similar.

The effect of CAP seems to have been consistent over all groups in one area, vocabulary. The group comparisons 2-3, 3-4,5, and 4-5 all appear to be similar.

The effectiveness of the program appears to decline with each subsequent group. The area of greatest decline appears to be in reading comprehension. Group 3 appears to be the apex with Groups 4 and 5 declining from it.

The areas of spelling and language appear to decline from Group 3 to Groups 4,5 combined. They then appear to stabilize between Group 4 and Group 5.

The most serious problem appears to be the decline in the area of reading comprehension. An investigation into why that decline is occurring would seem warranted. Perhaps the "mastery" approach to reading, the method used in the CAP program, is not appropriate in the area of improvement of reading comprehension.)

The question concerning a differential effect termed "Aptitude Treatment Interaction" of the CAP program was explored in Research Question 3. The answer appeared to be negative. If a student scored high on a standardized test

at the beginning of CAP, then s/he probably scored high on a standardized test at the end of the program. Likewise, if a student scored low on a standardized test at the beginning of CAP, then s/he probably scored low on a standardized test at the end of the program. (If CAP was viewed as a means of raising relative levels of performance of poorer students, then it does not appear to have succeeded. If CAP was viewed as a means for students to maintain a relative level of performance, then it appears to have succeeded.

Whether CAP should be continued, discontinued, or modified would depend on the goals of the school district. CAP is doing as well as the program it replaced, as measured by standardized tests. It does appear to be influencing a decline in the effectiveness of reading comprehension. CAP does not appear to be influential in changing the relative performance level of students.

The results of this study may provide some insights in one area of program effectiveness. The district could act solely on the information provided in this study or pursue further investigation of the program, as noted in "Recommendations for Further Research."

Implications for Public Schools in General

While most educators believe evaluation studies must be preplanned, this study provides ample evidence that there are data available in school districts for generating studies on an ex post facto basis. These studies do not require

more testing, but simply creative use of existing data. A commitment is needed on the part of the district to utilize the resources available.

While there is much concern about lack of "longitudinal" studies in education, this study points out that data are available in many school districts for research which covers a time period of several years. It is even apparent that in numerous school districts, there may be data which will make possible experimental and control group research designs.

Recommendations for Further Research

This study, as stated previously, evaluated only test scores in the reading and associated abilities of spelling and language. Other areas of the program which might be evaluated: (1) attitude changes in the staff since implementation of CAP, (2) attitude changes in students, (3) skills in the areas of writing and speech, (4) continuity between elementary and senior high English programs, (5) ease of implementing CAP, (6) expense of CAP, (7) staff communication, (8) staff motivation.

(An additional major area for further research would be the apparent decline from group to group in the reading comprehension abilities of the Traverse City students. Is the trend continuing or has it been reversed or stabilized? The school district continues to collect the same kinds of data

as used in this study. A continuation of this research program could be useful and is possible.

Is there a decline in reading comprehension occurring during the junior high school years, as evidenced by the data in this study? Is the pattern found in other school districts? A comparative study using non-equivalent control groups might provide useful information.

Summary

This study has been a summative evaluation of a language arts program in a northern Michigan school district. A curricular innovation has been investigated in depth to determine its impact upon student performance in basic skill areas, as measured by selected standardized tests. The school district has been provided with extensive research data on which it can base future decisions regarding termination, continuation, and/or modification of the program.

APPENDICES

APPENDIX A

TRAVERSE CITY PUBLIC SCHOOLS

FOURTH FRIDAY

COUNT INFORMATION

ENROLLMENT DATA
TRAVERSE CITY AREA PUBLIC SCHOOLS

LEVEL/ YEAR	1970	1971	1972	1973	1974	1975	1976	1977	1978
6	640	677	718	730	709	737	688	647	693
7	658	711	730	744	729	734	749	695	650
8	635	716	767	765	755	779	741	752	702
9	639	686	748	806	759	765	768	755	756
10	679	687	719	744	813	802	785	815	790
ELE	4542	4698	4760	4769	4737	4822	4936	4909	4905
JrHi	1932	2113	2245	2315	2243	2278	2258	2202	2108
SrHi	1821	1960	2054	2079	2160	2305	2248	2313	2265
SpEd	120	125	133	91	137	134	229	235	295
Total	8415	8896	9192	9254	9277	9539	9671	9659	9573

Source: Local District Summary: Fourth Friday Reports
Traverse City Area Public Schools

APPENDIX B

METHODS OF PROGRAM EVALUATION

AND

SOURCES OF EVIDENCE

Table 1. Purposes and General Methods of Program Evaluation

	Experimental Study	Quasi-Experimental Study	Correlational Study	Survey	Personnel or Client Assessment	Systematic "Expert" Judgments	Clinical or Case Study	Informal Observation or Testimony
I. To contribute to decisions about program installation								
A. Need								
1. Frequency								
a. Individual								
b. Society								
c. Other (that is, industrial, professional, governmental)								
2. Intensity								
a. Individual								
b. Society								
c. Other								
B. Program conception								
1. Appropriateness								
2. Quality								
3. Priority in the face of competing needs								
C. Estimated cost								
1. Absolute cost								
2. Cost in relation to alternative strategies oriented toward same need								
D. Operational feasibility								
1. Staff								
2. Materials								
3. Facilities								
4. Schedule								
E. Projection of demand and support								
1. Popular								
2. Political and financial								
3. Professional								



— Likely investigation method

I. To contribute to decisions about program installation

A. Need

1. Frequency

 a. Individual

 b. Society

 c. Other (that is, industrial, professional, governmental)

2. Intensity

 a. Individual

 b. Society

 c. Other

B. Program conception

1. Appropriateness

2. Quality

3. Priority in the face of competing needs

C. Estimated cost

1. Absolute cost

2. Cost in relation to alternative strategies oriented toward same need

D. Operational feasibility

1. Staff

2. Materials

3. Facilities

4. Schedule

E. Projection of demand and support

1. Popular

2. Political and financial

3. Professional

Table 1 (continued)

Experimental Study	Quasi-experimental Study	Correlational Study	Survey Study	Personnel or Client Assessment	Systematic "Expert" Judgments	Clinical or Case Study	Informal Observation or Testimony
--------------------	--------------------------	---------------------	--------------	--------------------------------	-------------------------------	------------------------	-----------------------------------

II. To contribute to decisions about program continuation, expansion, or "certification" (licensing, accreditation, and so on)

A. Continuing need

1. Frequency
 - a. Individual
 - b. Society
 - c. Other
2. Intensity
 - a. Individual
 - b. Society
 - c. Other

B. Global effectiveness in meeting need

1. Short-term
2. Long-term

C. Minimal negative side effects

D. Important positive side effects

E. Cost

1. Absolute cost
2. Cost in relation to alternative strategies to fill same need
3. Cost in relation to effectiveness

F. Demand and support

1. Popular
2. Political and financial
3. Professional

Table 1 (continued)

	Experimental study	Quasi-experimental study	Correlational study	Survey	Personnel or Client Assessment	Systematic "Expert" Judgments	Clinical or Case Study	Informal Observation or Testimony
D. Program context								
1. Administrative structure, auspices								
2. Program administration procedures								
3. Staff roles and relationships								
4. Public relations efforts								
5. Physical facilities and plant								
6. Fiscal sources and stability								
7. Fiscal administration procedures								
E. Personnel policies and practices								
1. Clients								
a. Recruitment								
b. Selection and placement, if any								
c. Evaluation, if any								
d. Discipline, if any								
e. Retention								
2. Staff								
a. Selection and placement								
b. In-service training								
c. Evaluation for promotion, guidance, retention, and so on								
3. Administrators								
a. Selection								
b. Evaluation for promotion, guidance, retention, and so on								
IV. To obtain evidence favoring program to rally support								
A. Popular								
B. Political and financial								
C. Professional								

V. To obtain evidence against program to rally opposition

- A. Popular
- B. Political and financial
- C. Professional

VI. To contribute to the understanding of basic processes

- A. Educational
- B. Psychological
- C. Physiological
- D. Social
- E. Economic
- F. Evaluation (methodology)

Table 2. Examples of Types and Sources of Evidence Frequently Associated with the Various General Methods of Investigation

Test Scores ^b	(Questionnaire or Interview Data ^c	Logs, Diaries ^d	Observations ^e	Ratings ^f	Clinical Examinations ^g	Records ^h	Social Indicators ⁱ	Expert Opinion	Hearsay, Chance Encounters
--------------------------	---	----------------------------	---------------------------	----------------------	------------------------------------	----------------------	--------------------------------	----------------	----------------------------

I. — Likely source of evidence

1. Experimental study

- A. Differences between performance of clients in the program and performance of nonclients
- B. Performance differences for clients exposed to program variations

Test Scores ^b	Questionnaire or Interview Data ^c	Logs, Diaries ^d	Observations ^e	Reviews ^f	Clinical Examinations ^g	Records ^h	Social Indicators ⁱ	Expert Opinion ^j	Hearsay, Chance Encounters ^k
--------------------------	--	----------------------------	---------------------------	----------------------	------------------------------------	----------------------	--------------------------------	-----------------------------	---

B. Report of program processes materials review or evaluation panel

C. Report of site visit to the program by a team of outside experts

VII. Clinical or case study

A. Analysis of program processes (implementation, management, evolution, and so on)

B. Phenomenological analysis of institutional change

C. Summary of impressions gained from examination of special client or personnel groups (for example, referrals)

VIII. Informal observation or testimony

A. Anecdotes about experiences of particular clients, staff members, and so on

APPENDIX C

GOALS AND OBJECTIVES OF CAP PROGRAM

TRAVERSE CITY JUNIOR HIGH COMMUNICATION ARTS
Reading Comprehension Objectives
Grades 7,8)

L
I
T
E
R
A
L

- A-1: Recognize and select the detail taken directly from the paragraph that best completes a sentence.
- A-2: Select a translation of a detail from the paragraph that best completes the sentence.
- A-3: Recognize and select the signal words relating to the paragraph pattern.
- A-4: Select the one response that best expresses the main idea of the paragraph.

I
N
T
E
R
P
E
T
I
V
E

- B-1: Complete a sentence by selecting a detail that is implied by the information given.
- B-2: Select a statement that tells the relationship among the ideas in the paragraph.
- B-3: Select the main idea that may be inferred from information in the paragraph.

A
N
A
L
Y
T
I
C
A
L

- C-1: Analyze a paragraph and select the question (main idea) that best states the problem.
- C-2: Select the hypotheses that best fit the situation.
- C-3: Distinguish between relevant and irrelevant details and select whichever is called for.

C
R
I
T
I
C
A
L

- D-1: Select the criterion that is the most appropriate basis for making a judgment.
- D-2: Select a judgment based on the criterion.

READING COMPREHENSION
LEVEL III
(Minimal: Grades 8-9)

These skills are minimal grades 7-9 and should be maintained above that grade level once mastered. If not mastered by the end of the 9th grade, the student should take courses that will allow him to master them before graduation.

Non-fiction: Journal, Essay, Biography, Autobiography,
Newspaper Article

Fiction: Short Story, Novelette, Novel
Drama

Poetry is dealt with as recreational reading and as craft-career ed.

-
- L
I
T
E
R
A
R
Y
- 1: The learner will recognize and select details taken directly from the fiction/nonfiction/drama material about characters/setting.
 - 2: The learner will recognize and select the signal words relating to the pattern of the material.
 - 3: The learner will recognize and select signals for alliteration and simile.
 - 4: The learner will select the one response that best expresses the main idea of the material.
-

- I
N
T
E
R
P
R
E
T
I
V
E
- 1: The learner will complete a sentence selecting a detail that is implied by the information in the material about characterization and setting.
 - 2: The learner will select a statement that tells the relationship among the ideas in selections of fiction/nonfiction/drama.
 - 3: The learner will identify the main idea of the plot that may be inferred from information in the fiction/nonfiction/drama material.
 - 4: The learner will identify and explain examples of figures of speech.
-

- 1: The learner will analyze fiction and nonfiction selections and choose the statement that best identifies the author's purpose (exclude theme).
- A
N
A
L
Y
T
I
C
A
L
- 2: The learner will distinguish between relevant and irrelevant details in identifying the author's purpose in nonfiction (include word usage).
- 3: The learner will be able to select the statement that best identifies the character's problem.
- 4: The learner will be able to distinguish between relevant and irrelevant details that can be used to predict the outcome of the problem.
- 5: The learner will be able to pick the statement that best identifies the solution to the problem.
-

- C
R
I
T
I
C
A
L
- 1: Given fiction/nonfiction/drama material, the learner will select the criterion that is the most appropriate basis for making a judgment.
- 2: The learner will be able to make a judgment based on external criterion.
- 3: Given two or more selections of nonfiction, the learner will identify the specified form.

TRAVERSE CITY JUNIOR HIGH SCHOOL
Communication Arts Objectives
Minimal Writing Objectives

Capitalization

1. Names of races (7i)
2. Political parties and their members (7i)
3. Organizations (7i)
4. Business firms (7i, 8)
5. Institutions and governmental departments (7i, 8)
6. Important historical events and documents (7i, 8)
7. Monuments (7i, 8)
8. First word in a sentence (7i, 8, 9)
9. First word in a quotation (7i, 8)
10. Word "I" (7i, 8)
11. Substitutes--Mother; Father (7i, 8)
12. First word and important words in title (7i, 8)
13. Specific course titles (7i, 8)
14. Section of country (7i)
15. References to religions, denominations, Bible and its parts, and Deity (7i, 8)
16. Proper nouns (7i, 8)

Mechanics

1. End marks: period, exclamation mark, question mark (7i)
2. Hyphens:
 - a. Compounds of a noun and prepositional phrase (7i)
 - b. Broken words at end of line (7i, 8)
 - c. Compound numbers (8i)
 - d. Compound fractions (9i)
 - e. Compound adjective before a noun (9i)
 - f. Capitalized adjective--last part capitalized (for example: pre-American) (9i)
3. Apostrophes:
 - a. Singular possession (7i, 8, 9)
 - b. Plural possession (7i, 8, 9)
 - c. Contractions (7i, 8, 9)
 - d. Special plurals--numbers, letters, etc. (7i, 8, 9)
4. Colons:
 - a. Time (7i)
 - b. Business letter (7i, 9)
 - c. A list (7i, 9)
 - d. In a play (9i)
5. Quotation marks:
 - a. Direct quotations (7i, 9)
 - b. Around titles, short poems, essays (7i, 9)
6. Underlining: to replace italicized writing (7i, 8)

7. Commas:
 - a. Mailing address (7i, 8, 9)
 - b. Various parts of dates (7i, 8, 9)
 - c. Salutation and complimentary close of a friendly letter (7i, 9)
 - d. Close of business letter (7i, 9)
 - e. Introductory words (well, yes, etc.) (7i, 8, 9)
 - f. Noun of address (7i, 8, 9)
 - g. Appositives (7i, 9)
 - h. Direct quotations (7i, 8, 9)
 - i. Words, phrases in a series (7i, 8, 9)
 - j. Compound sentences (8i, 9)
 - k. Parenthetical elements (9i)
 - l. Adverb clause (9i)
 - m. Participial phrase (9i)
 - n. Co-ordinate adjectives before a noun (9i)
8. Semi-colons
 - Coordinating conjunctions (9i)

Sentence Structure--words

1. Identification of parts of speech by function (7i, 8)
2. Nouns (7i, 8, 9)
3. Pronouns (7i, 8, 9)
4. Adjectives (7i, 8, 9)
5. Adverbs (7i, 8, 9)
6. Verbs (7i, 8, 9)
7. Prepositions (8i, 9)
8. Conjunctions (8i, 9)
9. Identification of parts of speech by definition (9i)
10. Interjections (9i)

Sentence Structure--sentences

1. Identification of simple sentences (7i, 8, 9)
2. Writing simple sentences (7i, 8, 9)
3. Identification of subjects (7i, 9)
4. Writing of subjects (7i, 9)
5. Identification of predicates (7i, 9)
6. Writing of predicates (7i, 9)
7. Identification of a statement (7i)
8. Identification of a question (7i)
9. Identification of a command (7i)
10. Identification of an exclamation (7i)
11. Identification of a compound sentence (8i, 9)
12. Writing of compound sentences (8i, 9)
13. Identification of complex sentences (9i)
14. Writing of adverb clause (9i)
15. Identification of noun clauses (9i)
16. Writing of noun clause (9i)

Sentence Structure--usage

1. Verbs--participial parts (8i, 9)
2. Verbs - subject agreement (8i, 9)
3. Nouns as predicate nominatives (9i)
4. Nouns as direct objects (9i)
5. Nouns as indirect objects (9i)
6. Nouns as objects of prepositions (9i)
7. Pronoun agreement with antecedent (9i)
8. Adjectives, predicate adjectives (9i)

Paragraph--Organization (readiness skill)

1. Topic sentence (7i, 8, 9)
2. Development of details (7i, 8, 9)
3. Concluding sentences (7i, 8, 9)
4. Organization by sequence, classification, comparison-contrast, cause-effect (7i, 8, 9)
5. Descriptive paragraph (8i)
6. Description of place (8i)
7. Description of person (8i)
8. Organization by definition, example, deductive, inductive (9i)
9. Expository paragraph (usually longer than one paragraph) (8i, 9)
10. Outlining (8i, 9)
11. Bibliography of expository paper (9i)
12. Narration paragraph(s), organized by sequence, classification, comparison-contrast, cause-effect, definition, example, deductive, inductive (9i))

Forms of Written Words

1. Journal (7i, 9)
2. Book report (7i, 8, 9)
3. Letter of application (9i)
4. Article--letters to editor (9i)
5. Research paragraph(s) (9i)
6. Plays (9i)
7. Dialogue--of a skit (9i)
8. Persuasive writing (9i)

CAP: A SUMMARY

CAP is a continuous communication arts pilot program from the seventh grade through the eighth grade. All students are placed at appropriate instructional levels as a result of an evaluation process. Each student learns skills in reading, writing, speaking, and listening through team teaching techniques and through the uses of multi-level materials. Students working on specific skills at the end of their seventh grade year continue on those same skills at the beginning of their eighth grade year.

The goals of the program are as follows:

To individualize instruction by:

- a. Using pretesting to place students in temporary instructional skill groups, focusing on a particular writing or reading skill within a hierarchy of objectives.
- b. Utilizing as many materials and approaches as possible to teach the skill while responding to style of learning and interests.
- c. Using post-testing to determine mastery (85%) and placement in the next skill group of the writing or reading design.

Each grade level (7th and 8th) teaching unit includes five full-time teachers, a reading resource teacher, and a clerical aide. Each teacher in the teaching unit has a daily common conference hour which is used regularly for planning teaching strategies and working on problems encountered by the unit teachers.

Each teacher in the unit follows the same teaching sequence:

- a. Pre-test
- b. Teach skill
- c. Apply skill
- d. Student - feedback
- e. Post-test
- f. In necessary - reassign to reteach

As each grade level team works more closely together, new ideas and methods are used. Each team is concentrating its skill teaching on twelve specific reading comprehension objectives and 35+ writing objectives. As these objectives are finalized, specific objectives in speaking and listening will be developed.

WHAT IS CAP?

CAP is a continuous communication arts program from the seventh grade through the eighth grade. All students are placed at the appropriate levels as a result of an evaluation process. Each student learns skills in reading, writing, speaking, and listening through team teaching techniques and through the uses of multi-level materials. Students working on specific skills at the end of their seventh grade year begin on those same skills when the eighth grade year begins.

Each grade level teaching unit has five full time communication arts teachers, one reading resource teacher, and one clerical aide. The teaching unit team carefully plans together each unit of study. Because of the unity of planning, much flexibility in scheduling is possible.

Below is a flow chart illustrating how a student might move through the CAP program.

Terms and symbols:

Skill cycle - The process that teacher and student go through for the student to learn skill.

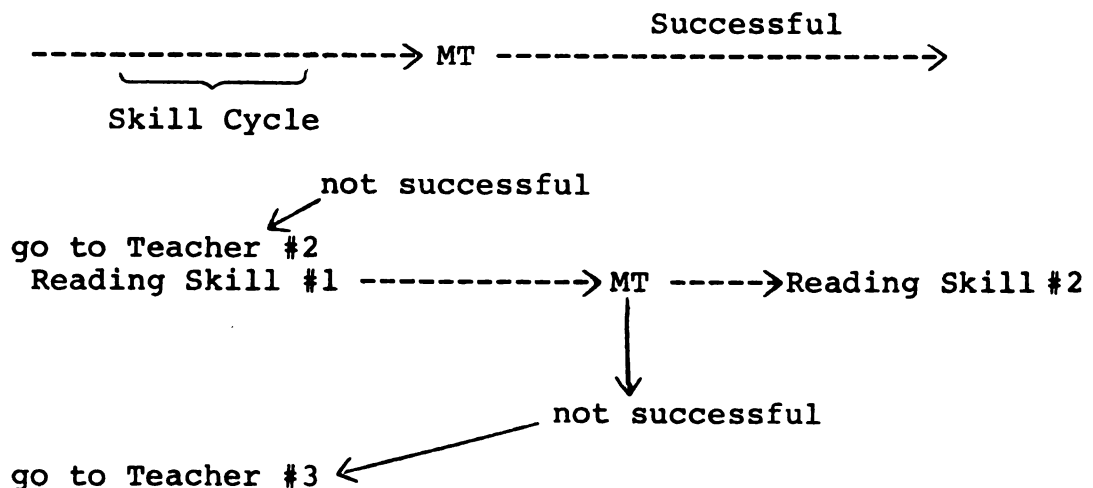
MT - Mastery Test - an examination which allows a student to proceed to another cluster of skills.

Successful - At least 85% of answers correct.

With Teacher #1

Reading Skill #1

Reading Skill #2



GOAL STATEMENTS 1977-78
8th grade Comm. Arts

A. Speaking

All students will complete one speech by the end of the school year.

B. Reading

1. During the first semester, students will go through one teach cycle of Croft B and Croft C reading skills. Those who do not pass the teach cycle will enter a minimum of one reteach cycle.
2. During the first semester, all students will be taught one cycle of setting.
3. During the second semester, students will go through one teach cycle of Croft C and Croft D reading skills. Those who do not pass the teach cycle will enter a minimum of one reteach cycle.
4. During the second semester, all students will be taught one cycle of characterization.
5. A minimum of one written book report will be required of each student by the end of the year.

C. Writing

1. The 8th grade unit will give common pretests in writing.
2. Areas covered in writing must include:
 - a. Parts of speech (nouns, verbs, adjectives, adverbs, pronouns, prepositions, conjunctions)
 - b. Sentence structure (run-ons, fragments, complete sentences; subjects, predicates, objectives)
 - c. Mechanics (commas, underlining, apostrophes, hyphens, quotation marks)

- D. All three areas of Communication Arts (reading, writing, and speaking) will be integrated whenever possible.

Progress of Program

The CAP program began in 1972-73. A team approach was used which included the following:

310 students (7th grade only)
 2 teachers
 1 Reading Resource Teacher
 (The 3 teachers involved had a common planning period.)

* In 1973-74

630 students (7th grade only)
 4 1/5 teachers (common conference hour)
 1 Reading Resource Teacher
 1 Aide

* In 1974-75

760 seventh graders	760 eighth graders
5 teachers (common conference hour)	5 teachers (common conference hour)
1 Reading Resource Teacher	1 Reading Resource Teacher
1 Aide	1 Aide

* In 1975-76. The Ninth Grade Project was established as a transition between the 7th - 8th grade pilot program and APEX. The structure during this year includes all junior high students enrolled in communication arts. The structure is as follows:

7th grade Teaching Unit (with common conference hour)
 1 unit leader
 1 Remedial Reading Teacher

8th grade Teaching Unit (with common conference hour)
 1 unit leader
 1 Reading Consultant

9th grade Teaching Unit (with common conference hour)
 1 unit leader
 1 Reading Consultant

2 aides assigned to Junior High Communication Arts Department to work with all grade levels.

* 4 1/2-days of preservice in 1973-74 (7th grade), 1974-75 (8th grade), 1975-76 (9th grade).

* 10 1/2-days were provided (to the communication arts department) each year for inservice.

Example of a Skill Cycle

OBJECTIVE: By target date teachers will have completed one complete cycle of:

- a. Teach
- b. Practice
- c. Feedback (student)
- d. Post-test

Steps in Planning

1. Pre-planning (includes inservice time) of unit committees
2. Set target date for completion
3. Skill group
4. Match up skill group with teacher strengths where possible
5. Work out management system within teaching unit
6. Begin cycle
7. Evaluation during skill cycle by teaching unit during unit meetings
8. Post-test
9. Evaluation by teaching unit of skill cycle

UPDATE

I. A. Goal Area: Continuity of Skills

There is a sequential list of performance skills* in reading and writing upon which Junior High Communication Arts instruction is based. Students progress through these skills at the rate most appropriate for them but within constraints of grouping management. The skills have developmental clusters which are identified for each grade level. These clusters represent the point at which most students receive initial instruction in these skills.

(*The skills list is a staff-developed adaptation, drawn from the K-12 Communication Arts Steering Committee list. The reading skills are based on the Croft Communication Reading Skills Design.)

B. Needs Assessment

--Finalize minimal list of objectives with grade level clusters.

--Solidify commitment to and understanding of common language; i.e., teach model, grammar, reading.

--Further development of instructional techniques and suggested materials for special groupings in certain areas (skill group isolates, enrichment groups, management needs).

--Clarification of Teach Model Components

C. Resources Necessary

--A three year commitment for released time for Inservice and Program development meetings on various topics (a minimum of six half days).

II. A. Goal Area: Developmental Groupings

The unit goal is to place students in developmental groupings in reading based on Standardized Achievement Testing (Gates or other standardized tests) and skill cluster groupings in writing based on departmental skills list and criterion testing.

Students are placed at their instructional level based on:

1. Reading--standardized achievement test scores or developmental levels (within a range of three developmental groups)
2. Writing--are placed in skills groups based on criterion testing.

B. Needs Assessment

- Continue testing program
- Work on continuous improvement of management system

C. Resources Necessary

- Inservice time
- Continuing money for testing correction
- One full day session per semester per unit to continuing level skills grouping

III.A. Goal Area:

Units will operate with a shared decision-making approach. Processes for arriving at decisions will be specified (majority; consensus; etc.) and unit members will be held accountable* for their goals and decisions.

- *
 1. Unit establishes goals through shared decision making.
 2. Unit goals approved by department head and principal.
 3. Unit goals approved by director of instruction.

Monitoring of implementation of goals is most effective first at the unit peer teacher level, but if problems cannot be resolved at that level, they will be handled through appropriate administrative channels.

B. Needs Assessment

- Continuing development of teaming skills necessary on the decision-making model.
- Continuing development of leadership skills for Unit Leaders.

C. Resources Necessary

- Four Inservice Half Days.

IV. A. Goal Area:

Individual student progress in specific reading and writing skills will be reflected in student record cards.

B. Needs Assessment

--Refinement of cards and system to suit needs for grouping and clarify what mastery "punch" indicates.

C. Resources Necessary

--Continuation of aide service to maintain records and grouping.

APPENDIX D

DEMOGRAPHIC INFORMATION

TABLE 34
Estimated Population
 Grand Traverse County

<u>1960</u>	<u>1970</u>	<u>Estimated*</u> <u>1979</u>	<u>Projected</u> <u>1985</u>
33,490	39,175	53,468	63,055
(Census)	(Census)		
% Change	17.0%	36.5%	17.9%
	(1960-70)	(1970-79)	(1979-85)
<u>Townships</u>	<u>1970</u>	<u>Estimated**</u> <u>1979</u>	<u>% Change</u> <u>1970-1979</u>
Acme*	1,662	2,612	57.2%
Blair*	1,677	4,590	174.0%
East Bay*	3,356	5,726	70.6%
Fife Lake	638	894	40.1%
Fife Lake Village#	(274)	(304)	(10.9%)
Garfield*	4,917	8,738	77.7%
Grant	507	717	41.4%
Green Lake*	1,206	2,110	74.9%
Long Lake*	1,584	3,111	96.4%
Mayfield	651	929	42.7%
Paradise*	1,434	2,544	77.4%
Kingsley Village	(632)	(1,127)	(78.3%)
Peninsula*	2,642	3,330	26.0%
Traverse City*	18,048	16,710	-7.4%
Union	57	101	77.1%
White Water	796	1,356	70.3%
Grand Traverse County and Elmwood Twp.	41,415	56,766	37.1%

* County estimates and designated townships are based on average of three factors: electric meters, voter registration and school age children.

** Townships not designated by * are based on average of two factors: voter registration and electric meter.

Township population includes village population as shown in ().

TABLE 36
Estimated Population By Age Group
 Grand Traverse County
 Spring 1979

<u>Age Group</u>	<u>1970 Census</u>	<u>% of Total</u>	<u>Spring 1979</u>	<u>% of Total</u>	<u>% Change</u>
Under 5	3,297	8.4%	3,900	7.3%	18.3%
5 - 9	3,796	9.7%	4,178	7.8%	10.1%
10 - 14	4,130	10.5%	4,438	8.3%	7.4%
15 - 19	4,050	10.3%	6,611	12.4%	6.3%
20 - 24	2,671	6.8%	5,238	9.8%	96.1%
25 - 29	2,420	6.2%	4,465	8.4%	84.5%
30 - 34	2,142	5.5%	3,322	6.2%	55.1%
35 - 39	2,129	5.4%	3,218	6.0%	51.2%
40 - 44	2,183	5.6%	2,968	5.6%	36.0%
45 - 49	2,296	5.9%	3,000	5.6%	30.7%
50 - 54	1,967	5.0%	2,255	4.2%	14.6%
55 - 59	2,042	5.2%	2,422	4.5%	18.6%
60 - 64	1,798	4.6%	2,060	3.9%	14.6%
65 - 69	1,410	3.6%	2,009	3.8%	42.5%
70 - 74	1,044	2.7%	1,299	2.4%	24.4%
75 - 79	834	2.1%	935	1.7%	12.1%
80 +	<u>966</u>	<u>2.5%</u>	<u>1,150</u>	<u>2.1%</u>	<u>19.0%</u>
TOTAL	39,175	100.0%	53,468	100.0%	36.5%

AGE DISTRIBUTION 1970 AND 1979

GRAND TRAVERSE CO. 1970

1979

% TOTAL

48.2

47.5

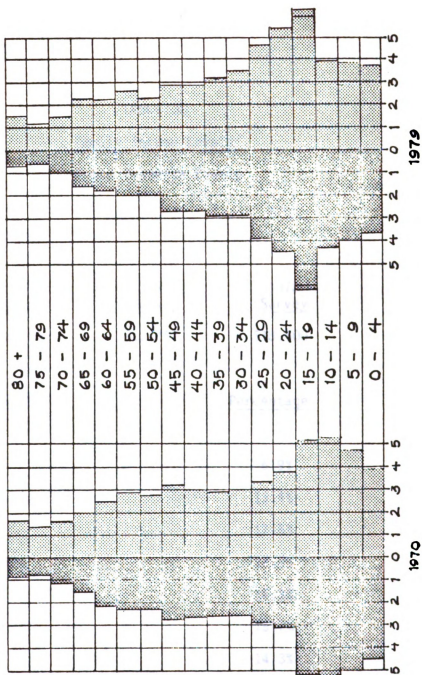
51.8

52.5

MALE

FEMALE

AGE



DATA RESEARCH CENTER, INC.

FIGURE 0

TABLE 38
Family Income Levels
 Grand Traverse County

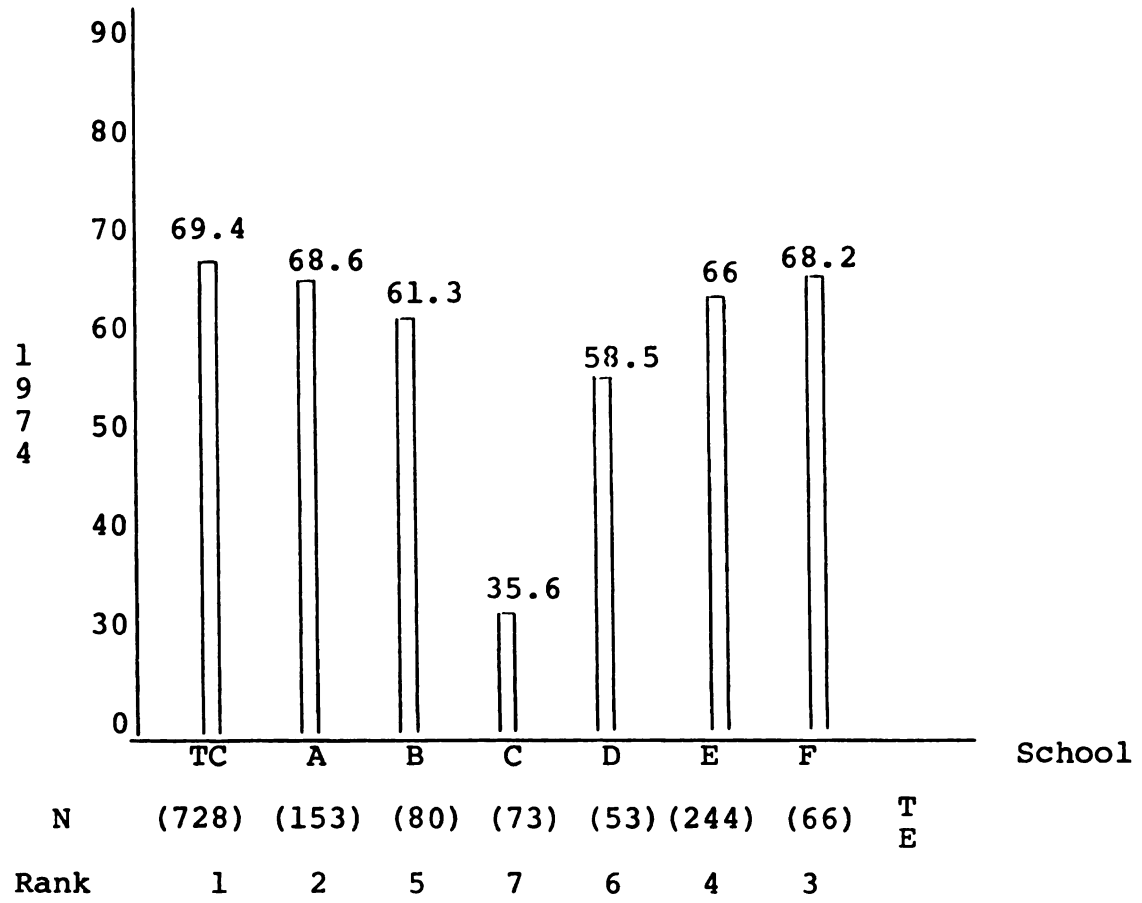
Median Family Income:

<u>1970 Census</u>	<u>1975 Survey</u>	<u>1978 Survey</u>
\$10,459	\$12,167	\$18,500

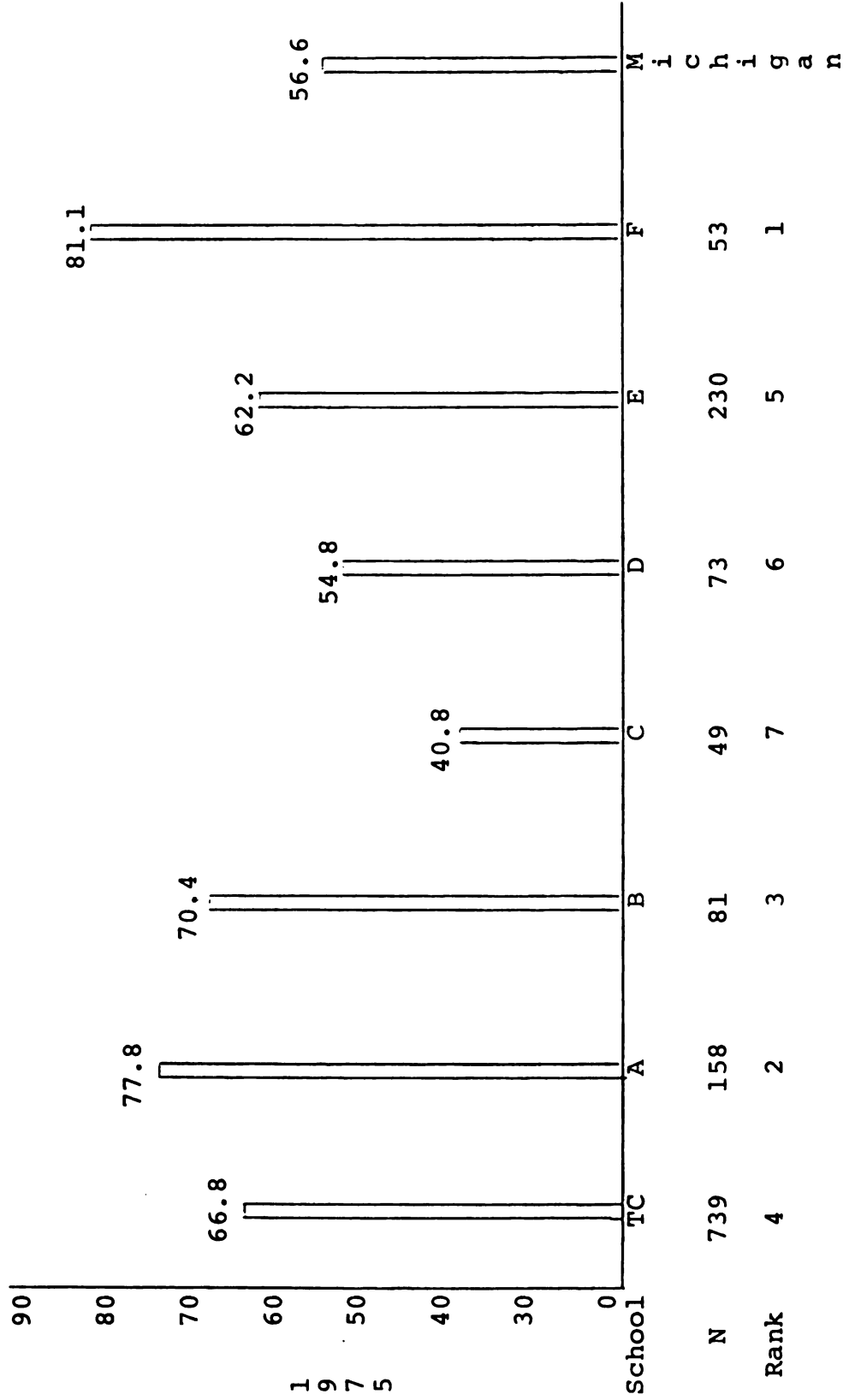
- Family Income, 1978:	<u>Percentage</u>
0,000 - 4,999	6.9%
5,000 - 9,999	13.4%
10,000 - 14,999	17.5%
15,000 - 19,999	17.9%
20,000 - 24,999	14.3%
25,000 - 29,999	15.7%
30,000 +	14.3%
 TOTAL	 100.0%

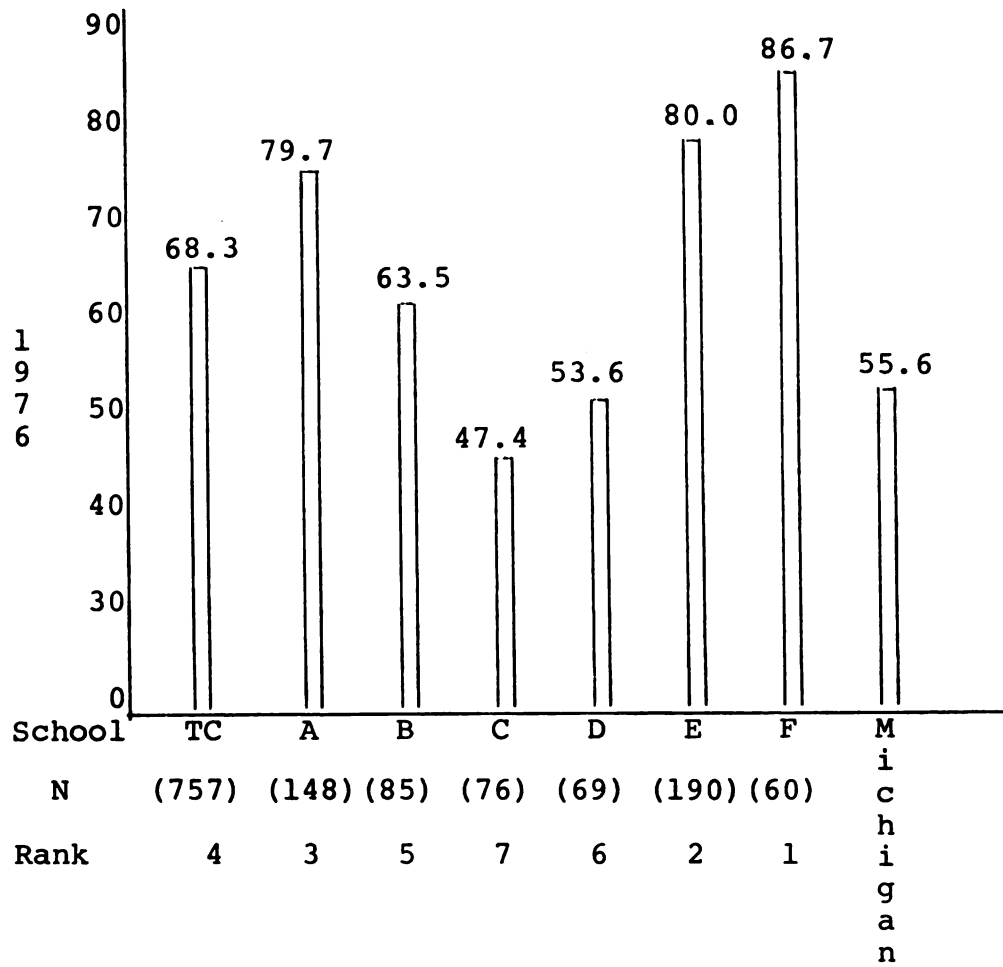
APPENDIX E

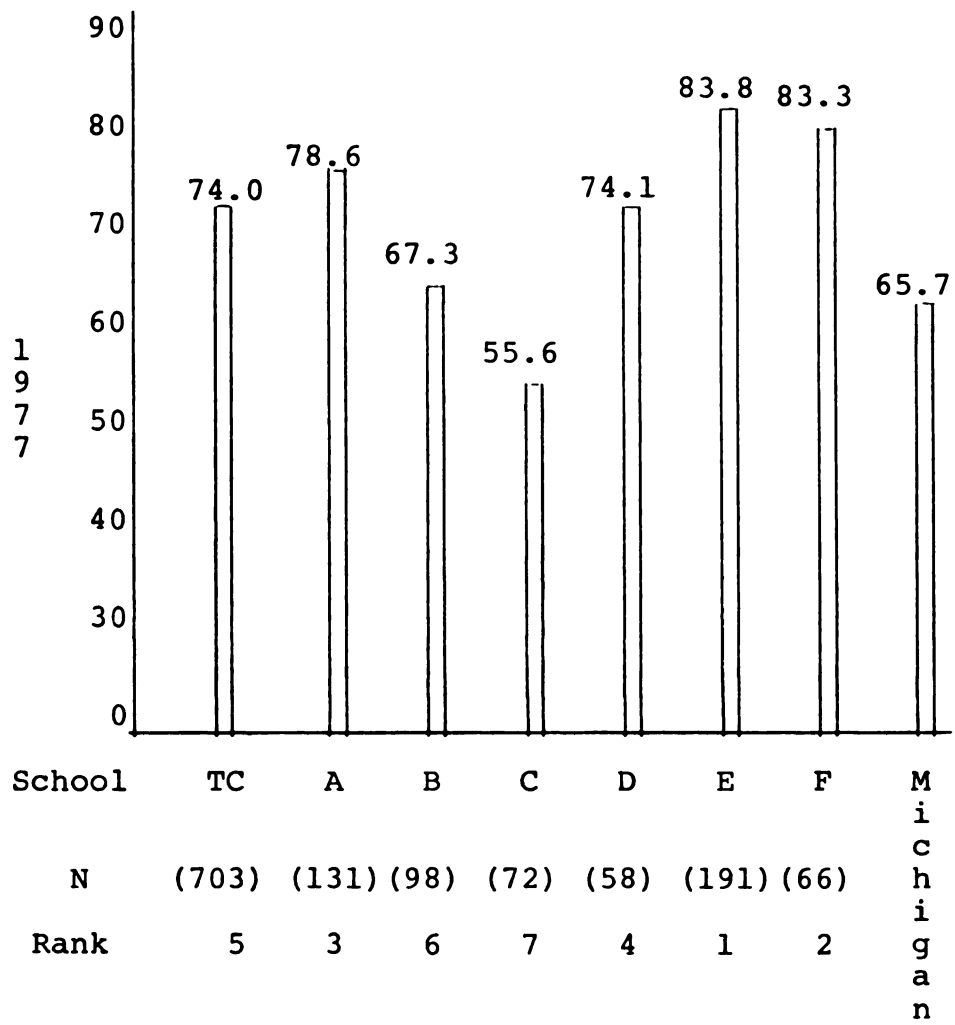
MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM
READING - SEVENTH GRADE
PROPORTION ATTAINING 75-100% OF OBJECTIVES
FOR SEVEN SCHOOL DISTRICTS IN
NORTHWEST MICHIGAN

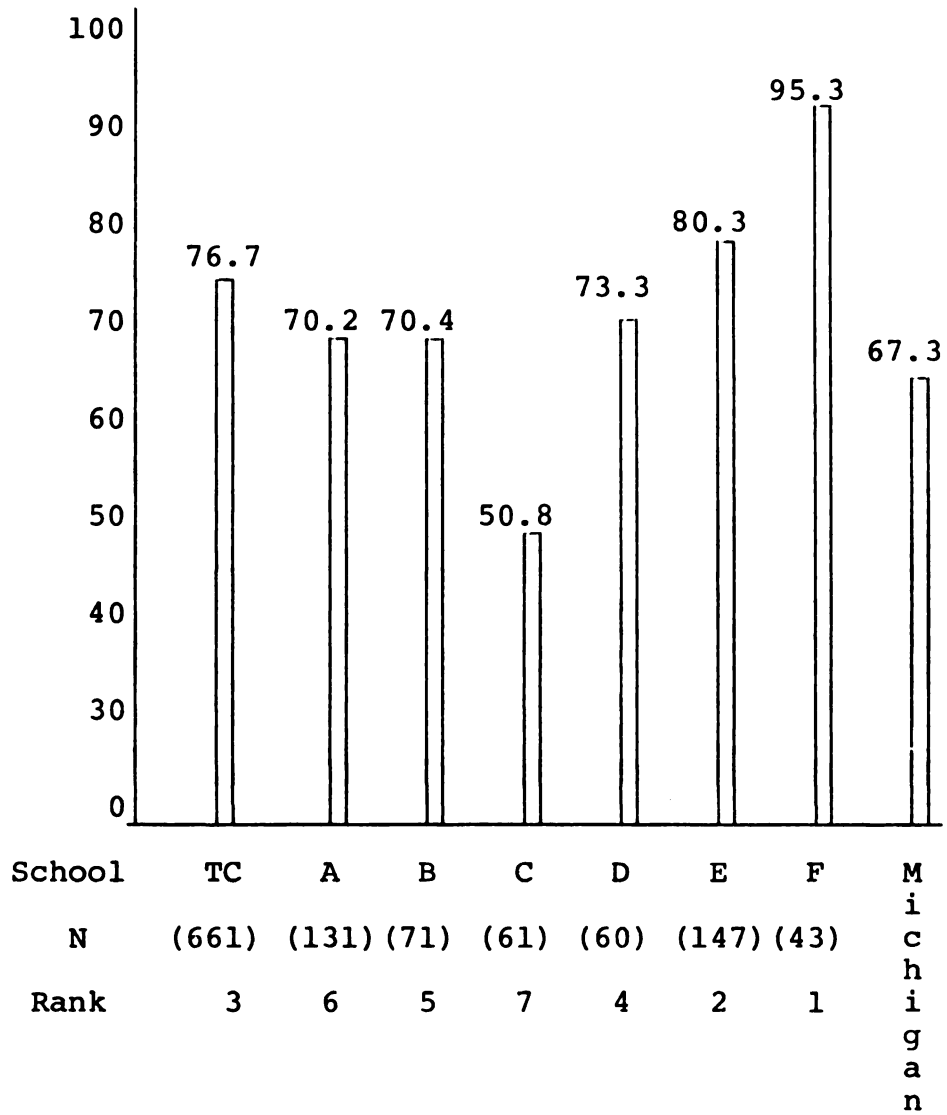


MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM
Reading - Seventh Grade
Proportion Attaining 75-100% of Objectives









BIBLIOGRAPHY

BIBLIOGRAPHY

- Alkin, M. C. , & Fitz-Gibbon, C. T. Methods and theories of evaluating programs. Journal of Research and Development in Education, 8 (3), 1975, 2-15.
- Anderson, S. B., & Ball, S. The profession and practice of program evaluation. San Francisco: Jossey-Bass Publishers, 1978.
- Barone, T. E. Reading, writing, and mastery learning: are they compatible? Educational Leadership, December, 1978, 187-191.
- Block, J. H., & Burns, R. B. Mastery learning. Review of research in education, Vol. 4. L. S. Shulman (Ed.) Itasca, Illinois: F. E. Peacock, 1976.
- Block, J. H. (Ed.). Mastery learning: theory and practice. New York: Holt, Rinehart, and Winston, 1971.
- Bloom, B. S. Human characteristics and school learning. Chicago: McGraw-Hill Book Company, 1976.
- Bloom, B. S. Mastery learning and its implications for curriculum development. Confronting curriculum reform. E. W. Eisner (Ed.). Boston: Little, Brown, and Company, 1971, 17-49.
- Borg, W. R., & Gall, M. D. Educational research. New York: David McKay Company, Inc., 1971.
- Buros, O. K. Sixth mental measurement yearbook, 1965. Highland Park, New Jersey: Gryphon Press, 1965.
- Buros, O. I. Seventh mental measurement yearbook, 1972. Highland Park, New Jersey: Gryphon Press, 1972.
- Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally College Publishing Company, 1963.
- Clymer, T., Barrett, T. C., & Burmeister, L. E. More sounds and syllables: the Ginn enrichment program. Ginn and Company, 1969.

- Combs, A. W. Forward in F. T. Wilhelms (Ed.), Evaluation as feedback and guide. Washington, D. C.: Association for Supervision and Curriculum Development, 1967.
- Cronbach, L. J. Comments on "mastery learning" and its implications for curriculum development. Confronting curriculum reform, E. W. Eisner (Ed.). Boston: Little, Brown, and Company, 1971, 49-55.
- Cronbach, L. J. Course improvement through evaluation. Teachers College Record, 64 (8), May, 1963.
- Cronbach, L. J., & Suppes, P. Research for tomorrow's schools: disciplined inquiry for education. New York: MacMillan, 1969.
- Declines in standardized test scores among secondary school students. ERS information aid. Arlington, Virginia: Educational Research Service, Inc., 1976.
- Discovering treasure: teachers' edition. Glenview, Illinois: Scott, Foresman, and Company, 1973.
- Dunning, S. The public and English teachers: an adversary relationship. English Journal, 64 (6), September, 1975, 9.
- Eisner, E. W. (Ed.). Confronting curriculum reform. Boston: Little, Brown, and Company, 1971.
- Farr, R., & Rosner, N. L. Reading assessment: a look at problems and issues. Journal of Reading, 17, May, 1974, 592-599.
- Fitz-Gibbon, C. T., & Morris, L. L. How to design a program evaluation. Beverly Hills, California: Sage Publications, Inc., 1978.
- Flesch, R. Why Johnny can't read. New York: Harper & Row, 1955.
- Good, C. V. (Ed.). Dictionary of education. New York: McGraw-Hill Book Company, 1973.
- Goodlad, J. I. The changing school curriculum. New York: The Georgian Press, 1966.
- Guilford, J. P. Fundamental statistics in psychology and education. New York: McGraw Book Company, 1965.
- Gussin, P. Developing the ninth grade objective--referenced test in communication skills: the Detroit experience.

- Detroit, Michigan: Detroit Public Schools, Spring, 1979.
- Harris, C. W. (Ed.). Problems in measuring change. Madison, Wisconsin: University of Wisconsin Press, 1963.
- Herzon, F. D., & Hooper, M. Introduction to statistics for the social sciences. New York: Thomas Y. Crowell Company, Inc., 1976.
- Holt, J. How children fail. New York: Dell Publishing Company, Inc., 1964.
- Isaac, S., Michael, W. B. Handbook in research and evaluation. San Diego, California: EDITS Publishers, 1971.
- Jennings, W., & Nathan, J. Startling/disturbing research on school program effectiveness. Phi Delta Kappan, 58 (7), March, 1977, 568-572.
- Kerlinger, F. N. Foundations of behavioral research. New York: Holt, Rinehart, and Winston, Inc., 1973.
- Klugh, H. E. Statistics: the essentials for research. New York: John Wiley & Sons, Inc., 1974.
- Kohn, H. 36 children. New York: The New American Library, Inc., 1967.
- Kozal, J. Death at an early age. Boston, Massachusetts: Houghton Mifflin Company, 1967.
- Leabo, D. A. Basic statistics. Homewood, Illinois: Richard D. Irwin, Inc., 1976.
- Local district summary: fourth Friday report. Traverse City, Michigan: Michigan Department of Education. Report of Traverse City Public Schools, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978.
- McNeil, J. D. Curriculum, a comprehensive introduction. Boston: Little, Brown, and Company, Inc., 1977.
- Moorish, I. Aspects of educational change. New York: John Wiley & Sons, Inc., 1976.
- Needham, P. Curriculum change: a study of the Traverse City Public School Communication Arts Pilot program, grades 7-8. Unpublished master's thesis, Central Michigan University, June, 1974.

- Petre, R. M. The '70's: decade of achievement? Journal of Reading, 16, December, 1972, 200-206.
- Population characteristics: ten counties, June, 1979. Developed by N. C. Hayward, Director. Traverse City, Michigan: Data Research Center, Inc., 1979.
- Sarason, S. B. The culture of the schools and the problems of change. Boston: Allyn & Bacon, Inc., 1971.
- Scriven, M. The methodology of evaluation. Educational evaluation: theory and practice. B. Worthen & J. R. Sanders (Eds.). Reprint from AERA monograph on curriculum evaluation, no. 1. Chicago: Rand McNally and Company, 1968.
- Stake, R. The countenance of educational evaluation. Teachers College Record, 68 (7), April, 1967.
- Strang, R., McCullough, C. M., & Traxler, A. E. The improvement of reading. New York: McGraw-Hill Book Company, 1967, 131-144.
- Taba, H. Curriculum development: theory and practice. New York: Harcourt, Brace, and World, 1962.
- Tuinman, J., Rowls, M., & Farr, R. Reading achievement in the United States: then and now. Journal of Reading, 19, March, 1976, 455-463.
- Tyler, R. Basic principles of curriculum and instruction. Chicago: University of Chicago Press, 1974.
- Van Til, W. (Ed.). Curriculum: quest for relevance. Boston: Houghton-Mifflin Company, 1974.
- Walker, H. M., & Lev, J. Elementary statistical methods. New York: Holt, Rinehart, and Winston, Inc., 1969.
- Wilhelms, F. T. Evaluation as feedback and guide. ASCD yearbook. Washington, D. C.: Association of Supervision and Curriculum Development, 1967.
- Williams, R., Bauer, M., Eddington, M., Needham, P., & Panter, J. A. Update on CAP program. Traverse City, Michigan: Traverse City Public Schools, May 5, 1978.
- Worthen, B. S., & Sanders, J. R. Educational evaluation: theory and practice. Belmont, California: Wadsworth Publishing Company, Inc., 1973.
- Zintz, M. V. The reading process. Dubuque, Iowa: William C. Brown Company Publishers, 1970.

Telephone and Personal Interviews

- Ailsworth, C., chair of science department, Traverse City Senior High School, in person, May 24, 1979.
- Albright, P., ninth grade teacher, Traverse City Junior High School, by telephone, February 25, 1979.
- Bacon, R., chair of social studies department, Traverse City Junior High School, in person, June 6, 1979.
- Crampton, S., parent. Traverse City, by telephone, March 15, 1979; April 1, 1979; April 16, 1979.
- Dodd, L., chair of social studies, Traverse City Senior High School, in person, May 31, 1979.
- Drury, C., counselor, Charlevoix Middle School, Charlevoix, Michigan, in person, May 9, 1979, and May 31, 1979.
- Forton, S., reading teacher, eighth grade, Traverse City Junior High School, by telephone, February 25, 1979, and March 15, 1979; and in person, March 26, 1979.
- Johnston, R. E. acting director of personnel and instruction, Traverse City Public Schools, by telephone, December, 1978, and March 15, 1979; in person, March 23, 1979, and April 11, 1979.
- Keating, W., assistant principal, Traverse City Senior High School, by telephone, March 29, 1979; in person, April 11, 1979.
- Lemcool, W., acting chair of the reading department, Traverse City Public Schools, by telephone, March 16, 1979.
- MacInnis, M., director of counseling, Traverse City Senior High School, by telephone, March 30, 1979.
- Moore, E., reading teacher, Traverse City Junior High School, in person, May 30, 1979.
- Needham, P., department chair of communication arts, Traverse City Junior High School, by telephone, March 16, 1979; in person, February 26, 1979, and April 2, 1979.
- Search, C., reading teacher, Traverse City Senior High School, by telephone, February 25, 1979.
- Thibedeau, J., chair of math department, Traverse City Senior High School, in person, May 31, 1979.

Van Wierner, G., superintendent, Charlevoix Public Schools,
Charlevoix, Michigan, in person, May 9, 1979.

Williams, R., reading teacher, ninth grade, Traverse City
Junior High School, March 16, 1979.