



RETURNING MATERIALS:

Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

JUL 10 1978

A STUDY OF THE EFFECTS OF EARLY RETENTION
ON FIFTH GRADE ACHIEVEMENT

By

Arthur Frederick Ebey

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

1981

6115748

ABSTRACT

A STUDY OF THE EFFECTS OF EARLY RETENTION ON FIFTH GRADE ACHIEVEMENT

By

Arthur Frederick Ebey

Purpose of the Study

The purpose of this study was to compare the achievement levels of low achieving fifth grade students who had experienced retention with low achieving fifth grade students who had not experienced retention. The academic areas of reading and mathematics were used in the comparison.

Design and Analysis Procedure

Forty-nine fifth grade students who had experienced one retention in kindergarten, first, second or third grade were selected. Their reading scores on the third grade Metropolitan Achievement Test were identified and formed a range of scores. A comparison group of 200 fifth grade students who had not been retained but who scored within the same range on the third grade reading test, was identified.

The two groups were then compared at the fifth grade level using the Metropolitan Achievement Test

sub-tests: Word Analysis, Total Reading and Total Mathematics.

A Multivariate Analysis of Covariance statistical program was used to compare the two groups and test the three main hypotheses and two sub-hypotheses.

Results

The testing of three hypotheses which attempted to verify differences of achievement levels between the two groups in the areas of reading vocabulary, total reading and total mathematics, confirmed the following statements:

1. Low achieving retained students cannot be expected to attain scores that are comparable with low achieving passed students on the Total Reading Sub-test of the Metropolitan Test.
2. Low achieving retained students may be expected to achieve at comparable levels with low achieving passed students on the Mathematics Sub-test of the Metropolitan Test.

The two sub-hypotheses which attempted to verify differences in functioning on the achievement test of males and females confirmed the following statements:

1. Passed and retained males and females may be expected to achieve similarly on the fifth grade subtests, Word Analysis, Total Reading and Total Mathematics of the Metropolitan Achievement Test.
2. The performance of passed and retained males and females may be expected to be consistent on the three sub-tests, Word Analysis, Total Reading and Total Mathematics of the fifth grade Metropolitan Achievement Test.

Copyright by

Arthur Frederick Ebey

1981

Dedicated to my wife Carolyn,
my children, Lance and Tara,
and my parents
Mr. and Mrs. Glenn Ebey

ACKNOWLEDGMENTS

I am deeply indebted to many people for their help with this dissertation. My greatest gratitude is due Dr. Ben Bohnhorst, my committee chairman. His constant encouragement, support and unwavering trust in me guided me through the research from the formative stages and to its completion.

I would like to express deepest appreciation to other members of my committee, Dr. Glen Cooper, Dr. Charles Blackman and Dr. John H. Suehr, whose suggestions and support were very helpful in this effort.

A special note of thanks to my patient family who worked quietly beside me, giving me encouragement and, at times, little shoves so that we may all see this effort completed.

My thanks are also rendered to the Forest Hills School District, Mr. Philip Schroo, Superintendent, and Mr. Kenneth Walnutz, Assistant Superintendent, for giving me needed time to work on this effort.

TABLE OF CONTENTS

List of Tables.....	vii
---------------------	-----

List of Appendices.....	viii
-------------------------	------

Chapter

I. Introduction.....	1
Statement of the Problem.....	5
Definition of Terms.....	6
Contribution of the Study.....	7
Assumptions on Which the Study is Based	8
Limitations of the Study.....	9
Design of the Study.....	10
Instrumentation.....	12
Sample Selection Procedure.....	12
Analysis of Outcomes.....	13
Organization of the Dissertation.....	13
II. Review of the Literature.....	15
Studies Concerned with Affective and Social Progress of Retained and Promoted Students.....	17
Studies Concerned with Retention and the Academic Achievement of Students.....	26
Some Aspects of the Retention-Promotion Question as It Relates to Achieve- ment.....	34
Summary and Conclusions.....	38
III. Methodology.....	39
Sample Population.....	39
Procedures for Collection of Data.....	41
Criteria for Selection of Sample.....	42
Procedure for Selection of Sample Groups.....	43
Instrumentation.....	44
Item Selection Statistical Information and Reliability of the MAT.....	47
Statement of Hypotheses.....	48
Program of Analysis.....	51

IV. Data Analysis.....	55
Data on Means and Correlations.....	55
The Influence of Sex Differences on Outcomes.....	57
Organization of Data Analysis as It Relates to the Hypotheses.....	59
Analysis of the Grand Means.....	60
Analysis by Dependent Variable, Uni- variate F Ratio--Method I.....	61
Analysis by Dependent Variable, Uni- variate Step Down F ratio.....	62
Testing the Hypotheses.....	63
Summary.....	65
V. Summary, Conclusions and Recommendations...	67
Grand Mean Difference.....	69
Influence of Sex on Outcomes.....	70
Observations of Trends Indicated by Raw Mean Percentile Score.....	73
Summary of Conclusions Drawn from Testing of the Hypotheses and Obser- vation of Trends from Raw Data.....	75
Recommendations Drawn from the Research in This Study.....	76
Questions for Future Research.....	80
Summary.....	83
Appendices.....	85
References.....	90

LIST OF TABLES

Raw Mean Percentile Score Difference Attained by Sex and Treatment Group Using Highest Score in Each Variable as Base of Zero.....	73
Mean Percentile Score Difference Compared by Sex.	75

LIST OF APPENDICES

Appendix

A	Observed Cell Standard Deviations.....	85
B	Error Correlation Matrix.....	86
C	Variance and Standard Deviation.....	87
D	Matrix of Correlations Adjusted Against Covariate.....	87
E	F-Statistic for Hypotheses I, II, and III.....	88
F	F-Statistic for Sub-Hypotheses A and B.	89

CHAPTER I

INTRODUCTION

Rising costs, energy availability and reduced student enrollments are forcing educators to become more introspective about the quality of the services they offer and about the types of emphases they pursue. The public is demanding evidence from institutional education that the tax dollars earmarked for educational services are indeed returning full benefits in the form of educated young people.^{1,2} Further, this pressure from the public seems to be demanding that those courses and activities which are not directly related to the cognitive development of students be eliminated from public education.

Institutional educators have responded with methods of evaluation which will support their methods of service. This defense seeking has led to the quite widespread adoption by educators of industrial models

¹Ben Lawrence, "Accountability Information: Progress, Pitfalls and Potential," New Directions for Higher Education, 20 (1979):51-56.

²B. Done, "A Call for Public Accountability, Great Britain," Times (London) Education Supplement, April 11, 1980.

of quality control.^{3,4}

The influence of Robert Mager's and Leon Lessinger's objective based industrial models in the late sixties and seventies on institutional educative processes is no longer novel or faddish but rather has become accepted as a means of evaluating success and quality in schools, classrooms and individuals.⁵

Since quality control ultimately focuses on the product, students' performance becomes the ultimate test of the system. Consequently, the debate over the benefits of passing or failing students has once more become a major consideration in the institutional educative process.^{6,7}

Schools and school systems are considering or have adopted policies which will require retention of students until they attain all necessary competencies at given levels.

³Y. Pai & Krueger, "Accountability at What Price?" Journal of Teacher Education, 22 (July-August 1979):20-1.

⁴A. Dittmer, "Assault with a Deadly Mandate," Media and Methods, 15 (September 1979):26-9.

⁵Leon M. Lessinger, Every Kid a Winner (New York: Simon and Shuster, 1970).

⁶Phi Delta Kappan, 59 (May 1979). Topic of this issue: accountability and minimum competency. Listed is the status of minimum competency laws by state.

⁷J. H. Sandberg, "K-12 Competency-Based Education Comes to Pennsylvania," Phi Delta Kappan, 61 (October 1979):119-20.

Are these retention policies effective? Do they make it easier for students to achieve? Is there a political expediency in the process that ignores the actual effectiveness of a person's learning? How is this policy change related to the response of critics who are demanding evidence of production in education?

The following arguments are given by those who espouse a strict posture toward retention. These proponents maintain that retention:

1. Helps support high academic standards.⁸
2. Enables pupils to work at levels more closely related to their achievement levels.
3. Maintains a smaller span of achievement levels with which a teacher must work. Consequently, teaching can be done more effectively.
4. Assists the low achiever by reducing frustrations they may encounter as they attempt to learn increasingly difficult material.⁹

⁸R. W. Templin, "A Check Upon Non-Promotions," Journal of Education, 123 (November 1940):259-60.

⁹"The Promotion/Retention Dilemma: What Research Tells Us," The School District of Philadelphia, Office of Research and Evaluation, December 1973.

Those who criticize strict retention policies argue that retaining students:

1. Creates undue social pressure on students by virtue of the stigma of failure which is attached to the procedure and the inherent problem of students being overage and/or oversize.^{10,11}
2. Does not significantly improve personal achievement levels.
3. Will not effectively improve or maintain academic standards of the institution.
4. Will not narrow the span of achievement levels with which educators must work.
5. Increases the incidence of school drop outs.
6. Creates discipline and behavior problems.¹²
7. Favors the higher socio-economic student.¹³

These arguments are profusely examined in the

¹⁰Ida A. Morrison and I. F. Perry, "Acceptance of Overage Children by Their Classmates," Elementary School Journal, 56 (1956):217-20.

¹¹W. H. A. Coeffield, "A Logitudinal Study of the Effects of Non-Promotion on Educational Achievement in Elementary School," Doctoral Dissertation, University of Iowa, Dissertation Abstracts, (1954):2291.

¹²Adolph A. Sandin, "Social and Emotional Adjustments of Regular Promoted and Non-Promoted Pupils," Child Development Monographs, 32 (1944).

¹³J. Rernherz and C. L. Griffin, "The Second Time Around," The School Counselor, 17 (1970):213-218.

literature and the issues have a bearing on any consideration of the retention question. Chapter II of this dissertation will examine in detail representative samples of contributions to the literature on retention.

Statement of the Problem

It will be the purpose of this study to investigate whether the retaining of a student in the early grades (Kindergarten-third) enables that student to perform at a level of achievement equal to that of his grade peers at the fifth grade level.

Essentially, the research will explore whether a student can maintain a comparable achievement level after a minimum time lapse of two years following the retention experience.

The average time lapse in this study is actually more than four years in duration (as will be reported in Chapter IV). This feature contributes to the uniqueness of the present study. As will be indicated in Chapter II, studies of retention to date have typically used smaller samples and shorter time lapses to investigate the effects of retention. The present study offers current data on the more durable effects of retention, if any.

Definition of Terms

Certain terms will be used throughout the study. It is necessary that these terms be defined.

Non-promotion. Non-promotion is here defined as the act of not assigning a student to the next higher grade at the end of the academic period or year. Further, it will imply that a student will not be given credit for any work done at a given grade level and consequently, will be prevented from participating in the succeeding grade level's instructional offerings.

Failure. For the purposes of this study, failure will be synonymous with non-promotion.

Retention. References to retention in this study imply non-promotion or failure of a total grade level. Unless so stated, retention will not connote the process of remembering or retaining knowledge.

Age peers. The term age peers refers to those students or children who are similar in age. Usually this similarity includes a band of months of approximately plus or minus six, from the age of a given individual. For the purposes of this study and in the context of the grade level system used in educational institutions, students who progress normally through the system after entering kindergarten or grade one, if no kindergarten is available, are considered age peers.

Grade peers. In this study, grade peers refers to those students who are assigned to a given grade level and whose achievement is compared with a specific grade level of achievement. Given the normal progression through the grade level system, students may be considered age peers as well as grade peers. However, upon failure or extra promotion a student's age peer group and grade peer group become different.

Low achiever or low achieving student. For the purposes of this study, low achieving students are considered to be those students who are achieving at a level which is perceived as too low to allow them to progress normally through learnings at given grade levels. Generally, low achievers would be classified as within the normal range of intelligence and would not possess any diagnosed development or learning disabilities.

MAT. Metropolitan Achievement Test-referred to throughout the study as MAT- is a nationally normed achievement test battery designed to measure student achievement. A further discussion of this test, what it is designed to do and its construction, will be found in Chapter III.

Contribution of the Study

This study will:

1. Make available some current data which may

be of assistance to teachers and administrators who are considering formulating retention policies.

2. Offer a decision base which may be useful to educators who are considering educational plans for individual students.
3. Contribute a longitudinal study to the literature which has certain controls not yet found in the research on retention.

Assumptions on Which the Study is Based

The fundamental assumptions which underlie this study and influence its approach, its methods and its recommendations will be stated at this point.

1. The grade level system is an integral part of the educational process in American institutional education.
2. Academic achievement of students may be effectively measured by nationally normed achievement tests.
3. Treatments students received in the public school classrooms connected with this study do not vary significantly from typical classrooms, schools or school districts. More concisely, the research in this study has been generated from sources which are

essentially comparable and represent a fairly adequate cross section of learning environments of institutional public education.

Limitations of the Study

This study is or may be limited for the following reasons:

1. The population sample may not represent the full range of student populations since it is derived from basically rural areas. However, the relative homogeneity of the sample in this study may represent an advantage, since it is the effects of the process of retention which this study attempts to isolate for investigation.
2. The time lapse between third and fifth grade may not be of adequate duration to examine the total progress a student may make throughout his/her subsequent academic career. That is to say, student achievement may accelerate, level off, or even decelerate in subsequent years, especially during or after puberty. Nevertheless, the time lapse employed in this study examines results over a longer interval than found in other studies to date.

3. There are no clinical controls, i.e., students' academic and personal environments have not been monitored or controlled. Therefore some variables in the delivery system which are unknown, may have been present to influence research outcomes. However, the size of the sample would appear adequate to support an assumption of "averaging out" of such variables.
4. Inherent errors in the construction and norming of the Metropolitan Achievement Test may be a limiting factor in the outcome of the results. But the MAT is a widely used instrument for assessing achievement.
5. Since the MAT was administered under classroom conditions, there were undoubtedly variations in timing of sub-tests, classroom decorum and other environmental factors. However, again the breadth of the samplings should assist in reducing the influence of these variables.

Design of the Study

Students were selected from rural school systems in the counties of Crawford, Oscoda, Ogemaw and Roscommon in north central Michigan. School districts

in these counties comprise the COOR Intermediate School District.

There were four criteria which governed the admission of data into the study sample. These criteria are as follows:

1. Students were in the sixth and seventh grades during the 1978-79 academic year. The rationale for selecting this specific year is that students in these classes were the last two classes on which complete MAT data were available. COOR school districts subsequently changed test instruments, and consequently there are no later comparison bases for a longitudinal study of achievement scores.
2. Test scores from the Metropolitan Achievement Test were available on each student at both the third grade and the fifth grade levels.
3. Students were considered to be within the normal intelligence range. They had not received help from any special education program for which processes such as an educational planning and placement committee (EPPC) are required for services. This criterion does not exclude those students who have received

speech therapy or titled reading program assistance.

4. There was no more than one retention experience for students included, and this retention experience was no later than grade three.

Instrumentation

The instrument used was the Metropolitan Achievement Test, 1970 Edition. Percentile scores from the following sub-tests were used:

Grade Three: Total Reading

Grade Five: Word Knowledge

Total Reading

Total Mathematics

Sample Selection Procedure

The sample groups in this study were determined by means of the following steps:

1. All students selected were identified per the four criteria stated above.
2. Then a "retained" sample was identified. These were those fifth grade students who had been retained in kindergarten, first, second or third grades. Throughout the data collection and analysis in this study, this group will be labeled group R.

3. Then the range of the reading scores for group R at the third grade level was determined.
4. Finally, those students who were not retained but whose third grade reading scores were within the range of group R were identified. The students in this group form the second sample and will be identified as group P, indicating "promoted" or "passed" students.

Analysis of Outcomes

The mean scores of groups R and P were compared using the appropriate sub-test achievement scores at the fifth grade level.

The MANCOVA model of analysis of variance was employed to test the levels of significance of differences between mean scores on the three sub-tests of the MAT, (Word Knowledge, Total Reading, and Total Mathematics).¹⁴

Organization of the Dissertation

Subsequent chapters will describe the study in more detail.

¹⁴Further discussion of the MANCOVA model and MAT will follow in Chapter III.

Chapter II will review current pertinent literature relative to promotion and non-promotion.

Chapter III will present the design and data collection procedure and details of the study along with the actual N's for groups R and P.

Chapter IV will discuss the findings with tabulated MANCOVA and levels of significance.

Chapter V will summarize and discuss the implications of the study and recommendations for further research.

Appendices are included which will set forth the various statistical details of the study.

CHAPTER II

REVIEW OF THE LITERATURE

This chapter will review literature relative to promotion and non-promotion of students. There are two broad categories of research which concern themselves with the topic: Research relative to the academic achievement and research related to social or affective growth. Although the concentration of the research in this dissertation deals quite specifically with the question of academic achievement, the literature dealing with both areas will be reviewed since several studies attempted to address both categories and the influence of affect may have a bearing on how one achieves in the school environment.

Gregg Jackson has trichotomized, by design type, the research on promotion/retention.¹ Each piece of research in this review will be classified according to Jackson's trichotomy.

¹Gregg B. Jackson, "The Research Evidence on the Effects of Grade Retention," Review of Educational Research, 45 No. 4:613-35.

Design Type I

Jackson's design type I uses an analysis in which retained students are compared with students who are promoted. This research design compares students who are retained or promoted under normal and existing policies and procedures.

Jackson cites two limitations of studies done according to design type I:

1. Although students were often matched according to IQ, achievement scores, academic grades, SES and/or teacher assessment, students who are retained apparently still experience difficulties not experienced by those who were promoted.
2. The wide variation in criteria for promotion among teachers and schools may influence the results of the study.

Design Type II

Design type II is a simple comparison of a student's condition before retention. This type design is the most frequently used by educators, since it is convenient and may employ data which are readily available, such as teacher assessments, grades, and/or placement test scores.

A major limitation of this design model is that

there are no data available which would show comparison with promotion as an alternative.

Design Type III

Jackson's type III model categorizes studies which compare experimentally or randomly, promoted and retained pupils. This is considered the most effective design type. However, the limitations are obvious by virtue of the fact that it is difficult to establish research settings which allow the experimental selection of students necessary for its consummation.

Studies Concerned with Affective and Social Progress of Retained and Promoted Students

Studies in this category generally address the following questions:

1. Do non-promoted students experience less frustration because they are working on learnings more commensurate with their achievement and ability levels?
2. What is the effect of retention on a student's personal-social adjustment?
3. Low achievement and maturity are highly correlated. Therefore, will students who are retained benefit by placement with a younger age-peer group?
4. What is the relationship between retention

and the school dropout rate?

McElwee's study in 1932 took into account the personality traits of children and what part this aspect played in the acceleration, promotion and retention of children. In this study which may be classified under Jackson's design model I, 300 children were examined. It was concluded that among the traits of quietness, calmness, quarrelsomeness, stubbornness, excitability and talkativeness, there was no difference among groups of children who were retained or promoted. However, most disobedient children were considered to be among the non-promoted group.²

Junior high pupils' adjustment to retention was examined in 1940 by R. O. Anfinson. In the study, which may be classified according to Jackson's design I, 116 non-promoted students were paired with 116 promoted students on the basis of school attendance, age, sex, intelligence and SES. Using the Symonds-Block Student Questionnaire and the Bell School Inventories to measure personality and adjustment, Anfinson concluded that no statements about destructive effects of non-promotion could be made. He found that there were poorly adjusted and well adjusted pupils in both groups. Additionally,

²E. W. McElwee, "A Comparison of Personality Traits of Accelerated, Normal and Retarded Children," Journal of Educational Research, 26 (1932):31-34.

it was his feeling that promotion and non-promotion are personal decisions and no patterns develop through the process which would show the benefits or dangers, affectively, of retention and promotion.³

N. M. Chansky's study conducted in Ulster County, New York, with first grade pupils (33 retained, 26 boys and 7 girls) used the California Test of Personality to assess adjustment. He concluded that there were no significant losses relative to personality growth among either retained and promoted students. Although both groups were still considered maladjusted, slight gains were observed for both but moreso for the promoted students.⁴

L. O. Briggs used design type I in his 1966 study of fifth and sixth grade male pupils who had been retained two years, by comparing them with randomly selected groups of successful males who had not been retained. The California Test of Personality revealed no significant difference between the groups. However, the teachers' perception of the retained pupils was that they were more withdrawn, lacked leadership

³R. D. Anfinson, "School Progress and Pupil Adjustments," Elementary School Journal, 41 (March 1941):507-14.

⁴N. M. Chansky, "Progress of Promoted and Repeated First Failures," Journal of Experimental Education, 32 (1964):225-37.

characteristics and tended to be more aggressive.⁵

A study by J. A. Chase in 1968 which followed Jackson's design type II, was conducted on 65 first and third grade students who had been retained. The students were given a battery of tests which included: Slossin Intelligence Test, Gesell Incomplete and Gesell Copy Forms, and Bender Visual-Motor Gestalt Test. Also questionnaires were completed by the retaining teacher, present or repeating teacher and the parents.

The results were:

1. Repeating a grade had met the needs of 75 percent of the children.
2. There seemed to be no emotional upset in 75 percent.
3. Ninety five percent of the parents were in favor of the retention.
4. There was no change in intelligence level.
5. The mean group performance on developmental tests was still nine months behind classmates.

He concluded that there are no negative social

⁵L. O. Briggs, "The Impact of Failure on Elementary School Pupils," Doctoral Dissertation, North Texas State University, (1966), Dissertation Abstracts 27 (1967):2719-A.

or emotional effects in children whose retention is based on immaturity. During the repeated year, the perceptual and motor abilities of children will approximate the expectations of the school system. First grade retentions are far more effective than second or third grade retentions.⁶

More clear cut results relative to the personal adjustment of retained and promoted students were reported by White and Howard. Their type I study of sixth grade children in six North Carolina school systems comprised of both rural and urban populations, showed that the most severe and consistent effects of failure on self concept occurred between those who were retained two or more times and those who were never retained. Using the Tennessee Self-Concept Scale they determined that there were four areas where significance occurred: 1. Total positive self concept, 2. self satisfaction, 3. family, and 4. social.⁷

Adolph Sandin's findings appear to be the most decisive among the research relative to affective growth and adjustment. He found that the difference between

⁶J. A. Chase, "A Study of the Impact of Grade Retention on Primary School Children," Journal of Psychology, 70 (1968):169-77.

⁷K. White and J. L. Howard, "Failure to be Promoted and Self-concept Among Elementary School Children," Elementary School Guidance and Counseling, 7 (1973):182-87.

the promoted and non-promoted students "contributed to bring out situations in which there was a barrier to good social relations."⁸ He also observed differences in interests, behavior, likes and dislikes. Further, non-promoted pupils exhibited less commendable behavior, feelings of discouragement and failure, dislike for school and school activities and were subject to criticism from teachers, parents and younger classmates.⁹

Sandin's study also determined that boys and girls were perceived differently by their teachers as to those children who were regarded as behavior problems. Retained girls were perceived as problems greater than six times more often than promoted girls. The cases of non-promoted boys perceived as behavior problems were less than four times greater than promoted boys. Totally, of all promoted students in Sandin's study 5.1 percent were regarded as behavior problems while 25 percent of the total non-promoted students were regarded as problems - a ratio of less than 1 to 5.¹⁰

Sandin's conclusions, which were researched

⁸Adolph A. Sandin, "Social and Emotional Adjustments of Regularly Promoted and Non-promoted Pupils," Child Development Monographs, 32 (1944):134.

⁹Ibid., 136.

¹⁰Ibid., p. 75.

according to Jackson's design type I, were reached through his administering an extensive series of questionnaires and rating surveys to students and teachers.

J. I. Goodlad's study done in Georgia in 1947-48 used design type I. He compared students in eleven elementary schools by matching them according to intelligence, achievement, social-economic status and location.

Students were compared using the CATP (California Test of Personality), Haggerty Olson-Wickham Behavior Rating Schedule, and peer rating data, which were collected on two questions: 1. Which three children in your classroom would you like to have as your best friends? 2. Which three children would you not like to have as best friends?

Goodlad found that the CATP did not yield any significant difference in favor of the promoted group on thirteen of the items, while seven items favored, significantly, the non-promoted students. Sociometric data favored the promoted group as did the social adjustment data.¹¹

11

J. I. Goodlad, "Some Effects of Promotion and Non-promotion Upon Social and Personal Adjustment of Children," Journal of Experimental Education, 22 (June 1954):301-28.

Twenty five grade one through three students, who were retained, were compared with 25 first through third grade students who were promoted in an urban school system by Paula Caplan. The focus of Caplan's study was classroom conduct. Although the scales were not sensitive enough to conclude anything relative to boys, Caplan concluded that promoted, low achieving girls behaved significantly better than did retained girls. Further misbehavior in girls influences grading and seems to be a major factor in decisions about promotion and retention. This was especially significant if the misbehavior was aggressive or disruptive.¹²

A final study concerned with the influence of retention on affective development of children was done by Shailer and Knudsen. It bears particular discussion here because it addresses the long term implications of retention.

Shailer and Knudsen examined the correlation between retention and the high school drop out rate. The study which was done in high schools in a southern city involved 100,000 students. They found that seven percent who had not repeated dropped out, while 27 percent of those who had repeated had dropped out.

¹²P. J. Caplan, "The Role of Classroom Conduct in Promotion and Retention of Elementary School Children," Journal of Experimental Education, 41 (Spring 1973):8-11.

Contrary to accepted belief, it was observed that the IQ was not a factor, either in the retention of students or in the drop out rate, i.e., students' ranges of achievement and IQ were not statistically different.

The consequences of non-promotion which seemed to influence the non-promoted pupils' decision to drop out were his/her relationship with family, relationships with peer groups, self concept and attitude toward school achievement.¹³

Although not conclusive as to the causative effects of the increased percentage of non-promoted dropouts, the study may indicate that there are negative long term implications of retention as it relates to school holding power.

The research evidence relative to the harmfulness or benefit of retention is not entirely conclusive. The studies which have been done which follow Jackson's design type I (Comparison of promotion and non-promotion) appear to favor promotion while those which may be categorized according to type II appear to favor retention. Although there is no overwhelming evidence, it is the opinion of this writer that, given the inherent difficulty with design type II to a point that studies

¹³Thomas Shailer, Dean D. Knudsen, "Relationship Between Nonpromotion and the Dropout Problem," Theory Into Practice, 4 (1965):90-94.

done under this design may be ignored or, at best, given less credence, the research would generally favor promotion if the affective growth of the student is the primary aspect to be considered.

Studies Concerned with
Retention and the Academic
Achievement of Students

Studies which have investigated the effects on achievement of promotion or non-promotion have considered the following questions to various degrees:

1. Will non-promotion raise standards of achievement?
2. Are teaching loads eased by virtue of the reduced variability of achievement levels of each grade?
3. Is the threat of failure a motivational factor in achievement?
4. Can a student's academic achievement be improved or maintained by retention?

Again, this review of the research relating to academic achievement is classified according to Jackson's trichotomy. Finally, studies which have explored the question of retention but have examined it from the perspective of different management systems or integration of other variables will be discussed.

Jackson's design type I was used by C. H. Keyes as he conducted a study in 1911 which included 5,000

students in the State of New York. Of the students, 20 percent showed improvement during the repeated year, while 40 percent did poorer work. There was no improvement demonstrated by the remaining 40 percent.¹⁴

A small study by E. B. Francis in 1939, which may be classified under Jackson's type I model, used a three based judgment of success: 1. Averaged report card grades of 70 percent or higher; 2. A score at grade level or above on the Gates Reading Test; 3. General achievement score at or above grade level on the Stanford Achievement Test. Of the sixty students in Glen Ridge, New Jersey Schools, eight showed no success; seven were successful on one of the three criteria; three were judged successful on approximately 50 percent of the criteria; seven demonstrated a 66 percent success rate while 35 were successful on 100 percent of the criteria.¹⁵

A. J. H. Gaite examined high school students who had failed a year of a course of study. He determined that the grades were better the second year; however, he felt that there was not significant enough improvement to warrant students' repeating a year and thus losing

¹⁴Charles H. Keyes, Progress Through the Grades of City Schools, (New York: Columbia University, 1911).

¹⁵E. B. Francis, "A Follow Up of Non-promotion," Journal of Education, 122 (June 1939):187-88.

time.¹⁶

Scott and Ames conducted a type II study in Connecticut in 1969. Twenty-seven elementary children who had been retained were examined the year after being retained in kindergarten, fourteen in first grade, three in second grade, three in third grade, one in fifth and one in sixth.

The criteria were based primarily on behavioral maturity and readiness rather than on achievement. Teacher grades and responses to parent and teacher questionnaires on adjustment toward repeating and attitudes of the child toward school were used as the measures of success.

They found that grades improved significantly and that attitudes were not negative and were significantly positive in most areas.¹⁷

W. H. A. Coeffield's study in 1954 using 190 matched seventh grade pupils concluded that failed students may make statistically significant educational gains while repeating a grade. However, they will probably not come up to grade norm. Failed and promoted

¹⁶A. J. H. Gaite, "On the Validity of Non-promotion as an Educational Procedure," (Madison: University of Wisconsin, 1969).

¹⁷B. A. Scott and J. B. Ames, "Improved Academic, Personal and Social Adjustment in Selected Primary School Repeaters," Elementary School Journal, 69 (1969): 431-39.

students who are compared in achievement at the time of failure, perform at about the same level when achievement of both is measured in the same higher grade. The promoted seem to perform better after a lapse of a common period of time.

Using the Iowa Test of Basic Skills standard score and grade equivalent units, Coffield specifically determined that progress of failed students is about four to six months less than matched promoted students. Further, during the two years following failure, the educational progress of failed students is not significantly greater than that made by promoted matchees during the year spent in the next higher grade.¹⁸

K. W. Ogden conducted a longitudinal study of one hundred high school students who were retained one year in elementary school. He wanted to determine whether these 100 students had reached the academic level of expectancy.

His conclusions:

1. Fifty percent of the students who were retained continued to have academic difficulty.
2. Age was a significant factor for

¹⁸W. H. A. Coffield, "The Effects of Non-promotion on Educational Achievement in Elementary School," 47 (1956):235-50.

initiating retention.

3. Intelligence was a key factor in differentiating retained and non-retained students.
4. The retained students, although judged academically successful, attained significantly lower scores on achievement tests.
5. Retained students did no better than did students who had been considered once for retention but then had been promoted.¹⁹

An interesting aspect of Ogden's study was his interviews with those students who had been retained in which 74 percent reflected a favorable attitude toward retention.

One of the most concise comparative studies carried out in recent years was done by Walter Worth in the Edmonton (Alberta, Canada) Public Schools in 1956-57. Worth's study was done on matched third and fourth grade low achievers and examined the achievement, social adjustment and personality traits of children as assessed by teachers.

Worth identified third grade low achieving students in the spring of 1955 and after the decision to retain students was made, he matched 66 retained students with

¹⁹K. W. Ogden, "An Evaluation of Non-promotion as a Method for Improving Academic Performance," Doctoral Dissertation, University of Southern California, Dissertation Abstracts International, 32 (1971):795-A.

66 who were promoted. In the spring of 1956, Worth administered the California Achievement Test and Gates Reading Test to both groups. Each teacher was also asked to rate the students on seven personality traits. A two point socio-metric questionnaire was given to classmates as a peer assessment.

Below is a summary chart of Worth's findings:

CALIFORNIA ACHIEVEMENT TEST

Sub-test	Favored non-promotion	Favored promotion	No. diff.
Reading Vocabulary		X	
Reading Comprehension			X
Total Reading		X	
Arithmetic Reasoning	X		
Mechanics of English			X
Spelling			X
Total Language			X
TOTAL ACHIEVEMENT BATTERY			X

GATES READING TEST

Sub-test	Favored non-promotion	Favored promotion	No. diff.
Paragraph Reading	X		
Word Recognition			X

PERSONALITY TRAITS AND
SOCIAL DESIRABILITY

Trait	Favored Non-promotion	Favored promotion	No diff.
Emotional Control			X
Creativeness			X
Judgment			X
Cooperation			X
Dependability		X	
Courtesy			X
Work Habits			X
Desirable Work Companion	X		
Undesirable Work Companion			X
Desirable Play Companion			X
Underdesirable Play Companion			X

Essentially, Worth's findings suggest that non-promotion does not necessarily improve achievement and socially, a retained student does not suffer adversely.

Although the majority of the areas of student behavior and environment examined by Worth were inconclusive relative to the benefits of retention or promotion, he cautions against wholesale or routine decisions in favor of retention because of the additional time commitment to schooling which is required for retained

students.^{20,21}

A significant limitation of Worth's very thorough study is the time span within which the study was done. Since the study examines achievement and adjustment the year following retention, the results may reflect either positively or negatively, the effects of repetition of curricular material as the retained students were recycled through the third grade.

The only significant type II study was done by Klene and Bronson in 1927. They examined the achievement of 100 randomly retained and promoted low achievers. Again, the assessment was made in the term following retention.

Briefly, their findings revealed:

1. The fifty promoted students showed greater progress as a whole.
2. Normal ability children gain more by promotion.
3. Below average ability children do not gain by repeating.
4. Pupils in grades 4-6 profit more from

²⁰W. H. Worth, "The Effects of Promotion and Non-promotion on Pupil Achievement and Social-Personal Development in the Elementary School," Doctoral Dissertation, University of Illinois, (1959) Dissertation Abstracts (1959) 20:3228.

²¹W. H. Worth, "Does Non-promotion Improve Achievement in the Language Arts?" Elementary English, 37 (1960):49-52.

promotion than do those in grades two or three.

5. Ninety percent of the promoted students were again promoted to the next higher grade.²²

Some Aspects of the
Retention-Promotion
Question As It Relates
To Achievement

Several studies have been conducted which have addressed specific arguments or questions relative to promotion or non-promotion.

Otto and Melby were concerned with what part the threat of failure played in motivating a child to achieve. Their 1934 study in Illinois focused on second and fifth grade students. Experimental students were assured that they would pass while control students were meted out the usual threats of failure if they did not achieve.

The results, based primarily on teacher opinion, showed that the elimination of threat did not affect, favorably or unfavorably, the quality of work done by the students.²³

²²V. Klene, E. P. Bronson, "Trial Promotion vs. Failure," Elementary School Journal, 29 (1929):564-66.

²³H. J. Otto, and E. O. Melby, "An Attempt to Evaluate the Threat of Failure as a Factor in Achievement," Elementary School Journal, 35 (April 1935):588-96.

The Omaha School system also examined the question of threat of failure and its effects on achievement in its study reported in 1942. Low achievers were not retained but rather left with their age peers and allowed to progress at their own rate. These failing students were termed "adjusted." Within five years 80 percent had attained normal progress. Teacher observations were that children were happier and more motivated when the fear of failure was lifted. More of them worked sincerely when the learning process was progressive, dishonesty was reduced and there seemed to be fewer social misfits.²⁴

Two studies addressed the question of variability of achievement levels and whether strict retention policies reduce that variability.

Coeffield in his study, which was cited earlier, concluded that there is no significant difference in the general level of variability between schools having high and low rates of failure. Further, he concluded that the high rates of failure lead to greater amounts of overage pupils.²⁵

²⁴F. Myers, "We Experiment with a Non-failure Program," Child Education, 18 (1942):205-09.

²⁵W. H. A. Coeffield, "A Longitudinal Study of the Effects of Non-promotion on Educational Achievements in Elementary Schools," Doctoral Dissertation: University of Iowa, Dissertation Abstracts, (1954):2291.

Cook selected from 148 Minnesota schools, nine which had the highest ratio of overage seventh grade students and nine which had the lowest ratio of overage pupils.

He concluded that:

1. The range of abilities with which a teacher must work is not significantly less in high ratio schools.
2. The high percentage of overage pupils in schools with high standards of promotion reduces intelligence of classes and significantly lowers the achievement average of the grades when compared with schools with more lenient standards of promotion.
3. There is evidence that adjustment of instruction to the pupils' ability is superior in every subject, except mathematics, in schools with low ratios of overageness.²⁶

Three policies seem to govern promotion. These are: 1. Grade standard; 2. Continuous promotion and 3. Continuous progress. The Phoenix (Arizona) School District Number 1 changed from a grade standard policy

²⁶W. Cook, "Some Effects of the Maintenance of High Standards of Promotion," Elementary School Journal, 41 (1941):430-37.

of promotion to a policy which was a combination of continuous progress. The board policy essentially stated that children were to be promoted on a basis of total adjustment rather than achievement alone.

Teachers were thus relieved of the guilt and frustration of the non- or low achieving child, consequently, they began to look more intently into what was good for each child.

The longitudinal data gathered from 1946-47 academic year through the 1955-56 academic year on fourth and sixth grade classes used the Iowa Every Pupil Achievement Test, Reading Sub-test, and the Kuhlman-Anderson Group Intelligence Test. Annually, 1240 fourth grade pupils and 1113 sixth grade pupils were tested.

The more lenient policy toward retention showed the following results:

Academic Year	Fourth Grade Mean Reading Grade Level	Fourth Grade Mean IQ
1946-47	3.47	93.84
1950-51	3.52	92.88
1955-56	3.40	91.67

The average age for fourth grade students went from nine years nine months in 1946-47 to nine years

two months in 1955-56.²⁷

Summary and Conclusions

The studies cited are considered to be the most significant studies done. Each one constituted original research and somewhat adequate designs and controls were employed. These studies have responded to the questions of affective harm or benefit of retention, motivational benefits and maintenance of academic standards. Also, the studies cited examined various aspects of achievement as they relate to the question of retention.

There still appears to be little or no decisive evidence as to what benefits retention accomplishes. Especially lacking is research relative to longer term effects on achievement. It is the intent of this dissertation to examine an aspect of student achievement over appreciable intervals of time.

²⁷William Hall and Ruth Demarest, "Effect on Achievement Scores of a Change in Promotional Policy," Elementary School Journal 58 (1958):204-07.

CHAPTER III

METHODOLOGY

Chapter III will present the design of the study, population and sampling procedures used in the study, instrumentation used, research hypotheses and collection of data.

Sample Population

Students were selected from rural elementary schools in the counties of Crawford, Oscoda, Ogemaw and Roscommon in north central Michigan. The counties comprise the COOR Intermediate School District.

The population from which this study drew its samples consists of small to moderate size school districts. The smallest district comprises a K-12 student enrollment of approximately 500 and is considered totally rural. The students in this district are from families of northern European descent. There seems to be strong family ties and deep religious bonds. Basic education is highly regarded. Consequently, the relationships between school and family is cooperative and the school's authority is greatly respected. Economically, families are considered low to middle income and generally dependent on earnings from the land,

agricultural related industries and small businesses. Although there has been a considerable amount of inward migration in recent years, of families from the industrial and more populous areas of the state, the demographic change is comparatively less than for other areas in Northern Michigan.

The larger school districts from which the sample is drawn have enrollments of 2,000 and 3,000 respectively in grades K-12. Geographically, these districts encompass towns with populations of approximately 2,000 as well as some totally rural areas. Economically, these districts draw from upper middle to lower income populations. Light industry, tourism and farming form the major occupational patterns. Again, during recent years these school districts have experienced some inward migration of families from the more populous areas of Michigan.

Generally, the COOR Intermediate School District appears to comprise a fairly homogeneous sample of the general population in two particular respects. The racial make up is more than 95 percent Caucasian, thus for all practical purposes, minority groups are largely absent. The second respect is that there is a much smaller ratio of students from upper socio-economic classifications than would be observed in areas nearer larger urban settings.

The advantage for selection of the COOR district's population are: 1. It contains a fairly homogeneous sample of students, i.e., socio-economically, middle to low and culturally rural to small town. 2. There are few sophisticated programs and educational services offered students which may influence results of the study. Philosophically, the districts which comprise COOR emphasize the basic learning of skills. 3. The testing program throughout COOR has been coordinated for several years which makes a longitudinal examination of students' progress practical.

Procedures for Collection of Data

School administrators were contacted in each local district and permission was granted to collect the necessary data.

During the summer and fall of 1980, visits were made to the schools and through inspection of the cumulative files of approximately 1100 students, the following information was obtained on each case contained in the sample:

1. The promotion record of each student.
2. Test scores from the MAT administered to each student at the third grade level.
3. Test scores from the MAT administered at the fifth grade level.

4. Determination of whether a student had participated in special education programs.

Total anonymity of the students who became part of the sample groups and also the individual schools and school districts is assured by the fact that no identifiers are used in the research analysis. The data on each student are considered as part of the whole with no statistical information of individuals being reported.

Criteria for Selection of the Sample

The following criteria were used for selection of the sample:

1. Students were in the eighth and ninth grades during the 1980-81 school year. Selection of students in these years was made because they were in the fifth grade the last two years the MAT was administered through COOR in all school districts. The sample is composed of students who were in fifth grade classes during the 1975-76 and 1976-77 school years.
2. There were MAT scores available on each student at both the fifth and third grade levels.
3. Students were considered to be within the

normal intelligence range, i.e., they must not have received service from any special education program except speech and titled reading programs.

4. Students had no more than one retention experience.
5. Any retention experienced by students must have been no later than third grade.

Procedure for Selection of Sample Groups

The final statistical analysis was done by comparison of two sample groups. These groups were determined by the following procedure:

1. Students in both samples were identified according to the sample selection criteria stated above.
2. Fifth grade students who had been retained in the kindergarten, first, second or third grades were classified as the retained sample, which then became group R.
3. Percentile scores attained by Group R students on the third grade MAT total reading sub-test were identified and a range was established. The scores ranged from 2 to 70. Since the highest score of 84 was a single score and was 14

percentile points above the next highest, it was decided to exclude this case from the sample as atypical and set the upper limit of the range at the seventieth percentile range. The total number of cases which became a part of the retained sample was 49.

4. The scores of the remainder of the students who met the original selection criteria were then compared with this 2 to 70 percentile range and those whose reading sub-test scores were within the range formed group P. The total number of cases comprising group P was established at 200.

Since students in groups R and P were within the same range, they are all considered to be low achievers. That is, all group P students could be considered potential R students. It should be noted further that although groups R and P were composed of grade peers, the average chronological age of group R would be approximately one year older than group P.

Instrumentation

Achievement test data used in this study were taken from scores on the Metropolitan Achievement Test, Form F.

Three sub-tests from the MAT were used as data:
Word Knowledge, Total Reading and Total Mathematics.

The sub-test, Word Knowledge, is basically a word recognition or vocabulary test. An analysis of the vocabulary of 11 series of readers was made and the words were tabulated according to the frequency of use and grade level. This test is "a representative sample of the words used in widely circulated reading series, which were shown to discriminate effectively between students of good and poor vocabulary."¹

The Total Reading sub-test has as its emphasis comprehension. The four aspects of reading comprehension which are measured are: Main thought, literal meaning, relationships among ideas and word meaning from context. The publishers constructed this sub-test so that the material would reflect achievement in each of the four areas mentioned above as the student encounters it in his/her reading instruction, as well as in reading situations occurring in other curricular areas at each grade level.²

¹Harcourt Brace Jovanovich, Inc., Manual for Interpreting Metropolitan Achievement Tests (New York: Harcourt Brace Jovanovich, 1962):33.

²Harcourt Brace Jovanovich, Inc., Metropolitan Achievement Tests Special Report Number 1, Content Development (New York: Harcourt Brace Jovanovich, Inc., June 1971):10.

The publishers observe that

One of the best indications of the validity of this test as a measure of reading ability as contrasted to a measure of intelligence is that students can be found with a high order of intelligence who do poorly on this test. In other words, they have specific disabilities in reading that are correctable under proper instruction.³

The Mathematics sub-tests cover three areas:
computation, concepts and problem solving.

The computation items are designed to give straightforward assessment of the pupils' ability to work with numbers. The levels progress in difficulty through the use of more difficult combinations with basic operations, use of fractions, decimals, and percents,....⁴

The concept test evaluates pupils' knowledge of fundamental principles and relationships in math.

Generally, this section tests what might be considered the "modern mathematics" programs.

Problem solving attempts to evaluate the pupils' total development ability in mathematics. It requires reasoning with numbers and operations. Each level includes problems demanding use of all four fundamental operations and some multiple step-problems.⁵

Also included are problems dealing with charts and graphs and number sentences.

The mathematics tests were reviewed in the

³Ibid., p. 35.

⁴Ibid., p. 11.

⁵Ibid., p. 11.

Eighth Mental Measurements Year Book by R. C. Allen.

His comments were:

The Metropolitan Math test is a solid middle of the road test which would be appropriate for the middle 40 to 50 percent of the children in school today.⁶

He feels that the test deserves "a better than most" label. The major limitation he felt the test had was that "too many items were tied to specific instructional practice."⁷

Item Selection, Statistical
Information and Reliability
of the MAT

Relative to test item inclusion there were seven criteria which the test publishers followed in their selection.

1. Average item difficulty had to be maintained at .55 for each intended grade.
2. The range of item difficulty had to be maintained between .20 and .90 in each test.
3. High item discrimination index items are given preference except when content specifications were involved.

⁶R. C. Allen, "Review of the Metropolitan Achievement Test, Mathematics Sub-test," in Eighth Mental Measurements Yearbook, Vol. II (New York: Gryphon Press, 1978), p. 424.

⁷Ibid., p. 424.

4. Grade progression in item difficulty preference was to be given.
5. Items were to have low inter-test correlations. Items must overlap between batteries to improve continuity of measurement across battery levels.
6. Each sub-test was to have a reliability of around .90 at each single grade level.⁸
The reliability estimates for all sub-tests were determined by the split-half, odd even method and corrected by the Spearman-Brown formula. Median split half reliabilities range from .93 to .96 at both levels of the test used in this study.⁹

Standard errors of measurement calculated in terms of grade equivalents range from .3 grade equivalent units for spring standardization groups.¹⁰

Statement of Hypotheses

The primary focus of this study is a comparative examination of fifth grade achievement test scores of

⁸MAT Special Report No. 1, p. 14.

⁹Harcourt Brace Jovanovich, Inc., Metropolitan Achievement Tests Special Report Number 10, Reliability Estimates and Standard Errors of Measurement, (New York: Harcourt Brace Jovanovich, Inc., June 1971):2.

¹⁰Ibid., p. 7.

passes and retained low achieving students.

The question: Will students who have a retention experience in grades K-3 maintain or attain achievement test scores which are similar to low achieving students who were not retained?

Three hypotheses were formulated which addressed the three tested areas at the grade five level. These areas corresponded to the three MAT sub-tests.

These three hypotheses, stated in the null form, are:

Hypothesis I: There is no difference in the mean percentile scores at the fifth grade level of retained and promoted students on the Word Knowledge sub-test of the MAT.

Hypothesis II: There is no difference in the mean percentile scores at the fifth grade level of retained and promoted students on the Total Reading sub-test of the MAT.

Hypothesis III: There is no difference in mean percentile scores at the fifth grade level of the retained and promoted students on the total mathematics subtest of the MAT.

The question stated above and the hypotheses attend themselves to the total population sample with no discerning between sexes. However, the factor of sex may have an influencing characteristic on the outcome. That is to say, that boys' mean scores may be higher or lower than the girls' mean scores on the various sub-tests

and therefore, the outcomes would not reflect the total population samples.

To address this problem two sub-hypotheses were generated. Sub-hypothesis A was designed to answer these questions: Is there interaction by sex, between the P and R groups? Will boys be expected to achieve similarly on the third grade measure and the three fifth grade measures? Is the effect of sex constant over both treatments of pass and retain?

The sub-hypothesis A stated in the null form is: There is no interaction, by sex, on the achievement measures of the retained and promoted students at the third grade and fifth grade levels.

Sub-hypothesis B has as its intent the examination of the difference between how boys and girls scored on the various measures. The questions to be answered: Is there a difference between boys' and girls' scores on the third and fifth grade tests? Do the scores of boys differ significantly from those attained by girls on any or all of the tests? Are retained boys' testing scores similar to retained girls' and passed boys' scores similar to passed girls' scores?

Stated in the null form, sub-hypothesis B is: There is no difference in the scores of males and females in groups R and P at the third and fifth grade levels.

Program of Analysis

An appropriate statistical program which would test all the variables was needed. The original intent to use a multivariate analysis of variance (MANOVA) proved to be inadequate because of the several variables introduced and also the anticipated interaction each would have on the other. A more concise multivariate analysis of covariance (MANCOVA) design was chosen. This analysis design would purify the interaction of the data and also allow for sample differences within the established range of third grade scores.

The following observations of Jeremy D. Finn, author of the MANCOVA model used in this study, demonstrate its appropriateness of use:

Multivariate analysis may be conceptualized in two ways. First, it is a means of analyzing behavioral phenomena. It is based upon the realization that hardly any form of human behavior worthy of study has only a single facet; that behind any measureable trait are components that covary only partially; that a "better" scientific description of any behavior is derived through some degree of finer analysis. Further, no observable behavior results from a single antecedent.

The second conceptualization of multivariate analysis is fitting a set of algebraic models to situations with multiple random variables, usually criterion or outcome variables, which are measures of the same sample(s) of subjects.

The class of experimental designs known as "repeated measures" designs denotes a multivariate situation. In this case each

subject is measured on a given scale at two or more points in time or under differing experimental conditions.

In each instance, the analysis of a single summary measure--for example, a total or average score--will result in the loss of the information conveyed by the individual scales. Statistical analysis of each of a series of measures separately will result in redundancy which in turn will threaten the validity of the interpretations drawn from the data. Use of the appropriate multivariate model will allow the researcher to retain the multiple scores and to treat them simultaneously, giving appropriate consideration to the correlations among them.¹¹

Specifically, as it relates to this study, the MANCOVA allows for refinement and/or compensates for these problems:

1. The inequalities between groups R and P on the third grade test (independent variable). Although all students in both R and P had achievement scores within the established range, it was anticipated that the mean scores of the groups would differ. Group P would have a mean higher since the N for this group was higher and statistically, this group would tend to produce a mean which would approach the fiftieth percentile level.

¹¹Jeremy D. Finn, A General Model for Multivariate Analysis, (New York: Holt Rinehart & Winston, 1974):vii.

2. The correlations between sub-test scores would be high, therefore statistical refinement would be necessary to make each sub-test statistically independent.

The specific computer program chosen to be used was one designed by Jeremy D. Finn entitled: Multivariate: Univariate and Multivariate Analysis of Variance, Covariance, and Regression--Version 6.1.¹²

Raw data were distributed by status (retained, passed) and sex (male, female). Four cells were established for these data. The cell breakdown and the N's for each cell is as follows:

Cell 1: Pass, male, N=85

Cell 2: Pass, female, N=115

Cell 3: Retained, male, N=32

Cell 4: Retained, female, N=17

Total N=249

It is noteworthy to observe the male-female ratio in the passed and retained groups. There is 25 percent more passed females than passed males while the retained group is composed of almost two times more males than females.

The third grade MAT total reading scores were

¹²J. D. Finn, Multivariate: Univariate and Multivariate Analysis of Variance, Covariance, and Regression, (Chicago: National Educational Resources, Inc., 1972), Version 6.1.

designated the independent variable (predictor variable or covariate). Dependent variables were identified as the three fifth grade measures: Word Analysis (vocabulary), Total Reading, and Total Mathematics.

Chapter III presented pertinent background on population samples, a discussion of instrumentation which is employed in the study, hypotheses and subhypotheses to be tested and the program of analysis employed.

CHAPTER IV

DATA ANALYSIS

Chapter IV will examine the results of the data collection analysis, give pertinent tabulated statistics, and discuss the findings with regard to the hypotheses. Further tabulation giving fuller details are contained in the Appendices.

Data On Means and Correlations

Means of the percentile scores for the four cells were calculated across the four test scores. These calculated means are as follows:

Treatment Status, sex	Grade 3 Tot. Rdg.	Grade 5 Wd. Analysis	Grade 5 Tot. Rdg.	Grade 5 Tot. Mth.
Pass, Male	.487765	.520353	.535059	.556000
Pass, Female	.512522	.469217	.476870	.515130
Retain, Male	.320625	.363437	.363437	.410000
Retain, Female	.354706	.327059	.288824	.306471

Those students who were included in the retained group (group R) experienced the retention treatment in the following grades:

Kindergarten	23
First	18
Second	5
Third	3

The average grade in which retention was experienced for the students in group R was between kindergarten and first grade or more than four years before the fifth grade testing.

On the third grade total reading measure, passed females scored slightly above the passed male group and the mean of the retained female group was slightly higher than that of the retained male group. Observation of the fifth grade Total Reading scores shows a change in the rank position of the groups. The passed male group attained a slightly higher score than the passed females and similarly, the retained male group attained a mean score above the retained female group.

This ranking position remained the same on the other two fifth grade measures--Word Analysis and Total Mathematics.

The differences between the highest and lowest mean scores were greater on the fifth grade measures than on the single third grade measure. Passed females' score of .5125 was 19 percentile points higher than the retained males' score of .3206 on the third grade total reading, however, the difference between passed males and retained females, the lowest scoring group on the fifth grade measures, increased 20 percentile points on the Total Reading and Math sub-tests.

The observed cell standard deviations for the

sixteen reported means were quite small. The smallest standard deviation of .150499 occurred on the third grade total reading measure for passed females. The largest standard deviation occurred in the measure of fifth grade total mathematics for retained females. This standard deviation was .279999.¹

Correlation data demonstrate a moderate to high correlation between the independent (third grade measure) and dependent variables (fifth grade measures). This correlation is expected, given the similarity of both tests. A correlation of .517 between total reading, third grade and word analysis, fifth grade, was also relatively high. Again, this correlation may be expected since curricular content is quite similar. The lowest correlation appeared between total reading, third grade and total math, fifth at .428.²

The Influence of Sex Differences on Outcomes

The question of sex differences and how this relates to or influences the examination of the three main hypotheses needs inspection before the main hypotheses can be reviewed.

The two sub-hypotheses established will be

¹See Appendix

²See Appendix

examined now.

Sub-hypothesis A: There is no interaction, by sex, on the achievement measures of the retained and promoted students at the third and fifth grade levels.

Essentially, this sub-hypothesis deals with the question of how consistently each sex performs over the measures. Did males and females score consistently, whether they were identified with the passed or retained group, or were there differences in their scores, when compared by sex, which would demonstrate an inconsistent performance by either one or both sexes?

A total F ratio of .6286 was produced for this test with a P of less than .5972. Statistically, the null hypothesis cannot be rejected. That being the case, there is no evidence for rejecting the following two assumptions: (a) there is no interaction between sexes, (b) there is consistency in performance by sex.

Sub-hypothesis B: There is no difference in the scores of males and females in groups R and P at the third and fifth grade levels.

The outcomes of the total samples--groups R and P--may not be accurate if those outcomes were weighted because of differences in the scores attained by the male or female groups.

A statistical comparison of the males and females showed an obtained F ratio of 1.241 and a level of

significance at less than .2956. This represents a probability greater than can be accepted. Therefore, the null hypothesis cannot be rejected, and there is no basis for rejecting the assumption that both groups performed similarly.

The assumptions drawn from testing sub-hypothesis A and B as stated above are such that it is logical to assume that sex is not a statistical factor in analysis of the data as they relate to the three main hypothesis.

Organization of Data Analysis as It Relates to the Hypotheses

The analysis program used, MANCOVA, required data to be analyzed and organized in a specific manner which would not permit independent application of statistical data to the three hypotheses. The interdependence of raw data, sample sizes and the need to equalize the population samples at the third grade level made it necessary to use MANCOVA as the analytical tool. However, the limitations of the program require some specific procedures for analysis.

Each variable can be examined using two statistical methods. Method 1: Univariate analysis. This method examines the relation between groups R and P by simply analyzing the differences on the three sub-tests. The

limitation of this method is chiefly that there is no adjustment for the high correlation among the three sub-tests. Outcomes under method 1 are very interdependent and would tend to reflect this in the final analyses.

Method 2: Covariate analysis or adjusted variance among the three variables and employment of a step down F ratio. This method permits the statistical outcomes among the three variables to be examined more independently. Further, this method refines the F ratio, thus enabling a more concise prediction of the statistical difference. However, in the process of refinement, it is not possible to completely separate or exactly identify and therefore test each hypothesis as it relates to its corresponding sub-test. The statistical output was organized in such a way that each adjusted comparison must be considered as part of the other score comparison.

Analysis of the Grand Means

A comparison of the grand mean of all three dependent variables between groups P and R showed a statistical difference of less than .0009 with an F ratio of 5.7411. This observation could well indicate that there is an overall statistical difference between groups P and R.

Analysis by Dependent
Variable, Using Uni-
variate F Ratio--
Method 1.

The sub-test scores data reflect statistical parts of the grand mean data reported above. Using the Type I error of $P=.05$ as the criterion for assessing levels of significance, a more concise determiner of the significance on the three sub-tests is the criterion established by dividing $P=.05$ over the three sub-test measures. Three is determined as the denominator because it represents the three sub-test measures. The new determiner for levels of significance then becomes .017 ($.05/3$).

Word Analysis sub-test comparisons of groups P and R show an F ratio of 8.7623 with P less than .0034. This probability is considerably less than the established criterion of $P=.017$, thus it would indicate that there is a statistical difference between groups P and R on the Word Analysis sub-test scores.

Similarly, the Total Reading sub-test scores of groups P and R are significant at the .0002 level with a univariate F ratio of 15.5432. This is much less than the criterion of .017. Again, this would indicate a difference, statistically, between the achievement scores of groups P and R.

The means of groups P and R differ less on the Math sub-test. The level of significance is 65 times

greater than that for sub-test Total Reading, with P less than .013. However, when compared with the adjusted Type I error level of significance of .017, mathematics scores could still be considered statistically different.

Analysis by Dependent
Variable, Using Multivariate,
Step Down F Ratio

The design for analysis using this method requires that the mathematics sub-test be observed first. This sub-test showed the lowest correlation and therefore could be termed the most independent. Adjusted F ratio (step down F) for this sub-test is 1.3464 with a P of less than .2471. The Type I error of .05 would indicate there is no significance between R and P groups on the mathematics sub-test.

The second piece of datum is a combination of the mathematics and total reading scores. However, since the statistical difference between R and P on the math scores was not significant, it is possible to consider the combination as a singular entity, reflecting only the total reading score.

The sub-tests, Total Reading and Mathematics total has an adjusted F ratio of 6.8807 and a level of significance of .0093. Using the Type I error of .05, significance can be established. Again, since the significance here does include the mathematics sub-test which was not significant, it can be established that

the significance represents the difference between R and P on the Total Reading sub-test only.

The final datum represents a total of the data from all three sub-tests, Mathematics, Total Reading and Word Analysis. Although the level of significance for this datum is .0034, and is statistically significant at the .05 level, it cannot be demonstrated as representing the Word Analysis sub-test outcome. This measure can be refined to the point that it represents both Total Reading and Word Analysis to the exclusion of the Mathematics sub-test since the Math sub-test demonstrated no significance and therefore can be excluded. However, the .0034 level of significance may be influenced by the Total Reading score and consequently, indicate a false conclusion. Because of this no inferences can be drawn about the sub-test Word Analysis.

Testing the Hypotheses

The hypotheses will be examined in light of the two methods described above.

Hypothesis I: There is no difference in the mean percentile scores at the fifth grade level of retained and promoted students on the word knowledge sub-test of MAT.

The analysis of the differences between the mean scores of groups R and P on the word analysis sub-test according to the univariate F ratio method does demonstrate

a statistically significant difference between groups R and P. On this basis the null hypothesis can be rejected. The level of significance for this sub-test and method is less than .0034 which is far less than the adjusted Type I error of .017 level of significance.

However the test of this hypothesis cannot be trusted using the adjusted F ratio (Method 1). Therefore, no statistical conclusions are possible relative to Hypothesis I.

Hypothesis II: There is no difference in the mean percentile scores at the fifth grade level of retained and promoted students on the total reading sub-test of the MAT.

Hypothesis II presents a clearer basis for analysis by virtue of the material presented earlier. Examination of the test for difference between groups R and P on the mean scores of the sub-test Total Reading using Method 1, univariate test, it is possible to conclude that the null hypothesis may be rejected. The level of significance of .0002 is again, far less than the adjusted .017 level of significance required for the Type I error.

Method 2, adjusted F ratio, also allows for the rejection of the null hypothesis. The level of significance attained, .0093, represents a level less than the Type I error of .05.

Statistically, there is a demonstrable difference between groups R and P on the sub-test, Total Reading. Given the fact that both methods demonstrate that there is a difference, Hypothesis II may be rejected.

Hypothesis III: There is no difference in the mean percentile scores at the fifth grade level of the retained and promoted students on the Total Mathematics sub-test of the MAT.

The difference in the means of groups R and P on the total mathematics sub-test was statistically different when analysis Method 1 was used. However, this difference was very close to the adjusted P of .017 and when rejecting the null hypothesis there is the possibility of introducing the Type I error.

Reviewing the mean difference after using an adjusted F ratio it is clear that there is no statistical difference between groups R and P. The obtained level of significance of .2471 is too great to accept at the .05 level of probability.

Relative to the above observations, the null hypothesis cannot be rejected. Statistically, there is no evidence of significant differences between groups P and R of the Mathematics sub-test.

Summary

In this chapter information and rationale regarding the use of the MANCOVA program of analysis was given.

Data on means, standard deviations, correlations and final outcomes were stated and explained. Sex variables were discussed and analyzed and the three hypotheses were tested.

Chapter V will present a discussion of the various outcomes, their implications, the limitations of the research and questions which may be answered in future research projects.

CHAPTER V

Chapter V will be a review and discussion of the statistical outcomes of the data as they relate to the three main hypotheses and two sub-hypotheses. It will also contain an examination of the trends indicated by collected raw data as well as inferences drawn from and conclusions about the outcomes. Recommendations for management systems in view of the conclusions and recommendations for future research will also be included.

This study was designed to examine the question of retention with consideration for two unique characteristics which have not been a part of previous studies. The characteristics are: 1. A comparison of the achievement of passed and retained students using the external measure of a nationally normed achievement test. 2. An examination of students' achievement several years after the retention has taken place.

Those students who were included in the retained group (Group R) experienced the retention treatment in the following grades:

Kindergarten	23
First	18
Second	5
Third	3

By far the greatest majority were retained in the first two years of their education experience, Kindergarten and first grade. The average number of years between retention and the fifth grade testing was 4.25. On the whole, this could indicate that the immediate short term effects of retention, whether positive or negative, are a negligible influence on the fifth grade measure. Therefore, the long range and lasting effects of retention, which are the more important to be considered, will influence the achievement outcomes.

Consequently, this longitudinal approach permits a look at retention from a perspective of realistically assessing future expectations of retained students.

All students examined in this study were considered to be low achievers by virtue of the fact that their third grade MAT total reading achievement scores were within the same range. Further, these scores were statistically equated during the data analysis to compensate for the differences in the mean score outcomes on the third grade Total Reading sub-test.

Given the above information, all outcomes of this study are directly related to the performance of low achievers.

There are three possible outcomes in this study for each of the three hypotheses tested:

Outcome 1: The mean percentile of group P is significantly greater than the mean percentile score of group R.

$$(M_P > M_R)$$

Outcome 2: There is no statistically significant difference between the mean scores of groups R and P.

$$(M_R = M_P)$$

Outcome 3: The mean percentile score of group P is statistically significantly less than the mean percentile score of group R.

$$(M_P < M_R)$$

Ideally, it might be intended that the group which received the treatment of retention would benefit from that treatment and therefore attain achievement scores above the nontreatment or passed group. Thus, the most hoped for outcome, given the introduction of the treatment of retention, would be outcome 3, the mean percentile score of group P is statistically less than group R. $(M_P < M_R)$

Grand Mean Difference

This statistical information was not part of the original emphasis of the study. However, it was reported in the research analysis and does represent

an overall indicator which may be relevant to the study.

The grand mean difference is a composite of the mean scores of the three fifth grade sub-tests, Word Analysis, Total Reading and Total Mathematics. There was a statistical difference between the grand mean of group R and group P. This difference was significant at P less than .007 with an F ratio of 5.7411. The grand mean of the passed group (group P) was higher than the retained group (group R). Therefore, the groups were statistically different with group P attaining the higher grand mean score.

Influence of Sex on the Outcomes-Sub-Hypotheses A and B

In Chapter V the results of the outcomes of the two sub-hypotheses relating to sex were reported. From a statistical perspective, neither variation in consistency of performance by males and females nor the difference in performance between males and females can be observed to be significant. Although there were no statistical differences between the sexes and their performance, some indications were evident from mean raw score data. These indications will be discussed later.

Now to address the original question and basis of this study in the light of the hypotheses: Is

retention of students an effective method of maintaining and/or improving achievement levels of the low achiever?

The material presented in Chapter IV regarding the testing of the three hypotheses obtained the following results:

Hypothesis I (Word Analysis-vocabulary)

The program used for analysis rendered inconclusive evidence regarding this hypothesis. Consequently from a statistical position no further discussion of this hypothesis is possible.

Hypothesis II (Total Reading)

This hypothesis cannot be accepted in the null form. There was a statistically proven difference between the mean scores of groups R and P. This difference was such that the mean score of group P was greater than group R. ($M_P > M_R$)

Hypothesis III (Total Mathematics)

Hypothesis III cannot be rejected in the null form.

The means of groups R and P were not statistically significantly different on the fifth grade mathematics test.

The similarity was such that

$$M_R \approx M_P.$$

None of the proofs of these hypotheses resulted in the ideal outcome of the treatment group attaining a significantly higher mean percentile score than the non-treatment group.

The mean percentile score attained by group R was significantly lower than that achieved by group P on the total reading sub-test (Hypothesis II). It could well be, therefore, that retention has inhibiting, detrimental influence on students' capacity to achieve successfully in reading.

Hypothesis III demonstrated that there was no statistical difference in the achievement levels of retained and promoted students. Given the fact that both groups were initially classified as low achievers, the treatment of retention does not seem to be an effective method of assisting students to improve their achievement.

Observations of the
Trends Indicated by
Raw Mean Percentile
Scores

Although not conclusive, the raw mean percentile score data present some patterns and trends which may indicate the effectiveness of the treatment of retention and therefore, be used for producing decision making rationales regarding retention.

The following table shows the raw mean percentile score differences as attained by each sex and treatment group on the four tested measures:

RAW MEAN PERCENTILE SCORE DIFFERENCE
ATTAINED BY SEX AND TREATMENT GROUP
USING HIGHEST SCORE IN EACH VARIABLE
AS BASE OF ZERO

Treatment Group and Sex	Variable (Sub-test)			
	Third Grade		Fifth Grade	
	Total Reading	Word Analysis	Total Reading	Total Math
Passed, Male	-2(49)	0(52)	0(54)	0(56)
Passed, Female	0(51)	-5(47)	-6(48)	-4(52)
Retained, Male	-19(32)	-16(36)	-18(36)	-15(41)
Retained, Female	-16(35)	-19(33)	-25(29)	-25(31)

As indicated by the above table, passed males attained the highest scores on the three fifth grade

variables and passed females scored a few mean percentile points below.

The greatest difference in each variable was between passed males and retained females. In two variables, total reading and total mathematics, the difference was 25 percentile points or one quartile.

It is interesting to note that on the third grade test both passed and retained male groups attained observed mean percentile scores slightly below their corresponding female group, while on all three fifth grade measures, the positions were reversed with males attaining observed mean scores slightly above their corresponding female groups.

Observation of mean percentile score differences, compared by sex, shows the females having the greatest differences between the pass and retain groups. The difference between the retained and passed female scores increased three mean percentile points from third grade total reading to fifth grade total reading, while the passed and retained males' mean score difference decreased by four percentile points on the same two measures.

The difference between passed and retained males and females on the total mathematics sub-test was only one percentile point.

The following table displays these differences:

MEAN PERCENTILE SCORE DIFFERENCES
COMPARED BY SEX

Treatment Group and Sex	Total Reading 3rd	Word Analysis 5th	Total Reading 5th	Total Math 5th
Passed, Female	51	47	48	52
Retained, Female	<u>35</u>	<u>33</u>	<u>29</u>	<u>31</u>
Difference	16	14	19	16

Passed, Male	49	52	54	56
Retained, Male	<u>32</u>	<u>36</u>	<u>36</u>	<u>41</u>
Difference	17	16	13	15

Although these differences cannot be verified statistically, they do offer some suggestive trends for consideration. The effects of a retention treatment might possibly be more negative for females than males in the curricular area of reading. Mathematics achievement is somewhat consistent for each treatment group, i.e., the differences between passed and retained males are essentially similar to those for females.

Summary of Conclusions
Drawn from Testing of the
Hypotheses and Observations
of Trends from Raw Data

Retention does not appear to be an effective means of raising or maintaining achievement.

There were no comparative gains for the retained group in total reading which would demonstrate that the retention treatment was actually beneficial to their achievement growth. On the contrary, it appears that the retained group gained less in achievement growth.

Although mathematics achievement for the retained group was comparable to the low achieving passed group, retention does not apparently represent any benefits to the retained group. Perhaps the advantage of maintaining a comparative level of achievement does not offset an additional year of schooling and the possible difficulties brought on by being placed with younger students.

There are indications that low achieving females stand to benefit the least from a retention experience in that their achievement levels seem to fall below both the retained males and passed females.

Recommendations Drawn
from the Research in
This Study

The processes of individual academic growth are apparently very complex processes. They may include factors relating to individual learning styles, specific emotional needs, value systems, school environments and varying delivery systems.

It was not the purpose of this study to examine each or any specific variable which may influence academic achievement. Instead assumptions were made relative to the samples which assumed that all students were receiving a typical elementary school education experience. Further, because of the sample size, it was assumed that variations in students' individual life space experiences would be averaged out and not warrant any special attention as they functioned within the sample.

The following recommendations are made in view of the research reported in this study and studies reported in Chapter II:

1. Retention, with the singular focus of raising individual achievement scores is not recommended, recognizing that individuals may have needs of other sorts which might warrant attention.
2. Retention consideration for females should be done with the understanding that such retention may tend to reduce their chances to achieve.
3. Retention, if employed, should be based on factors which are not primarily related to achievement.

Some of these factors may be:

- Physical maturity, including such considerations as physical size, projected growth, and psycho-motor development.
- Chronological age. Students who are having difficulty and are younger than their grade peers may benefit from a retention experience.
- Social Immaturity. This factor is difficult to assess. However, socially immature students may gain by placement with students who are chronologically younger.
- Developmental immaturity. Students who have difficulties in perceptual, conceptual or language development may benefit from retention. This might best be used as a consideration at the kindergarten and first grade levels only.

In view of the research and considering the recommendations stated above, an alternative structure to the current management system used in schools which would eliminate the grade level structure and the emphasis it places on unified or "lock-step" movement in achievement, would appear to be more effective in dealing with the low achiever. A continuous progress system which approaches individualized instruction through a series of achievement levels or objectives would appear to enable learning to be gained more consistently. Elimination of the grade level structure, which is largely maintained for the convenience of the institution and the creation of a curricular program which addresses the nature and needs of the

learner may be a more effective system of education.

These recommendations appear to be supported by findings reported in Chapter II in the works of Cook, Myers, and Klene and Bronson. Work done in Brenard County, Florida and reported by D. T. Sheurer¹ and others would also demonstrate that this approach may have some viability. Curriculum at the Melbourne High School in Brenard County was organized to reflect the needs of the learner through changes in approaches to instruction, instructional management systems, and assessment. The grade level structure was minimized and students were allowed to achieve according to their needs and in smaller learning segments. Although research on the success of this management system is not thorough there seem to be merits to its institution.

Educators are frequently under pressure to produce achievement statistics which show academic gains, and consequently, they may be concentrating their efforts on teaching children chiefly content to the relative neglect of individual developmental growth.

¹D. T. Sheurer, "The Placement of Students in Viable Learning Situations Through the Use of Achievement Tests and Systems Engineering Rather than Through Annual Promotion and Retention," (Washington, DC: USOE, ERIC Document Reproduction Service, ED 057 068, 057 067, 1970).

A continuous progress, developmental approach need not eliminate accountability measures, nor need it deemphasize productivity in the classroom. On the contrary, it could become even more important for educators to focus on those learnings they perceive as necessary and important for children's future success. It would appear desirable if educators might become even more concise in educational objectives regarding developmental behavior and children's social growth. Increased precision might subsequently allow children to grow more efficiently in academic achievement and at a pace fitted to their individual ability and background.

Questions for Future Research

There are several questions generated by this study which appear to deserve further research.

1) What are the implications of retention in more urban settings, given their more sophisticated diagnostic programs and additional curricular alternatives, both within the classroom and supplemental to the classroom?

Since the population sample of this study was drawn from primarily rural school systems which have a comparatively lower per pupil expenditure, and

therefore fewer services to assist in learning, such as intervention programs or alternative delivery systems, a replication of this study in a metropolitan area may produce different and additionally illuminating results.

2) Would different outcomes be observed if the study were replicated using a broader population base?

The essentially rural, middle to low income, ethnically white population sample of this study may not reflect results which might be obtained if the population sample were broader and included students who are from higher income families and/or more diverse ethnic backgrounds.

3) How would the achievement data of both low achieving groups in this study compare with those from regularly achieving students?

The design of this study did not include any comparison of low achievers with those termed regular achievers. Future studies may explore this relationship with the focus of determining what differences exist between groups over a similar educational time period.

4) What influence does a retention experience have on a child's affective development given similar time frames?

This study was limited to academic measures.

Conclusions relative to attitudes, social behavior, and social development cannot be drawn from this study. Future research may include instruments which would attempt to measure the long term effects retention has on specific affective areas of children's growth.

5) Can similar achievement patterns be observed in other curricular areas?

Reading and mathematics were the focus of this study. However, achievement in social studies, in science, and in curricular areas such as music, art, and physical education could be observed and compared to see if similar results are evident.

6) Will outcomes be comparable after a longer span of time?

7) What factors precipitate the disproportionate ratios of males and females in the passed and retained groups?

The educational institutions' response to maleness or femaleness may make males more liable to failure. Perhaps there is indeed a physiological and/or emotional difference between the sexes which allows girls to achieve more readily at this level and therefore not subjected to retention treatments.

Differences between passed and retained students' achievement at later testings may confirm or deny the

outcomes of this study. Perhaps an examination of achievement scores when the students are in junior high school or high school may produce different outcomes.

Summary

Chapter V has presented the conclusions of this study in view of given data analysis.

These conclusions are:

1. There is no statistical evidence that males and females function differently when given the treatment of retention.
2. There is no evidence that passed and retained students achieve differently in mathematics.
3. Reading achievement is statistically lower for retained students than for passed students.
4. There is no indication that retention benefits students in the curricular area of mathematics.
5. The results of this study indicate that retention may have an adverse effect on a students' ability to achieve in reading.

Also presented in Chapter V were some

recommendations for possible changes in curriculum delivery systems which may enable low achieving students to learn more efficiently.

Finally, several questions for future research on the topic of student retention were discussed.

APPENDICES

APPENDIX A

OBSERVED CELL STANDARD DEVIATIONS

Status and Sex	Third Grade		Fifth Grade	
	Total Reading	Word Analysis	Total Reading	Total Math
Passed, Male	.163215	.190685	.197428	.249934
Passed, Female	.150449	.179742	.169294	.210152
Retained, Male	.196451	.236406	.246820	.238598
Retained, Female	.199503	.255606	.239397	.279999

APPENDIX B
ERROR CORRELATION MATRIX

	Total Reading Third	Word Analysis Fifth	Total Reading Fifth	Total Math Fifth
Total Reading Third	1.000000			
Word Analysis Fifth	.517431	1.000000		
Total Reading Fifth	.599007	.882899	1.000000	
Total Math Fifth	.427538	.450178	.509808	1.000000

APPENDIX C
VARIANCE AND STANDARD DEVIATION

Variable	Error Mean Squares	Standard Deviation
Total Reading, Third	.027148	.1648
Word Analysis, Fifth	.038838	.1971
Total Reading, Fifth	.038151	.1953
Total Math, Fifth	.054290	.2330

APPENDIX D
MATRIX OF CORRELATIONS
ADJUSTED AGAINST COVARIATE
(Total Reading, Third Grade)

	Word Analysis	Total Reading	Total Math
Word Analysis	1.000000		
Total Reading	.836165	1.000000	
Total Mathematics	.295973	.350491	1.000000

APPENDIX E

F-STATISTIC FOR HYPOTHESES I, II, AND III

Hypothesis	Mean Square	Univariate F	P Less Than	Stepdown F	P Less Than
Hypothesis I	.2502	8.7623	.0034	8.7623	.0034
Hypothesis II	.3818	15.5432	.0002	6.8807	.0093
Hypothesis III	.2795	6.2739	.0130	1.3464	.2471

88

Degress of freedom for hypothesis = 1, Degrees of freedom for error = 244.

F-STATISTIC FOR GRAND MEAN

Degrees of freedom = 3 and 242

F Ratio = 5.7411

P Less than .0009

APPENDIX F

F-STATISTIC FOR SUB-HYPOTHESES A AND B

Sub-hypothesis A (Sex interaction)	Degrees of freedom=3 and 242 F ratio=.6286 P less than .5972
Sub-hypothesis B (Sex differences)	Degrees of freedom=3 and 242 F ratio = 1.2407 P less than 12956

REFERENCES

REFERENCES

A. BOOKS

- Finn, Jeremy D. A General Model for Multivariate Analysis. New York: Holt Rinehart & Winston, 1974.
- Harcourt Brace Jovanovich, Inc. Manual for Interpreting Metropolitan Achievement Tests. New York: Harcourt, Brace, Jovanovich, 1962.
- Lessinger, Leon M. Every Kid a Winner. New York: Simon and Shuster, 1970.

B. JOURNALS AND PERIODICALS

- Anfinson, R. D. "School Progress and Pupil Adjustments," Elementary School Journal 41 (March 1941):507-14.
- Caplan, Paula J. "Role of Classroom Conduct in Promotion and Retention of Elementary School Children," Journal of Experimental Education 41 (Spring 1973):8-11.
- Coeffield, W. H. "Effects of Non-promotion on Educational Achievement in Elementary Schools," Journal of Educational Psychology, 47 (1956):235-50.
- Chansky, N. M. "Progress of Promoted and Repeated First Failures," Journal of Experimental Education, 32 (1964):225-37.
- Chase, J. A. "A Study of the Impact of Grade Retention on Primary School Children," Journal of Psychology, 70 (1968):169-77.
- Cook, W. "Some Effects of the Maintenance of High Standards of Promotion," Elementary School Journal, 41 (1941):430-37.
- Dittmer, A. "Assault With a Deadly Mandate," Media and Method, 15 (September 1979):26-9.
- Done, B. "A Call for Public Accountability, Great Britain," Times (London), Educational Supplement, 11 April, 1980.

- Francis, E. B. "A Follow Up of Non-promotion," Journal of Education, 122 (June 1939):187-88.
- Gaite, A. J. H. "On the Validity of Non-promotion as an Educational Procedure," University of Wisconsin, 1969.
- Goodlad, J. I. "Some Effects of Promotion and Non-promotion Upon Social and Personal Adjustment of Children," Journal of Experimental Education, 23 (June 1954):301-28.
- Hall, W. F. and Demarest, Ruth. "Effect on Achievement Scores of a Change in Promotional Policy," Elementary School Journal, 58 (1958):204-07.
- Jackson, Gregg B. "The Research Evidence on the Effects of Grade Retention," Review of Educational Research, 45 No. 4.
- Keyes, C. H. "Progress Through the Grades of City Schools," Bureau of Publications, Teachers College, New York, Columbia University.
- Klene, Vivian and Bronson, E. P. "Trial Promotion vs. Failure," Elementary School Journal, 29 (1929):564-66.
- Lawrence, Ben. "Accountability Information: Progress, Pitfalls and Potential," New Directions for Higher Education, 20 (1979).
- McElwee, E. W. "A Comparison of Personality Traits of 300 Accelerated, Normal and Retarded Children," Journal of Educational Research, 26 (1932):31-4.
- Morrison, Ida and Perry, Ida F. "Acceptance of Overage Children by Their Classmates," Elementary School Journal, 56 (1956):217-20.
- Myers, F. "We Experiment with a Non-failure Program," Childhood Education, 18 (1942):205-09.
- Pai, Y. and Krueger, J. "Accountability What Price?" Journal of Teacher Education, 22 (June-August 1979):20-21.
- Phi Delta Kappan 59 (March 1979). Total issue on topic of accountability.

- Rernherz, H. and Griffin, C. L. "The Second Time Around," School Counselor, 17 (1970):213-18.
- Sandberg, J. H. "K-12 Competency Based Education Comes to Pennsylvania," Phi Delta Kappan, 61 (October 1979):119-20.
- Sandin, A. A. "Social and Emotional Adjustments of Regularly Promoted and Non-promoted Children," Child Development Monographs, 32 (1944).
- Scott, B. A. and Ames, J. B. "Improved Academic, Personal and Social Adjustment in Selected Primary School Repeaters," Elementary School Journal, 69 (1969):431-39.
- Shailer, Thomas and Knudsen, Dean D. "Relationship Between Non-promotion and the Drop Out Problem," Theory Into Practice, 4 (1965):90-94.
- Templin, R. S. W. "A Check Upon Non-promotion," Journal of Education, 123 (November 1940):259-60.
- White, K. and Howard, J. L. "Failure to be Promoted and Self Concept Among Elementary School Children," Elementary School Guidance and Counseling, 7 (1973):182-87.
- Worth, W. H. "Promotion vs. Non-promotion: The Earlier Research Evidence," Alberta Journal of Educational Research, 5 (1959):77-86.
- Worth, W. H. "What Research Says About Promotions," Canadian Education, 15 (1960):61-70.
- Worth, W. H. "Does Non-promotion Improve Achievement in the Language Arts?", Elementary English, 37 (1960):49-52.

C. REPORTS, YEARBOOKS AND SPECIAL WORKS

- Boros, W. Eighth Mental Measurements Yearbook, Vol. II, New York: Gryphon Press, 1978.
- Finn, Jeremy D. Multivariate: Univariate and Multivariate Analysis of Variance, Covariance and Regression. Chicago: National Educational Resources, Inc., 1972.

Harcourt, Brace and Jovanovich, Metropolitan Achievement Tests, Special Report Number 10--Reliability Estimates and Standard Errors of Measurement. New York: Harcourt, Brace and Jovanovich, June 1971.

Harcourt, Brace and Jovanovich, Inc. Metropolitan Achievement Test Special Report Number 1--Content Development. New York: Harcourt, Brace and Jovanovich, Inc., 1971.

School District of Philadelphia, "The Promotion/Retention Dilemma, What Research Tells Us," Office of Research and Evaluation, December 1973.

Sheurer, D. T. The Placement of Students in Viable Learning Situations Through the Use of Achievement Tests and Systems Engineering Rather than Through Annual Promotion and Retention. Bethesda, MD: ERIC Document Reproduction Service, ED 057 067, ED 057 068.

D. UNPUBLISHED DISSERTATIONS

Briggs, L. D. "The Impact of Failure on Elementary School Pupils", Doctoral dissertation. North Texas State University, 1966. Dissertation Abstracts, 27 (1967):2719-A.

Coeffield, W. H. A. "A Longitudinal Study of the Effects of Non-promotion on Educational Achievement in Elementary School," Doctoral dissertation. University of Iowa, 1954. Dissertation Abstracts, (1954):2291.

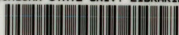
Ogden, K. W. "An Evaluation of Non-promotion as a Method for Improving Academic Performance," Doctoral dissertation. University of Southern California, 1971. Dissertation Abstracts, 32 (1971):795-A.

Worth, Walter. "The Effects of Promotion and Non-promotion on Pupil Achievement and Social-Personal Development in the Elementary School," Doctoral dissertation. University of Illinois, 1959. Dissertation Abstracts 20 (1959):3228.

E. GENERAL REFERENCES

- Bulmer, M. G. Principles of Statistics. New York: Dover Publications, 1967.
- Elzey, Freeman F. A Programmed Introduction to Statistics. Belmont, California: Brooks/Cole Publishing Co., 1965.
- Kowitz, G. J. "The Effects of a Promotional Policy on Academic Achievement," Elementary School Journal, 61 (1961):435-43.
- Sackett, E. B. "Record of 25 Social Promotions," Elementary School Journal, 36 (1936):776-80.

MICHIGAN STATE UNIV. LIBRARIES



31293010731838