

INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.
4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.
5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

REEVES, Ronald Roland, 1941-
THE EFFECT OF A SPECIAL PROGRAM ON THE
READING AND ARITHMETIC ACHIEVEMENT OF
STUDENTS SELECTED FOR PARTICIPATION.

The University of Oklahoma, Ph.D., 1974
Education, guidance and counseling

Xerox University Microfilms, Ann Arbor, Michigan 48106

© 1974

RONALD ROLAND REEVES

ALL RIGHTS RESERVED

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED.

THE UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

THE EFFECT OF A SPECIAL PROGRAM ON THE READING
AND ARITHMETIC ACHIEVEMENT OF STUDENTS
SELECTED FOR PARTICIPATION

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
degree of
DOCTOR OF PHILOSOPHY

BY
RONALD ROLAND REEVES

Norman, Oklahoma

1974

THE EFFECT OF A SPECIAL PROGRAM ON THE READING
AND ARITHMETIC ACHIEVEMENT OF STUDENTS
SELECTED FOR PARTICIPATION

APPROVED BY

Gerald K. Kautz
H. Wright James
W. K. Korte
Gene P. Pugh

DISSERTATION COMMITTEE

ACKNOWLEDGMENTS

It is not possible to name all the many people who have given encouragement and advice throughout my educational experiences. However, my chairman, Gerald T. Kowitz, has been exceptionally helpful with his advice. Special thanks also goes to William H. Graves of my committee for the time he has devoted to the progress of this study. To the other members of my committee, Wayne Rowe and G. Gene Pingleton, my thanks for your willingness to serve on my committee and for the advice and assistance you have given.

My appreciation is also expressed to the school district which allowed me to use the data from the program for my research. Although they desire to remain anonymous, they have helped immensely. Special thanks goes to the Superintendent, the Coordinator of Elementary Education, the Director of Elementary Guidance Programs and the Elementary Counselors.

Another person has been instrumental in my endeavors to complete the research--my wife, Molly. She has been understanding and has given much encouragement throughout the research effort. In addition, she has ably typed the manuscript, all the time withstanding the pressures.

To the many others, thank you for your encouragement and support.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	iii
LIST OF TABLES	v
CHAPTER	
I. INTRODUCTION	1
Special Program	3
Purpose of the Study	5
Statement of the Problem	6
Definition of Terms	6
Hypotheses	6
Limitations of the Study	7
II. REVIEW OF THE LITERATURE	9
Program Evaluations	10
Instruments Used in the Study	18
Summary	24
III. METHODOLOGY	25
Subjects	25
Schedule of Data Collection	26
Instruments	28
Statistical Design	34
IV. DATA ANALYSIS AND RESULTS	41
Testing of the Hypotheses	44
V. CONCLUSIONS AND RECOMMENDATIONS	55
Conclusions	55
Recommendations for Research	62
REFERENCES	64
APPENDICES	70

LIST OF TABLES

Table		Page
1.	Means and Standard Deviations for Groups on Full Scale Score of the VKT	41
2.	Analysis of Variance for Full Scale Scores of VKT	42
3.	Mean Ages (in Months) of Subjects in the Three Groups	43
4.	Summary Table for Analysis of Variance of Age (in Months) at Time of Last Test	43
5.	Summary Table for Chi Square	44
6.	Univariate F Tests of the MRT Variables	45
7.	Means and Standard Deviations of the MRT Variables	46
8.	Univariate F Tests Conducted on VKT Subscale Scores	48
9.	Results of the Univariate F Tests Among Groups on the Reading and Arithmetic Scores of the Stanford Achievement Tests	49
10.	Means and Standard Deviations of SAT Variables	49
11.	Results of the Multivariate Analysis of Covariance for Groups 1 and 2 on the Stanford Achievement Test Variables	52
12.	Analysis of Covariance on SAT II Reading Scores Adjusted by SAT I Reading Scores	53
13.	Analysis of Covariance on SAT II Arithmetic Scores Adjusted by SAT I Arithmetic Scores	53

THE EFFECT OF A SPECIAL PROGRAM ON THE READING
AND ARITHMETIC ACHIEVEMENT OF STUDENTS
SELECTED FOR PARTICIPATION

CHAPTER I

INTRODUCTION

The prevention of situations that result in students leaving the educational system prematurely is one of the most pressing problems educators face today. Two major considerations confront a school district in establishing a program to combat the drop-out problem--considerations which have proven to be sources for dispute in past programs and studies. One of the considerations is early identification of children who will have difficulty in school (Smith, 1966; Haring and Ridgeway, 1967; Webb and Pate, 1970; Weininger, 1972). The other consideration is the development of a program that is both feasible within the existing structure of the school system, and can be conducted at an effective time in the child's development.

In an effort to discover how a student progresses, and by implication where difficulty begins, Anderson and Maier (1963) initiated a study designed to study achievement and intellectual functioning of students in grades five through twelve in hopes of discovering keys to the drop-out problem.

Shortly after the study was begun, people at the Educational Testing Service (E.T.S.) questioned the study in the belief that some of their findings might be wrong or incomplete without knowledge of the years prior to the fifth grade. However, at that time, on the national level, there was not a commitment to early education, nor was there a satisfactory judgment about the adequacy of measurements with younger children (E.T.S., 1969).

The emphasis shifted from the adequacy of measurement to program development in 1965 when Head Start was initiated.

. . . Head Start was predicated upon the intuitively compelling assumptions that the earliest years are generally the most important in shaping people's lives and that an important segment of the population is likely to have "inadequate" early shaping experience (E.T.S., 1969, p. 8).

Head Start provided experiences for the economically and culturally deprived segment of the population. But what of those children who were not eligible for participation in Head Start because of the family's economic status but who still suffered from environmental deprivation due to a lack of stimulation, who lived in a community which had no Head Start program, or who had a learning disability? These children have since been identified as needing assistance.

Initially established as a short-term program, Head Start was extended to a full-year effort with stimulation being given at the same time as for children who were not deprived and not included in the Head Start program. Upon completion

of Head Start the children were placed in a regular classroom. Placement in regular classrooms could create difficulties such as different educational approaches and larger class size. To avoid these difficulties, Reidford and Berzonsky (1967) recommended that short-term programs be replaced by "long-term programs which extend their methodologies up through the early elementary grades (p. 7)." Grotberg (1971) also called for an extension to provide for continuity.

Several authors (Smith, 1966; Haring and Ridgeway, 1967; Weininger, 1972) indicated that the important part of helping children with difficulties is designing a program to help them overcome whatever is interfering with their learning. The program may, of necessity, be started after the pupils enter school. In fact, from the cognitive-developmental position Kohlberg (1968) argued that

. . . certain cognitive-enrichment programs should be timed later for culturally disadvantaged children because of cognitive retardation, rather than attempting to provide enrichment programs for these children at the age at which more advanced middle-class children are presumed to be receiving parallel stimulation (p. 1048).

Special Program

During the 1969-70 school year a program was initiated for the purpose of combating school drop-outs by a school district in the metropolitan Oklahoma City area. It was an effort to provide enrichment to the students through a special approach within the existing structure. Eventually the techniques and

methodologies which proved to be effective would be moved to the regular classrooms of the system. Thus an opportunity was provided to evaluate the effects of an in-school program over a period of two years to determine if appreciable gains had been made by the students who participated.

Description of the Program

The program under study had an overall objective of early identification of academic, emotional and social needs of students and subsequent treatment of the students identified as needing assistance. The ultimate goal of the program was to help the students before they became so handicapped that they could not function in school. A primary method of dealing with a student's needs used individualized instruction prescribed by the project staff after an evaluation of each student. The program was also designed to assist the regular classroom teacher in gaining a better understanding of the students and providing them with approaches which allowed the children in the program to succeed during the period they were in the regular classroom. The project staff assisted the regular classroom teacher through discussion of the child as well as through an exchange of teaching ideas and materials.

Assistance was given to participating students in three developmental areas: academic, emotional and social. The academic help was provided in reading and arithmetic by teachers specialized in these areas. The emotional assistance

was carried out by a counselor in the program and also by each teacher who had contact with the student. The social help was given throughout the program, but was primarily the responsibility of the teacher in the Human Relations Division, who conducted discussions designed to enable the students to function more adequately with their peers, teachers and family.

The program differed from the regular classroom in two major areas. First, the program was conducted in what the project staff terms a modified clinical approach. Thus, the members of the team had the capabilities of diagnosing the children and teaching to their weaknesses. The process was repeated until the student reached the level of expected functioning for his age-grade placement. Secondly, the students were primarily given individual instruction which was designed for them after the diagnostic stage. The individual instruction was augmented by the use of small groups for instruction if the children were identified as having similar needs.

Purpose of the Study

The present study was designed to examine the effects of an educational program developed to offer stimulation to students identified as having a problem which could prevent them from learning in the regular classroom. The study is an evaluation limited to the measurement of academic

achievement of the students selected for participation. The present study should contribute to the overall evaluation of effectiveness of the treatment provided the participants.

Statement of the Problem

The study was undertaken in order to evaluate the effect of the program on reading and arithmetic achievement, with emphasis on the effect as measured by achievement tests in the two areas. Groups having differing lengths of time in the program were compared to estimate the effect of longer treatment. A comparison of the groups was made using the screening instruments to determine initial differences between the groups.

Definition of Terms

The comparison groups used in the study were:

Group 1.--Students randomly selected from those children who were screened but were not selected for program participation because of space limitations.

Group 2.--Students who participated in the program for one year or less.

Group 3.--Students who participated in the program for more than one year.

Hypotheses

The following hypotheses were developed in order to explore the effect of the program.

Hypothesis 1.--There was a significant difference among groups prior to starting the program on the subtests measured by the Metropolitan Readiness

Tests; the students selected for participation scored lower on all subtests than those eligible but not selected.

Hypothesis 2.--There was a significant difference between the two groups selected for the program on the Vocabulary and Perceptual-Motor subscales of the Vane Kindergarten Test; the students who remained in the program beyond an academic year were in the group scoring lower on the Perceptual-Motor subscale.

Hypothesis 3.--The students who returned to the regular classroom during or at the end of the first year showed higher achievement than the other two groups.

Hypothesis 4.--When the effects for previous achievement and intellectual functioning are taken into consideration, the students who remained in the program longer than one year showed higher achievement than those who returned to the regular classroom during or at the end of the first grade.

Hypothesis 5.--Groups 1 and 3 showed higher achievement than Group 2 at the third grade test time after the scores were adjusted for second grade achievement.

Limitations of the Study

Since the study was designed after the program had been in effect for a period of two years, it was in nature an ex post facto study--with all the ramifications inherent in such. The major limitation of an ex post facto study is the lack of controls on the independent variables. In this study, the criteria used in the initial selection procedure were not firmly established.

Second, the subjects for the study could not be assigned randomly for the groups. In most school programs administrators are not willing to use random assignment of

students to experimental programs because of the risks involved, either real or imaginary, such as angering school patrons. The lack of randomization for experimental programs has created much discussion over the evaluation of groups not randomly selected (Campbell and Erlebacher, 1970a; Cicirelli, 1970; Evans and Schiller, 1970). Their discussion centers around whether the sophisticated analysis techniques adequately control variables or cause a bias in the results. Campbell and Erlebacher (1970b) recommend that "no scientific evaluation be done at all in ex post facto situations and in other situations where none (scientific evaluation) is possible (p. 224)." Evans and Schiller (1970) point out the political difficulties with randomization in program application. Although the argument set forth by Campbell and Erlebacher is the ideal, it appears that much information will be lost while the public education called for is being conducted. In addition, some understanding seems necessary regarding the operational effects of programs being conducted.

Another limitation was the lack of control for differences in program application among classroom teachers. The high mobility of students from school to school in the district dictated the lack of control of this variable.

The limitations may deny conclusions either in defense of or in opposition to the program. However, the study can give indications of the operational effect of the program as measured by the achievement of participating students.

CHAPTER II

REVIEW OF THE LITERATURE

For a little over a decade there has been much debate over whether the readiness factor for learning could be, and even whether it should be, manipulated (Weininger, 1972). Several programs have been specifically designed to manipulate the factor, an example being the national Head Start effort. At times much controversy accompanied the efforts. These programs have been designated as compensatory.

The program under study is compensatory. It was based on the idea that if students who were expected to have difficulty in school were provided compensatory instruction they would be able to progress satisfactorily through the remainder of their educational careers.

At the present time there are three types of early compensatory education which are discussed. The review which follows uses the three categories as organizational guidelines; however, it does not follow the chronological order in which the approaches were developed. Also included in this chapter is a brief review of the research conducted on the instruments which were used in the study for evaluation of the students.

Program Evaluations

The three types of compensatory education used for organizational purposes are: (1) the National Head Start Program; (2) compensatory programs started through local initiative at the preschool level; and (3) programs begun in the primary grades of the regular school program. In actuality, types one and two developed in reverse chronological order.

Head Start Programs

Most of the evaluation studies on the effects of stimulation for learning have been conducted at the preschool level. The majority were designed to evaluate Head Start. Anderson and Temp (1967) showed that children involved in Head Start showed "some gains in intelligence test scores" over a short period of time. At the same time they discovered that there were wide variations in programs studied. They could not tell what the long-term effects would be. Another study (Reidford and Berzonsky, 1967) showed that the Berciter-Englemann Preschool Curriculum, used over a long term, increased intelligence quotient levels as measured by the Stanford-Binet. They also found that it "stimulated development in reasoning ability, language facility, and understanding (pp. 6-7)."

The Head Start research through 1969 has been reviewed by Datta (1969). She reported that many of the studies

on the summer program found significant improvement in the general ability as measured, although the national norms were not reached. On the other hand, the research on full-year programs showed that the scores reached the national average. These studies indicated that the level of achievement was a function of the length of time in the program.

In the same report, Datta reviewed studies which showed that the rate of development found in the measurements were not sustained after the children entered the primary grades of school. She stated, "What appears to happen is that the rate slows down for the Head Start children while their non-Head Start counterparts sooner or later catch up (p. 13)." This catching-up appears to be accomplished by the end of the first year in school, whether that is kindergarten or first grade.

At the time of Datta's review, additional research on Head Start was continuing. One study (Moore and Ogletree, 1973) examined the readiness level of Head Start and non-Head Start children. The results showed that Head Start children had higher readiness scores.

Another study that includes many parts and actually began before Datta's review, is continuing. This is a longitudinal study designed by Educational Testing Service (E.T.S.) and Head Start (E.T.S., 1968; 1969; 1970; 1973). However, the study is still in the analysis stage and few results are available.

Other Preschool Programs

Although Head Start has received much of the national attention, there were programs in operation several years prior to 1965. In 1959, a study was initiated based upon "providing a special ten-week intervention program for deprived children just before they entered first grade (Klaus and Gray, 1969, p. 1)." The results from the evaluation of the program showed that the experimental children gained eight points in Stanford-Binet IQ while the control group gained three points. The children were also tested on the Peabody Picture Vocabulary Test with similar results. The children were followed through to the end of the first grade when they were given the Metropolitan Achievement Tests. The results indicated that no significant differences existed between the two groups. Klaus and Gray stated that despite attempts to make the groups similar the control groups had been significantly higher than the experimental group on the Stanford-Binet pretest. This fact would partially obscure the effects of the program. However, "the pilot study certainly failed to demonstrate any positive results at the end of the first year (p. 2)."

The 1959 pilot study led to the planning for the Early Training Project which was reported in detail by Klaus and Gray (1968). This project was conducted over a five-year period which began in 1961. The Early Training Project

was designed to provide ten-week summer sessions for two groups of children and follow-up visits to the homes during the remainder of the year. The two groups were involved in the program for three years and two years. Two control groups were also used in order to measure the effect of the program on the siblings of participating children.

Studies of the Early Training Project showed that the children who participated in the program performed better at entrance into public school than did the control groups. In addition, Klaus and Gray stated that their findings

. . . substantiate a positive effect across several measures upon the performance of the experimental children, an effect that was sustained through the last assessment period, at the end of the second year of public schooling (p. 52).

In addition it was found that there was a diffusion effect on the siblings. The younger siblings of the experimental children showed higher performance than the younger siblings of children not in the program.

Before the report on the Early Training Project was published, Herbert Sprigle started a program in Jacksonville, Florida, called "Learning to Learn," evaluated and reported by Van De Reit and Van De Riet (1967). The program was designed around a developmental sequence of growth and development. It concentrated on manipulating, organizing, classifying and ordering materials so that the children would be led to internalized thought and effective verbal expression.

Van De Riet and Van De Riet found that the experimental group was superior to other groups on developmental measures at the end of the nine-month program. They also found that the experimental group's superiority remained at the end of the first grade; a finding which is different from most follow-up studies. When they compared the traditionally trained group and the no-training group they found that the traditional group was higher at the end of the first nine months. However, at the end of the first grade there was little difference. Overall, the researchers found that the groups had moved closer together at the end of the first grade primarily because the no-training group had improved more than the other groups, with the experimental group improving the least. Grotberg (1971), in reviewing the program, made the point that the "wash-out effect" might be the result of improved learning for the no-training group due to the interaction with the children who were participants in the program.

Another program similar to Head Start was conducted and reported by Turner and DeFord (1970). It was a follow-up study of the Early Childhood Education Project (ECEP) in Richmond, Virginia. The investigators compared three groups of children: (1) a random sample of children who had no organized preschool experience prior to entering kindergarten; (2) pupils who participated in the regular ECEP session; and

(3) pupils who participated in both the ECEP and the Summer Head Start programs.

The results of the study indicated that

. . . The longer the duration of the preschool educational experience, the higher the scores on the Metropolitan Achievement Test, reading and arithmetical concepts and skills gaining most. Results of both standardized tests and teacher ratings indicated that preschool experiences improved academic readiness and performance (p. 7).

Turner and DeFord's results are comparable to most studies conducted on compensatory education programs. However, the study did not extend the evaluation over a long enough period of time to examine the stability of the changes indicated by the evaluation.

Data are also available on the Infant Educational Research Project conducted in Washington, D. C. After a period of tutoring, the intelligence level of a group of Negro boys was raised significantly when compared to a control group. However, there was a decrease from the peak IQ level one year after the tutoring was ended (Yahreas, 1973).

In-School Programs

Additional information regarding long-term effects of special programs is available from reading education. One of the more extensive coverages of a follow-up approach was presented by Newman (1972). Newman's study was concerned with the effect on later learning of a differential instruction program for low-reading, first-grade pupils. Newman

conducted the follow-up after the pupils had reached the fifth or sixth grade. The original study was conducted in Cedar Rapids, Iowa, and consisted of seven groups, which included 309 pupils. Newman was able to use 230 students in her follow-up study.

The results of Newman's study showed no significant differences among the treatment groups at either the first- or sixth-grade levels. She did find that first-grade reading achievement test results were a strong predictor of later achievement and were also more reliable than the measures of first-grade readiness. No clear decision could be made regarding the correlational structure of achievement-related variables.

Also in Newman's report, a question was raised concerning the lasting effect of a special program. However, she stated that a much more profitable use of differential instruction might be made if it were possible to look at children with a multivariate approach and that until this is done any grouping should be flexible in the initial grades.

There have been other, more broadly based, programs designed for students after they have entered school. One such program is being conducted in Hackensack, New Jersey, and is known as Project LEM (Learning Experience Module). The program incorporates several features found in elementary education into one program. Among these features are: multi-age and ethnic grouping; individually prescribed

instruction; differentiated staffing; and a core curriculum. The program is designed to meet several specific objectives. Among these is the improvement of "the reading and arithmetic skills of students (Hackensack Public Schools, 1972, p. 3)." According to the school system report the program has been successful.

Another study pertaining to in-school programs was conducted by Buckland and Balow (1973). It was primarily interested in the effect of specific materials developed for visual perceptual training, particularly as the training effected the low-achieving first-grade pupil's "readiness and word recognition skills (p. 299)." Two groups were given equal amounts of time by the teachers, with an experimental group using a workbook while a control group listened to and discussed stories. The results of the study failed to show statistically significant differences between the two groups or between pre- and post-testing.

Most of the programs mentioned thus far have been designed for compensatory purposes. However, there is a report by Hoffman (1962) which concerns itself with enrichment for the highly intelligent child. The purpose of the study was to show that the highly intelligent child would benefit from a special program within a small group arrangement. The evaluation was conducted on the student's evaluative ability and his ability to make generalizations and to draw conclusions from them. The results of the evaluation showed that

after one year there were significant differences. However, at the end of the second and third years no differences were observed.

Instruments Used in the Study

The following section reviews a sample of the research which has been conducted using the various instruments utilized in the present study. The review follows the order in which the instruments were given to the subjects in the study.

Metropolitan Readiness Tests

The Metropolitan Readiness Tests (MRT) have been used for many research studies. Several of these studies have been reported by Hildreth, Griffiths and McGauvran (1969). They used the tests to establish research groups for further research and also for checking the performance of the readiness tests themselves.

The use of the MRT in studies of other instruments has continued since 1969. Proger, McGowan, Bayuk, Mann, Trevorrow and Massa (1971) utilized the MRT in their study on the predictive and construct validities of the Otis-Lennon Mental Ability Test, the Lorge-Thorndike Intelligence Test, and the Metropolitan Readiness Tests. They found that the Otis-Lennon predicted achievement as well as the Lorge-Thorndike or Metropolitan Readiness Tests.

In a study conducted by Goolsby and Frary (1970), the predictive validity of the MRT for white and Negro students was investigated. They reported "no reason to question the validity of the MRT for Negro students . . . (p. 447)." On the other hand they questioned the validity of the tests for the white students because of the inaccuracy of prediction for later achievement.

Another study which investigated the predictive validity of the MRT has been reported by Lessler and Bridges (1973). They concluded that the MRT alone predicted achievement in the first and second grades as well as several other tests used in combination.

In research reported by Lederman and Blair (1972), the MRT was used in the validation of another instrument. They were interested in the validity of ratings obtained from teachers and mothers on the Preschool Attainment Record. They correlated the ratings with the MRT and discovered that the teachers' ratings were "fairly good" while those ratings obtained from the mothers were biased in the direction of higher ratings.

Vane Kindergarten Test

Researchers have used the Vane Kindergarten Test (VKT) (Vane, 1968) for a few studies. D'Angelo, Walsh, and Lomangino (1971) attempted to extend the norms to include Negro children and to test the advisability of lowering the

bottom age to four and one-half years. The results of their study indicated that it was not advisable to change the lower age limits. However, they found that the girls showed higher scores than the boys on all the subtests except the Vocabulary subtest.

In another study, Walsh and D'Angelo (1971) evaluated the use of the VKT with Puerto Rican children. They found that the Puerto Rican children generally scored significantly higher on the Perceptual-Motor (P-M) and the Draw-a-Man (Man) subtests of the VKT, and showed no significant difference on the Full Scale IQ score. Unlike the study using Negro children, this study found no significant differences between boys and girls.

McKnab and Fine (1972) in their study found that the VKT was probably not an appropriate instrument to be used with an above-average sample of children. Their results indicated that the Vocabulary subtest did not offer enough items of varying difficulty to adequately measure the above-average student. The VKT could be limited if there are a large number of above-average students.

Another limiting factor on the VKT is the effect of a bilingual background on the Vocabulary subtest found by Walsh, D'Angelo and Lomangino (1971).

However, Willis (1970) found that the VKT discriminated between SES levels with middle-class children scoring better than lower-class students. She also concluded that

the VKT would be an effective screening test which could reduce considerably the testing time compared to the administration of other tests.

It therefore appears that the VKT is an appropriate instrument for screening children from a mono-lingual culture in the average or below-average group. The sample for the present study satisfies these requirements.

Stanford Achievement Tests

The Stanford Achievement Tests (SAT) (Kelley, Madden, Gardner, Rudman, 1964) have been in existence since 1923. After fifty years the tests have been used by many schools to evaluate the progress of students. They have also been used for many research studies.

The SAT has been utilized as a criterion measure in a study predicting achievement based on IQ scores (Mitchell, 1963). The author found that the predictions of achievement from IQ were not significantly different when the regression equation was based on individual or on school means. With the finding the author had developed expectancy tables for the tests, thus enabling schools to make more accurate predictions for program planning.

Studies have also been conducted to compare achievement scores on the SAT with measures of intellectual functioning. Merenda, Novack and Bonaventure (1972) compared the SAT and the California Test of Mental Maturity. After

analyzing the data on the two tests for primary-grade children, the authors concluded that there was negligible overlap in attributes measured by the two tests. Apparently, the two instruments can be used together for student evaluation in order to obtain a more accurate understanding of pupil functioning.

A study closely related to those comparing achievement with intellectual functioning was reported by Scott (1963). The study was designed to see if there was a relationship between either intelligence or gain in reading and gains on the subtests of the SAT. The author concluded that the relationship between intelligence and gains on the subtests of the SAT is difficult to determine, since many pupils of lower capacity achieved greater gains than pupils with higher capacity.

In a study designed to examine lower-class black and white children's achievement, Musgrove (1972) used the SAT with second-grade pupils. He was specifically interested in whether race or sex created significant differences on the achievement test. The findings indicated that significant differences occurred between black and white boys in all six areas, with the white boys scoring significantly higher. Scores on only two of the tests showed significance when black and white girls were compared. Girls scored higher on all tests than did the boys.

Ashcraft (1971) reported an interesting use of the SAT, using it to evaluate the effect of clinical treatment for emotional handicaps on school achievement. The study was conducted over a five-year follow-up period. The subjects were all diagnosed in the clinical setting as emotionally handicapped. The experimental group had been treated for at least six months while the control group did not return for treatment. The results of this study failed to show significant differences between the two groups at the end of the five-year period. However, the experimental subjects did show gains during the first two years but then dropped off during the next three years. At the end of five years, the experimental subjects were performing at a level below that of the control subjects. Another finding was that neither group achieved the expected nine-month gain on the SAT.

The SAT has been used as a covariate to adjust for achievement. Askow and Fischbach (1973) utilized the SAT in their effort to determine attitude toward reading. In their study they were also looking at a new instrument for measuring attitude toward reading.

Finally, Jacobs (1968) used the SAT in his evaluation of the Frostig Visual Perceptual Training Program. The results indicated that the Frostig Program does not guarantee higher reading achievement.

Summary

Several conclusions are suggested by the literature review pertaining to programs and their evaluation. These conclusions are:

1. Gains in functioning can be obtained through compensatory educational programs.
2. The amount of gain is dependent upon the length of time the children participate in the program.
3. The gains observed in children who participated in the programs are less visible relative to the children not in the programs in the years following participation. It is speculated that the loss is a matter of discrepancy in measurement. It is more a reflection of increased learning by those children who have not participated in the programs either through training in school or through interaction with children who have been in the experimental groups.

The type of intervention program to be employed remains a question. It may be that one type will be helpful with certain children while it will not be useful with other. Continued study and experimentation will be necessary.

The literature indicates that there are enough unanswered questions that evaluation studies should continue. Therefore, the study reported in the following chapters should add to our knowledge of programs being made available in our educational system.

CHAPTER III

METHODOLOGY

The effects of a special academic program were examined in the present study by comparing students who had participated in the program for varying lengths of time and by comparing the participants with non-participants. The comparisons were made using information collected on the students prior to, during, and for those still attending the schools, after they had left the program. Using multivariate statistical designs, the effects of the program as measured by achievement scores were examined to determine if the hypothesized differences existed.

Subjects

The subjects for study were selected from those students in one large school system in the metropolitan Oklahoma City area. The pupils were enrolled and attending school in the district. The majority had attended school in the district since they began kindergarten during the academic year 1970-71; however, a few had transferred in and out at the beginning of the first grade.

During the 1970-71 school year there were 711 children enrolled in the kindergarten classes with 31 students transferring

into the district the next year. Of the 742 students, 215 scored below the fiftieth percentile of the national norms on the Metropolitan Readiness Tests. From this group, 83 students were identified as in need of the assistance available through the program and began the program in the first semester of their first-grade year. At the time of the last test, 53 of the students were still in the school district, with 27 of them being tested in the third grade. Of those not tested 24 had been retained in the first or second grade and 2 were unavailable for testing. A random sample of 47 was taken from the remaining students who had scored below the fiftieth percentile on the MRT to be used as a comparison group. The n's of the groups were therefore as follows:

Group 1--non-participants	n = 47
Group 2--one year or less	n = 23
Group 3--over one year	n = 30
Total N	= 100

Schedule of Data Collection

The students were selected to participate in the program on the basis of screening in April and May of 1971. In April, 1971, the Metropolitan Readiness Tests, Form A (Hildreth, et al., 1969) were administered to the entire kindergarten population of the school district. Those pupils whose overall readiness score was at or below the fiftieth percentile, according to national norms, were selected for further evaluation to decide whether placement in the program would be appropriate.

An additional evaluation was conducted in May, 1971. The pupils were administered the Vane Kindergarten Test (Vane, 1968) by the school counselors. Although the VKT is designed for group administration the personnel involved in planning the program decided to give the test to the students on an individual basis. The effect of revising the test administration was not known. However, in most instances the existing norms did not apply. The counselors were thus able to observe the children to evaluate their approach to the test. The procedure was used to allow for differences in test-taking ability.

Using the results on the VKT, final decisions were made on which children would participate. The selection was based partially on the test scores; however, there was not a fixed criterion for selection. The kindergarten teachers also made recommendations on children with whom they had worked and whom they felt would have extreme difficulty in the regular first-grade classes, allowing for additional flexibility. They were also able to place students in the program who were judged by the counselors and teachers to need social or emotional help. No estimate is available regarding the comparative accuracy between the teachers and the test results.

In the first semester, 1971, the students selected were given the Stanford Early School Achievement Test, Level 1 (SESAT I) (Madden and Gardner, 1969). The results of the SESAT I, combined with the results of the screening instruments

mentioned above, enabled the staff to decide what special help each child should receive while in the program. During the time of participation the students were evaluated periodically by the staff. When the staff believed that a child was ready to leave the program another achievement test was given. If the student achieved at grade level when his scores were compared to the norms provided in the test manual, he was returned full-time to a regular classroom.

The school had a district-wide testing program at the second-grade level. Therefore, the students who participated in the program were tested in the fall of 1972 when they were in the second grade using the Stanford Achievement Test, Primary I Battery, Form W (Kelley, et al., 1964a).

In November, 1973, those students who had participated in the program and who were still enrolled in the school district were again tested. They were given the Stanford Achievement Test, Primary II Battery, Form W (Kelley, et al., 1964b). In addition, the random sample of those students who had scored at or below the fiftieth percentile on the MRT in kindergarten but were not in the program were also given the SAT Primary II Battery.

Instruments

The data for the study was from the files of the school district except the final achievement test which was administered as a special test by the school counselors. A brief

discussion of the instruments is included on the following pages. The information is taken from the manuals for administration for each test and from Buros (1972).

Metropolitan Readiness Tests

The Metropolitan Readiness Tests (Hildreth, et al., 1969) contain six tests designed "to measure the extent to which school beginners have developed in the several skills and abilities that contribute to readiness for first-grade instruction (p. 2)." They are: (1) Word Meaning, (2) Listening, (3) Matching, (4) Alphabet, (5) Numbers, and (6) Copying. Through early classification of the children the tests enable teachers to manage instruction by homogeneously grouping pupils for instructional purposes.

Hildreth, et al., (1969) report validity coefficients of .76 with the Pintner-Cunningham Primary Mental Ability Test and .80 with the Murphy-Durrell Reading Readiness Analysis. In addition they list correlations with tests measuring general intelligence ranging from .41 to .72 (pp. 16-17). Also indicated is a predictive validity at the "level of at least .60, a value that must be considered as very good for test results of five- and six-year-old children who in almost every instance are taking their first group-administered test (p. 23)."

The reliability of the MRT was obtained using both the split-half and test-retest techniques. The split-half reliabilities reported by the authors ranged from .90 to .95.

Results from the test-retest technique show total correlations of .89 and .93 (Hildreth, et al., 1969, p. 27-29).

Vane Kindergarten Test

The Vane Kindergarten Test (Vane, 1968) was designed

. . . to evaluate the intellectual and academic potential and behavior adjustment of young children.
. . . It is based on the assumption that the measurement will provide clues as to the child's functioning ability in school related areas (p. 1).

The VKT consists of three subtests. These subtests are: (1) the Perceptual Motor (P-M) subtest, (2) the Vocabulary (VOC) subtest, and (3) the Draw-A-Man (Man) subtest. It is recommended that the P-M and Man subtests be given to small groups at one time while the VOC subtest is to be administered individually to each child. However, as mentioned earlier (p. 27), this procedure was modified by the counselors giving the test to each child individually. In addition to the scores obtained from the subtests, provisions exist for obtaining a Full Scale score.

In the original publication regarding the test, Vane reported two test-retest reliabilities for the subtests ranging from .71 to .97 (1968, p. 23). Vane also reported a validity coefficient of .76 with Stanford-Binet IQ's for 212 children (p. 24). In addition, correlation coefficients of .60 and .59 have been reported between the VKT and scores from separate groups respectively, one group on the Stanford Achievement Test and another group on the California Achievement Test (p. 21).

Stanford Early School
Achievement Test, Level I

The Stanford Early School Achievement Test, Level I (Madden and Gardner, 1969) consists of four parts and is designed to provide an estimate of a child's early school cognitive abilities. The test is an effort at measuring the "input" present when the child enters school. As the authors state, "Most knowledge at this age is learned spontaneously, but much is learned as a result of rather direct adult intervention (p. 12)." The four parts of the test are: (1) the Environment; (2) Mathematics; (3) Letters and Sounds; and (4) Aural Comprehension.

SESAT I was standardized for use with kindergarten and first-grade children. The standardization was conducted in October, 1968, using 11,106 pupils proportionally distributed with regard to geographic areas and socioeconomic levels according to the 1960 census.

Madden and Gardner (1969) report split-half reliability coefficients for the first-grade sample ranging from .77 on Aural Comprehension to a .89 for Letters and Sounds. They also list standard error of measurements of 2.4 for the Environment; 2.0 for Mathematics; 2.0 for Letters and Sounds; and 2.1 for Aural Comprehension. These standard errors of measurement are reported in raw score form (p. 20). The authors report a correlation of .74 between the total score of SESAT I and the Otis-Lennon Mental Ability Test (p. 20).

Stanford Achievement
Test Primary I Battery

The Primary I Battery of the Stanford Achievement Test (Kelley, et al., 1964a) was designed as the first of a series of achievement batteries for pupil evaluation. It was developed for use from the middle of first grade to the middle of second grade. The battery consists of six tests which measure reading, arithmetic, and spelling, "the three abilities to which greatest attention is devoted in the instructional program of the primary grades (p. 4)." The six tests are: (1) Word Reading; (2) Paragraph Meaning; (3) Vocabulary; (4) Spelling; (5) Word Study Skills; and (6) Arithmetic. Of these, only the tests measuring reading and arithmetic achievement were used in the present study. Reading is measured by the Word Reading, Paragraph Meaning, and Word Study Skills tests. The arithmetic measurement is obtained by means of one test consisting of three parts.

The reliability for the Primary I Battery has been calculated using both split-half reliability and the Kuder-Richardson Formula 20. These reliability measures were calculated for 1000 children from grade 1.6. The split-half reliability coefficients ranged from .79 to .95 and the Kuder-Richardson from .83 to .95.

Kelley, et al., (1964a) did not list validity coefficients for the instrument. They did, however, state that an attempt was made to insure construct validity by reviewing the

prevalent curricula. The ultimate decision regarding validity lies with the users of the test in the evaluation of their pupils' achievement levels. Apparently the decision must be based on the content of the test compared to the instructional content.

Stanford Achievement
Test Primary II Battery

The Primary II Battery of the Stanford Achievement Test (Kelley, et al., 1964b) is the second in the battery series. The Primary II Battery was developed for use with children from the middle of the second grade through the end of the third grade. The Primary II Battery consists of eight tests: (1) Word Meaning; (2) Paragraph Meaning; (3) Science and Social Studies Concepts; (4) Spelling; (5) Word Study Skills; (6) Language; (7) Arithmetic Computation; and (8) Arithmetic Concepts. Once again, only those tests used for measuring achievement in reading and arithmetic were used in the present study. The tests of Word Meaning, Paragraph Meaning and Word Study Skills measure reading ability. Arithmetic Computation and Arithmetic Concepts are the measures of arithmetic skills.

The reliability for the Primary II Battery was calculated by the authors using the split-half reliability technique on each test included in the battery. The correlations ranged from .79 to .94. The validity of the Primary II Battery is based on the same criteria as that claimed for the Primary I Battery.

Statistical Design

The statistical methodology used to examine the effectiveness of the program was based on the multivariate analysis of variance (MANOVA) approach to data analysis. The MANOVA approach uses a design which accommodates several dependent measures or variables. The variables are combined into means vectors and the analysis is conducted to test for significant differences among these vectors. The concern of the research is whether the population centroids are located at different places in the measurement space represented by the dependent vector variable (Cooley and Lohnes, 1971; Overall and Klett, 1972; Kerlinger, 1973). The approach is especially useful in studies where it is not possible to control for all of the important sources of variance (Overall and Klett, p. xvii).

The multivariate analysis was conducted using the multivariate analysis of variance computer program developed by Cramer and Thurston of the University of North Carolina (unpublished). "The program performs univariate and multivariate analysis of variance, covariance, and regression (Cramer and Thurston, unpublished, p. 1)."

The program reports the overall tests of significance using Wilks' Lambda and the univariate F tests for each variable. In addition the program also produces as output the analysis of regression for the covariates as well as the

overall test of significance using Wilks' Lambda and the univariate F tests.

Prior to analyzing the data for student achievement two preliminary analyses were conducted to obtain an understanding of the comparability of the groups. It was felt that the analyses were important since there was a lack of controls on the selection of students for program participation.

The first analysis was performed on the Full Scale scores for the Vane Kindergarten Test. The original information obtained from the program personnel stated that the pupils selected for participation in the program were in the average range of functioning as measured by the VKT. The analysis was performed using a one-way analysis of variance technique (Ferguson, 1966). The hypothesis tested stated that there was no difference among the groups on the VKT Full Scale score.

The second of the preliminary analyses was made on the ages in months at the time of the last test. The analysis also used the one-way analysis of variance approach. The analysis tested the hypothesis that there were no significant differences among groups when compared on age in months.

Primary Analyses

The approach in the following section is the identification of the hypothesis and a discussion of the statistical analysis employed to test each hypothesis.

Hypothesis 1.--Hypothesis 1 compared the three groups across subtests on the MRT to determine if the groups differed on the measured readiness factors prior to entry into the special program.

Hypothesis 1.--There was a significant difference among groups prior to starting the program on the subtests measured by the Metropolitan Readiness Tests; those students selected for participation would score lower on all subtests.

Stated in the null form the hypothesis is:

H₀₁: No difference will be observed among the groups on the subtests of the Metropolitan Readiness Test administered prior to participation.

Hypothesis 1 was examined using the MANOVA computer program cited earlier. The independent variables were the three groups; the dependent variables were the raw scores for the six subtests and the total score obtained on the Metropolitan Readiness Tests (Hildreth, et al., 1969).

Hypothesis 2.--Hypothesis 2 was formulated to examine the Vane Kindergarten Test (Vane, 1968), since the VKT was used as an additional screening instrument and as a tool in the decision regarding the treatment the children received.

Hypothesis 2.--There was a significant difference between the two groups selected for the program on the Vocabulary and Perceptual-Motor subscales of the Vane Kindergarten Test; those students who remained in the program beyond an academic year were in the group scoring lower on the Perceptual-Motor subscale.

Hypothesis 2, stated in null form, is:

H_{02} : No difference will be found between the means of Groups 2 and 3 on the subscales of the Vane Kindergarten Test taken before the start of the program.

The hypothesis was tested using the MANOVA program. It was logical to think that one of the subtests would indicate a need for longer exposure to the treatment provided in the program. The MANOVA approach provided a method whereby the groups involved were compared across all of the subtests and on each individual subtest reported by the univariate F tests. The MANOVA program thus assisted in the decision of whether one group scored significantly different from the other on the dependent variables.

Hypothesis 3.--The effect of the program on academic achievement in the areas of reading and arithmetic was evaluated in Hypothesis 3.

Hypothesis 3.--The students who returned to the regular classroom during or at the end of the first year showed higher achievement than the other two groups.

In null form, Hypothesis 3 states:

H_{03} : There will be no difference among groups when achievement test scores are compared for the second and third grades.

Hypothesis 3 was explored through the use of the MANOVA approach described earlier (p. 34). The dependent measures were the grade scores obtained on the reading and arithmetic sections of the SAT in the second and third grades. Another analysis was a pairwise comparison of means for each

group (e.g., Group 1 vs. Group 2; Group 1 vs. Group 3; and Group 2 vs. Group 3) using the S-method if significant differences were found. The method was developed by Scheffé¹ and followed the procedure recommended by Glass and Stanley (1970).

Hypothesis 4.--Hypothesis 4 was designed to study the effects of previous achievement and intellectual functioning on achievement in the second and third grade.

Hypothesis 4.--When the effects of previous achievement and intellectual functioning are taken into consideration the students who remained in the program longer than one year show higher achievement than those who returned to the regular classroom during or at the end of the first year.

Stated in null form, Hypothesis 4 is:

H_{04} : There will be no difference between Groups 2 and 3 on achievement after accounting for previous learning and ability.

Hypothesis 4 was tested through the use of a multivariate analysis of covariance (MANCOVA) for Groups 2 and 3. The reason for using the MANCOVA approach was to partial out the effects of the variables that were present prior to the program treatment. Analysis of covariance is designed to account for variables which may affect the dependent variables being studied but cannot be controlled through randomization. However, to use the analysis technique there must be an assurance that the covariates have not been affected by the treatment under study. In the case of the present

study the covariates were measured prior to beginning the program. The VKT was given while the students were in kindergarten and the SESAT I was given at the time of entrance into the program. Thus there was no effect on the scores which were directly related to the program treatment. The dependent variables for the analysis were the grade scores obtained on the SAT at both test times. The covariates used were the total of the three subtest scores on the VKT and the total raw scores obtained on the SESAT I.

Hypothesis 5.--Hypothesis 5 was designed to study the gains in achievement from the second to the third grade. From the literature review it was believed that the students remaining in the program would gain more than the group who were returned to their regular classroom during or at the end of the first year. It was also believed that the group of non-participants would achieve more gain.

Hypothesis 5.--Groups 1 and 3 showed higher achievement than Group 2 at the third grade test time after the scores were adjusted for second-grade achievement.

Stated in null form for testing, Hypothesis 5 was:

H₀₅: There will be no difference in achievement for the three groups after the third-grade achievement scores are adjusted for second-grade achievement.

The hypothesis was tested using analysis of covariance included in the MANOVA computer program. The dependent measures for the analysis were the scores for reading and

arithmetic measures obtained in the third grade. The covariates for the analysis were the scores available from the second grade. The analysis was conducted on the reading scores separate from the arithmetic scores. Thus, the reading scores were adjusted for previous reading achievement and the arithmetic scores were adjusted for previous arithmetic achievement.

If significant differences were found, pairwise comparisons would be made on the adjusted means for the three groups. The adjustments would be made following the procedure recommended by Winer (1971), after obtaining basic statistical information for each group through the use of the BMD02D computer program (Dixon, 1970).

CHAPTER IV

DATA ANALYSIS AND RESULTS

The data analyses conducted in the study followed the chronological sequence from the initial evaluation at the kindergarten level through the administration of the last instrument in the third grade. There were additional analyses conducted to ensure a modicum of comparability between the three groups involved in the study.

The first of the preliminary analyses was conducted using the Full Scale score from the VKT to determine if any difference existed among the three groups before involvement in the program under study. The means and standard deviations for the three groups are shown in Table 1.

TABLE 1
MEANS AND STANDARD DEVIATIONS FOR GROUPS ON
FULL SCALE SCORE OF THE VKT

Group	n	\bar{X}	s.d.
1	17	104.94	8.81
2	23	101.35	12.10
3	30	98.87	9.63

The scores were analyzed by a one-way analysis of variance (Ferguson, 1966). The results of the analysis showed no significant difference with $F_{2,67} = 1.88$. The summary of the analysis is presented in Table 2.

TABLE 2
ANALYSIS OF VARIANCE FOR FULL SCALE SCORES OF VKT

Source of Variation	Df	Mean Square	F
Among	2	200.82	1.88
Within	67	106.74	
Total	69		

Because only seventeen of the forty-seven subjects in Group 1 had VKT scores there was no way of determining if they were representative of the total sample. In addition, since no significant difference was found a decision was made to not include the scores for the seventeen subjects in any later analysis of the VKT.

The next analysis was conducted to determine if any differences existed among the groups on mean age. The analysis of variance was performed on the subject's age in months at the time of the last test. The elimination of ten subjects from this analysis was necessitated by conflicts in the records used for gathering the data. Two sources were used; in some instances the date of birth was different.

The mean ages and standard deviations for the groups are shown in Table 3.

TABLE 3
MEAN AGES (IN MONTHS) OF SUBJECTS IN THE THREE GROUPS

Group	n	\bar{X}	s.d.
1	38	101.39	4.51
2	23	101.00	3.77
3	29	99.86	3.37

The analysis of variance (Ferguson, 1969) among mean age by groups produced an F value of 1.26 for df of 2 and 87. The value was not significant. The summary of the analysis is presented in Table 4.

TABLE 4
SUMMARY TABLE FOR ANALYSIS OF VARIANCE OF
AGE (IN MONTHS) AT TIME OF LAST TEST

Source of Variation	df	Mean Square	F
Among	2	19.94	1.26
Within	87	15.87	
Total	89		

A final preliminary analysis was conducted to determine if Groups 2 and 3 differed significantly on the number

of pupils retained. The analysis was conducted by the use of Chi square. The analysis produced a Chi square of 9.09 which is significant at the .01 level. More students were retained in Group 3 than Group 2. Table 5 presents the data for the analysis.

TABLE 5
SUMMARY TABLE FOR CHI SQUARE

	Retained	Not Retained	
Group 2	5	18	23
Group 3	19	11	30
	24	29	53

Testing of the Hypotheses

Hypothesis 1

Hypothesis 1 was formulated to evaluate the differences between the groups used in the study on the readiness factors as measured by the MRT prior to program involvement.

Hypothesis 1 stated:

There was a significant difference among groups prior to starting the program on the subtests measured by the Metropolitan Readiness Tests; the students selected for participation would score lower on all subtests.

Stated in the null form, Hypothesis 1 is:

No difference will be observed among the mean scores of the subtests of the Metropolitan Readiness Tests administered prior to participation.

A MANOVA was performed using a total of 163 subjects. The tests of significance using Wilks' Lambda yielded an approximate overall F of 2.65. The probability of this F occurring by chance is less than .001. Therefore, there was a significant difference among groups on the MRT. Hence, the null hypothesis was rejected.

The data were also analyzed by the program using univariate F tests for each variable of the MRT. Table 6 shows the univariate F tests.

TABLE 6
UNIVARIATE F TESTS OF THE MRT VARIABLES

Variable	Mean Square	F _(2,160)	p
Word Meaning	26.84	5.10	.007
Listening	26.54	4.17	.017
Matching	6.49	0.99	.375
Alphabet	57.43	5.02	.008
Numbers	88.81	10.71	.001
Copying	18.79	2.09	.127
Total Score	1036.66	14.19	.001

The results displayed in Table 6 indicated that while there was an overall difference among groups there was not a significant difference for each of the variables as analyzed by

the univariate F tests. Therefore, it is possible to reject the hypothesis that those students participating in the program scored lower on all the subtests of the MRT.

Comparison of the means of the groups shown in Table 7 allowed for further refinement of the results of the F tests.

TABLE 7
MEANS AND STANDARD DEVIATIONS OF THE MRT VARIABLES

Subtest	Group 1 (n = 95)		Group 2 (n = 40)		Group 3 (n = 49)	
	\bar{X}	s.d.	\bar{X}	s.d.	\bar{X}	s.d.
Word Meaning	7.42	2.38	6.20	2.17	7.00	2.32
Listening	9.21	2.25	8.13	2.83	8.88	7.93
Matching	5.93	2.52	5.30	2.38	5.76	2.67
Alphabet	8.06	3.40	7.30	3.52	6.16	3.61
Numbers	9.42	2.58	7.55	3.59	7.41	2.65
Copying	6.02	3.06	5.73	3.12	4.90	2.99
Total Score	46.11	6.18	40.20	11.69	40.12	10.36

The comparisons were conducted using the S-method. The results showed that the means for Group 1 on Word Meaning, Listening, Numbers, and Total Score were significantly higher than the means for Group 2. They also showed that the means for Group 1 on Alphabet, Numbers, and the Total Score were significantly higher than the means for Group 3. No significant differences were found between the means for Groups 2

and 3. Tables showing the differences between group means may be found in Appendix A.

Hypothesis 2

Hypothesis 2 tested the differences between Groups 2 and 3 on the Vane Kindergarten Test. Hypothesis 2 was stated:

There was a significant difference between the two groups selected for the program on the Vocabulary and Perceptual-Motor subscales of the Vane Kindergarten Test; the students who remained in the program beyond an academic year were in the group scoring lower on the Perceptual-Motor subscale.

The null hypothesis to be tested was:

No difference will be found between the means of Groups 2 and 3 on the subscales of the Vane Kindergarten Test taken before the start of the program.

Hypothesis 2 was examined by the same MANOVA program which was used for Hypothesis 1. For the analysis, scores were available and used for all of the eighty-three students who had participated in the program from the first semester of the first grade.

The test of significance using Wilks' Lambda yielded an approximate F value of 1.58 which is not significant ($p > .20$). In addition, the univariate F tests conducted on the subscale score showed no statistically significant differences between the groups. The univariate F test results are displayed in Table 8. On the basis of the results obtained in the analysis of the VKT, which showed no statistically

significant differences, the null hypothesis cannot be rejected.

TABLE 8
UNIVARIATE F TESTS CONDUCTED ON VKT SUBSCALE SCORES

Variable	Mean Square	F _(1,81)	p
P-M	203.15	.72	.397
VOC	345.82	1.24	.268
MAN	313.01	1.51	.222

Hypothesis 3

The effect which the program has had on the reading and arithmetic achievement of the participating students was analyzed in Hypothesis 3. Hypothesis 3 proposed.

The students who returned to the regular classroom during or at the end of the first year showed higher achievement than the other two groups.

In the null form Hypothesis 3 was:

There will be no difference among groups when achievement test scores are compared for the second and third grades.

Hypothesis 3 was explored through the use of the MANOVA program. The MANOVA on the Stanford Achievement Test revealed an approximate F of 1.69 using Wilks' Lambda; the value is not significant ($p > .11$). However, on all four of the univariate F tests significance was found. Thus there

were differences among the groups established on the basis of the length of program participation. Table 9 displays the results of the univariate F tests.

TABLE 9
RESULTS OF THE UNIVARIATE F TESTS AMONG GROUPS ON THE
READING AND ARITHMETIC SCORES OF THE
STANFORD ACHIEVEMENT TESTS

Tests	Mean Square	F (2,68)	p
<u>SAT I</u>			
Reading	146.08	3.85	.026
Arithmetic	40.95	4.25	.018
<u>SAT II</u>			
Reading	300.96	5.66	.005
Arithmetic	92.92	3.28	.044

Table 10 reports the means and standard deviations for the four SAT variables.

TABLE 10
MEANS AND STANDARD DEVIATIONS OF SAT VARIABLES

Tests	Group 1		Group 2		Group 3	
	\bar{X}	s.d.	\bar{X}	s.d.	\bar{X}	s.d.
<u>SAT I</u>						
Reading	22.26	6.77	22.13	5.81	16.40	2.27
Arithmetic	19.54	3.26	18.73	2.94	16.40	2.50
<u>SAT II</u>						
Reading	31.76	7.01	31.33	8.99	23.30	5.50
Arithmetic	28.52	5.52	27.20	5.32	23.80	4.26

As stated earlier, the overall test showed no significant differences among the three groups. However, with the significant results shown in Table 9, it appears that there were interactions occurring which caused the nonsignificant results. To decide if Group 2 showed higher achievement than the other two groups, pairwise comparisons were made using the S-method. The results of these comparisons showed that Group 2 did not differ significantly from Group 1 on the four measures. However, there was a significant difference between Group 2 and Group 3 on the SAT I Reading and the SAT II Reading measures with Group 2 scoring higher. Hypothesis 3 was rejected since Groups 2 and 3 did score significantly different. Tables showing the differences between group means may be found in Appendix B.

Hypothesis 4

Hypothesis 4 dealt with the effect of the program under study on the level of achievement at the second and third grades, with the effect of prior achievement as measured at entrance to the program and ability as measured by the VKT taken into consideration. Hypothesis 4 stated:

When the effects of previous achievement and intellectual functioning are taken into consideration, the students who remained in the program longer than one year show higher achievement than those who returned to the regular classroom during or at the end of the first year.

The hypothesis stated in null form for testing was:

There will be no difference between Groups 2 and 3 in achievement after accounting for previous learning and ability.

Hypothesis 4 was tested through the use of the MANCOVA included in the computer program. The achievement tests were analyzed in two stages so that pupils with scores on only one of the tests could be included in the analysis. The division of the analysis was necessary because a total of thirty one students had scores on the SAT I and only twenty-six pupils had scores on the SAT II. Eight out of both groups had one but not the other of the results available.

As stated earlier there was a significant difference among groups on both SAT I and SAT II reading scores. The results of the MANCOVA for Groups 2 and 3 for the SAT I variables indicated that when the prior achievement and ability were accounted for the groups showed no difference on either the reading or arithmetic scores. However, on the SAT II there were significant differences for both the reading and arithmetic measurements. Therefore, the null hypothesis was rejected. The results of the analysis are shown below in Table 11.

It should be pointed out that the number of cases in the groups is small (SAT I: $n = 31$; SAT II: $n = 26$). Also, in the MANCOVA the covariates do not appear to have contributed to the differences which exist on reading at the third grade, since the MANOVA also showed significant differences.

TABLE 11

RESULTS OF THE MULTIVARIATE ANALYSIS OF COVARIANCE
FOR GROUPS 1 AND 2 ON THE STANFORD
ACHIEVEMENT TEST VARIABLES

Tests	df	F	Mean Square	p
<u>Overall Tests of Significance</u>				
SAT I	4,52	1.57		.195 N.S.
SAT II	4,42	3.29		.020
<u>Univariate F Tests</u>				
SAT I				
Reading	2,27	2.40	54.17	.110 N.S.
Arithmetic	2,27	2.15	14.59	.136 N.S.
SAT II				
Reading	2,22	5.18	179.34	.014
Arithmetic	2,22	5.46	97.69	.012

Hypothesis 5

Hypothesis 5 was formulated to study the gains in achievement from the second to the third grade as measured by the two SAT instruments. Hypothesis 5 proposed:

Groups 1 and 3 showed higher achievement than Group 2 at the third grade test time after the scores were adjusted for second grade achievement.

Stated in null form, Hypothesis 5 is:

There will be no difference in achievement in the third grade when the scores are adjusted for second-grade achievement.

The data were analyzed through the analysis of covariance technique (ANCOVA) which is included in the MANOVA program used throughout the present study. The results from the analysis show that after the scores were adjusted there were no significant differences among the three groups. The analysis of covariance showed an F of 1.79 ($p > .18$) on the reading variable and an F of .54 ($p > .58$) on the arithmetic variable. The results of the analysis on the reading scores are summarized in Table 12. Table 13 shows the results for the analysis of the arithmetic scores.

TABLE 12

ANALYSIS OF COVARIANCE ON SAT II READING SCORES
ADJUSTED BY SAT I READING SCORES

Source of Variation	df	Mean Square	F	p
Within Cells	67	28.41		
Covariate	1	1712.20	60.26	.001
Among Groups	2	50.85	1.79	.175

TABLE 13

ANALYSIS OF COVARIANCE ON SAT II ARITHMETIC SCORES
ADJUSTED BY SAT I ARITHMETIC SCORES

Source of Variation	df	Mean Square	F	p
Within cells	67	19.68		
Covariate	1	611.12	31.06	.001
Among groups	2	10.67	.54	.584

Since there were no significant differences there was a failure to reject the null hypothesis. It is possible to reject the original hypothesis that Group 1 and 3 would show higher adjusted achievement in the third grade.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

The following chapter summarizes the findings of the study reported. It includes the conclusions which have been drawn based on the results after considering the limitations which were inherent in an ex post facto research study. The final section of the chapter lists the recommendations for additional research stemming from the present study.

Conclusions

The analysis of the scores from the two screening instruments showed different pictures. The analysis of the Metropolitan Readiness Test showed that Group 1 was initially different from Groups 2 and 3 but that there was no initial difference between Groups 2 and 3. On the tests where there were differences, Group 1 (non-participants) scored significantly higher than Groups 2 and 3 (program participants). These differences indicated the use of the MRT for screening was effective in selecting students for program participation.

Since the groups were initially different on four individual tests of the MRT, an analysis using the MRT to adjust the achievement scores on the SAT I and the SAT II should be conducted using all three groups. Such an analysis

would show the effect of readiness measured in kindergarten on subsequent achievement.

The analysis of the Vane Kindergarten Test showed no significant differences among groups. The results of the analysis between Group 1 and Groups 2 and 3 cannot be taken as conclusive because of the small number of children in Group 1 with VKT scores. Further research is needed with subjects not in the participating group having VKT results available for analysis. Such research would show whether the VKT added sufficient information to make its administration worthwhile.

It was believed that the scores on the subtests of the VKT would discriminate between Groups 2 and 3. The discrimination might be expected since the VKT is purported to measure ability, and ability should affect the level of achievement. The absence of significant differences between Groups 2 and 3 on the VKT indicates that the test has little power in the discrimination of the level of achievement, and therefore gives no assistance in estimating the duration of treatment required.

The evaluation of the achievement variables, although giving indications regarding the effectiveness of the program, must be viewed with caution. Prudence is required because of the decline in the number of subjects remaining in the district who had results available for analysis. Even more important, it should be remembered that twenty-four

students (i.e., 45.2 per cent) were retained either in the first or second grade and were not included in the analysis. The Chi square analysis conducted indicated a highly significant difference on the number of students retained between the two groups who had participated in the program in the first grade. Group 3 had more students retained than Group 2. These students, if included, might have led to different results than those obtained in the present study. Therefore, the results can only be interpreted within the operational limits to describe the trend for the program and directions for further research. For the purpose of future evaluation of the program, data should be collected on those students who have been retained which is equivalent to that for students not retained. This would allow for the inclusion of retained students in future evaluations of the program.

Reading and Arithmetic Results

As the students progressed they were returned to the regular classroom on a full-time basis. Group 2 was composed of those students who scored at grade level on the SESAT test, and were, therefore, returned during or at the end of the first year. From the results obtained in the analysis of Hypothesis 3, it was shown that at the beginning of the second grade and at the beginning of the third grade, those students in Group 2 compared favorably with those students

who had not participated in the program. The achievement scores as measured by the SAT I and the SAT II were not significantly different for Groups 1 and 2.

On the other hand, the members of Group 3 did not fare as well through the efforts of the program. Their achievement scores were significantly lower than scores for those not involved in the program on the SAT I and the SAT II achievement scales used in the present study at the beginning of both the second and third grades.

Group 3 showed a significant difference on the SAT I reading variable and on the SAT II on both the reading and arithmetic variables when compared to Group 2, with Group 2 scoring higher. When comparing Groups 2 and 3 on the SAT I arithmetic variable, Group 2 scored higher, but the difference was not significant. From this it is concluded that the additional academic year in the program did not raise the students in Group 3 to the level of achievement of those students in Groups 1 and 2. Therefore, the additional year of the program had no measured effect. A question is raised about the efficacy of the program beyond a point not yet determined.

It would be tempting to conclude that Group 2 achieved at a higher rate than Group 1 since there was an initial difference in readiness between Groups 1 and 2. Yet, Group 2 scored significantly higher in achievement than Group 3 and there were no significant differences on the MRT. It is

concluded that in the present study readiness, as measured by the MRT, was not a predictor of the level of achievement that would be attained in the first and second grades.

There is an apparent paradox in the analysis conducted on the achievement scores at the second and third grades. A non-significant F was found on the overall test of significance but significant F's were found among groups on each of the univariate F tests. Tatsuoaka (1970) states that

The danger of getting a distorted picture of the group differences tends to increase as the correlations among the variables become larger (although we cannot assert that the greater the correlations, the greater the distortion will necessarily be) (p. 2).

The variables used in the present study are correlated, thus if a difference exists on one variable there is the possibility that part of any difference found on a second variable should be contributed to the first. The confounding of the results leads to possible confusion in the interpretation. To avoid the confusion in the present study the MANOVA approach was taken.

The additional analysis using the covariance technique adjusted the achievement scores for initial achievement and ability as measured by the SESAT I and the VKT, respectively. Since neither SESAT I nor VKT scores were available for the non-participants, only Groups 2 and 3 were included in the analysis. Before adjusting for initial achievement

there was a difference on SAT I reading, after adjustment there were no significant differences. This implies that the differences in achievement at the beginning of the second grade should be attributed to ability and level of achievement at the beginning of the program.

However, after the adjustment of the SAT II scores was accomplished using the same covariates, significant differences were found. The finding that the difference at the third-grade level remained after adjustment but had disappeared at the second-grade level might indicate that involvement in the second year of the program caused the students to lag behind. Since the differences in achievement at the third-grade level were unchanged by adjustment for initial ability or achievement, it can be concluded that the difference can be attributed to the additional year of participation. Therefore, the year spent in the regular classroom by Group 2 appeared to contribute more to achievement scores than the second year in the program for Group 3.

The final analysis adjusted the third-grade achievement scores for achievement in the second grade through the use of the reading and arithmetic scores on SAT I as covariates for the reading and arithmetic scores on the SAT II respectively. These results seem to conflict with the results discussed in the previous paragraph. It can be interpreted that since achievement in the second grade was proportional for the three groups, the program did not hamper the

achievement of those students who spent an additional year in the program. In the light of these findings the interpretation of the analysis conducted to adjust for ability and initial achievement must be reexamined. Even though members of Group 3 gained in achievement proportional to members of Group 2 in the second grade, they did not gain as much. This would account for the negation of the differences at the second grade but not the differences at the third grade when adjusting for initial ability and achievement.

Neither ability nor readiness as measured by the VKT and the MRT respectively seem to account for differences in achievement on the SAT II for Groups 2 and 3. This difference may be attributable to social or emotional factors. Further research is indicated to determine if emotional or social factors are involved.

Overall, the following summary statements may be made regarding the conclusions drawn from the present study.

1. There were differences among groups on the Metropolitan Readiness Tests.
2. The two groups in the program did not differ on the subtests of the Vane Kindergarten Test.
3. Neither the VKT nor the MRT predicted the level of achievement nor the duration of treatment.
4. The group of children who were returned to the regular classroom full-time during or at the end of the first year had progressed in a manner not different from those students who did not participate.

5. Even though Groups 2 and 3 scored proportionally in the second grade, Group 2 achieved more than Group 3.
6. The program appears to have moved the higher achievers back into the regular classroom, thus accounting for the differences on the SAT II but not SAT I after adjustment for initial ability and achievement.
7. Since differences in achievement cannot be attributed to either readiness or ability, there must be other factors involved.

Recommendations for Research

The following recommendations for further research have been made on the basis of the present study.

1. Since the groups were initially different on the readiness factors of the MRT, an analysis using the MRT to adjust the achievement scores on the SAT I and the SAT II should be conducted using all three groups.
2. Further research is needed with subjects not in the participating group having VKT results available for analysis to find out if the VKT added sufficient information to make its administration worthwhile.
3. For the purpose of future evaluation of the program, data should be collected on those pupils who have been retained equivalent to that for pupils not retained.
4. To evaluate any possible differences in ability to manage relationships with peers, a comparison is needed between pupils involved in the program and those not involved. Such research would assist in understanding the social development of the participants.
5. Further research is needed to determine if social or emotional factors are involved in the level of achievement for the pupils who participated in the program.

6. Research is needed to compare the students starting the program in the first semester to those beginning the program in the second semester of the first grade.
7. The ultimate research is a long-term follow-up on the students involved in the program to determine the effectiveness of the program in preventing drop-outs.

The present study has been concerned with the achievement in the academic areas of reading and arithmetic as measured by the Primary I and Primary II Batteries of the Stanford Achievement Test. The results cannot be generalized beyond groups similar to those studied, but may give indications regarding the effectiveness of the program for those students who have participated.

REFERENCES

REFERENCES

- Anderson, S. B. and Maier, M. H. Thirty-four thousand pupils and how they grew. Journal of Teacher Education, 1963, 14, 212-216.
- Anderson, S. B. and Temp, G. Project Head Start, Summer 1966. Final report, section 3, Educational Testing Service, Contract OEO 1359, Office of Economic Opportunity, May 1967.
- Ashcraft, C. W. Later school achievement of treated and untreated emotionally handicapped children. Journal of School Psychology, 1971, 9, 338-342.
- Askow, E. N. and Fischbach, T. J. Investigation of primary pupils attitudes toward reading. Journal of Experimental Education, 1973 (Spring), 41, 1-7.
- Buckland, P. and Balow, B. Effects of visual perceptual training on reading readiness. Exceptional Children, 1973, 39(4), 299.
- Buros, O. K. The Seventh Mental Measurements Yearbook. Highland Park, New Jersey: The Gryphon Press, 1972.
- Campbell, D. T. and Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), Disadvantaged Child, Vol. 3. New York: Brunner/Mazel, Inc., 1970.
- Campbell, D. T. and Erlebacher, A. Reply to the replies. In J. Hellmuth (Ed.), Disadvantaged Child, Vol. 3. New York: Brunner/Mazel, Inc., 1970.
- Circirelli, V. G. The relevance of the regression artifact problem to the Westinghouse-Ohio evaluation of Head Start: A reply to Campbell and Erlebacher. In J. Hellmuth (Ed.), Disadvantaged Child, Vol. 3. New York: Brunner/Mazel, Inc., 1970.
- Cooley, W. and Lohnes, P. Multivariate Data Analysis. New York: John Wiley and Sons, Inc., 1971.

- Cramer, E. and Thurston, L. The University of North Carolina Manova Program. Chapel Hill, North Carolina: Psychometric Laboratory, University of North Carolina. Unpublished.
- D'Angelo, R.; Walsh, J.; and Lomangino, L. IQ's of Negro Head Start children on the Vane Kindergarten Test. Journal of Clinical Psychology, 1971, 27(1), 82-83.
- Datta, L. A report on evaluation studies of Project Head Start. Washington, D. C.: Office of Economic Opportunity, 1969.
- Dixon, W. J. (Ed.) Biomedical Computer Programs: University of California Publications in Automatic Computation. No. 2. University of California Press, 1970.
- Educational Testing Service. Disadvantaged children and their first school experiences--ETS-OEO longitudinal study. Theoretical considerations and measurement strategies. Appendices related to measures. Princeton, New Jersey, December 1968.
- Educational Testing Service. From theory to operations. Disadvantaged children and their first school experiences. ETS-Head Start longitudinal study. Princeton, New Jersey, August 1969.
- Educational Testing Service. Disadvantaged children and their first school experiences. ETS-Head Start longitudinal study. Princeton, New Jersey, August 1970.
- Educational Testing Service. The sample. Children tested in 1969 operations in the Head Start year. Disadvantaged children and their first school experiences. ETS-Head Start longitudinal study. Princeton, New Jersey, February 1973.
- Evans, J. W. and Schiller, J. How preoccupation with possible regression artifacts can lead to a faulty strategy for the evaluation of social action programs: A reply to Campbell and Erlebacher. In J. Hellmuth (Ed.), Disadvantaged Child, Vol. 3. New York: Brunner/Mazel, 1970.
- Glass, G. V. and Stanley, J. C. Statistical Methods in Education and Psychology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.
- Goolsby, T. M., Jr., and Frary, R. B. Validity of the Metropolitan Readiness Test for white and Negro students in a southern city. Educational and Psychological Measurement, 1970, 30, 443-450.

- Grotberg, E. H. Early intervention for the disadvantaged: Does it influence reading achievement. Paper presented at the meeting of the International Reading Association, Atlantic City, April 1971.
- Hackensack Public Schools. Overview and evaluation of Project LEM. Innovative education in action. Hackensack, New Jersey, 1972.
- Haring, H. G. and Ridgway, R. W. Early identification of children with learning disabilities. Exceptional Children, 1967, 33, 387-395.
- Hildreth, G. H.; Griffiths, N. L.; and McGauvran, M. E. Manual of Directions--Metropolitan Readiness Test. New York: Harcourt, Brace and World, Inc., 1969.
- Hoffman, R. B. Evaluation of Elementary School Enrichment Program (Experimental Program 94), Wantagh Public Schools, Wantagh, New York, 1962.
- Jacobs, J. M. Follow-up evaluation of the Frostig Visual Perceptual Training Program. Educational Leadership, 1968, 26, 169-175.
- Kelley, T. L.; Madden, R.; Gardner, E. F.; and Rudman, H. C. Stanford Achievement Test: Directions for Administering, Primary I Battery. New York: Harcourt, Brace and World, Inc., 1964.
- Kelley, T. L.; Madden, R.; Gardner, E. F.; and Rudman, H. C. Stanford Achievement Test: Directions for Administering, Primary II Battery. New York: Harcourt, Brace and World, Inc., 1964.
- Kerlinger, F. N. Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1973.
- Klaus, R. A. and Gray, S. W. The Early Training Project for disadvantaged children: A report after five years. Monographs of the Society for Research in Child Development, 1968, 33 (Serial No. 120).
- Kohlberg, L. Early education: A cognitive developmental view. Child Development, 1968, 39, 1013-1062.
- Lederman, E. and Blair, J. R. Comparison of the level and predictive validity of preschool attainment record ratings obtained from teachers and mothers. Psychology in the Schools, 1972, 9, 392-395.

- Lessler, K. and Bridges, J. S. Prediction of learning problems in a rural setting: Can we improve on readiness tests? Journal of Learning Disabilities, 1973, 6, 90-94.
- McKnab, P. A. and Fine, M. J. Vane Kindergarten Test as a predictor of first grade achievement. Journal of Learning Disabilities, 1972, 5, 503-505.
- Madden, R. and Gardner, E. F. Stanford Early School Achievement Test: Directions for Administering. New York: Harcourt, Brace and World, Inc., 1969.
- Merenda, P. F.; Novack, H. S.; and Bonaventure, E. Relationships between the California Test of Mental Maturity and the Stanford Achievement Test battery at the primary level. Educational and Psychological Measurement, 1972, 32, 1079-1087.
- Mitchell, B. C. A comparison of the achievement-intelligence relationships for pupils with that for school systems. Journal of Educational Research, 1963, 57, 172-180.
- Moore, R. C. and Ogletree, E. G. A comparison of the readiness and intelligence of first grade children with and without a full year of Head Start training. Education, 1973, 93(3), 266.
- Musgrove, W. J. Follow-up study of black and white kindergarten children on academic achievement. Academic Therapy, 1971-72, 7, 123-129.
- Newman, A. P. Later achievement study of pupils underachieving in reading in first grade. Reading Research Quarterly, 1972, 7, 477-508.
- Overall, J. and Klett, C. Applied Multivariate Analysis. New York: Holt, Rinehart and Winston, 1967.
- Prager, B. B.; McGowan, J. R.; Bayuk, R. J., Jr.; Mann, L.; Trevorrow, R. L.; and Massa, E. Relative predictive and construct validities of the Otis-Lennon Mental Ability Test, the Lorge-Thorndike Intelligence Test, and the Metropolitan Readiness Tests in grades two and four: A series of multivariate analyses. Educational and Psychological Measurement, 1971, 31, 529-538.
- Reidford, P. and Berzonsky, M. Field test of an academically oriented preschool curriculum. Paper presented at AERA Convention, New York, February, 1967.

- Scott, C. M. Intelligence and gain in reading as related to gains in the sub-tests of the Stanford Achievement Test. Journal of Educational Research, 1963, 56, 494-496.
- Smith, H. T. Role of the elementary grades in developing the potentials of disadvantaged children: Kalamazoo County Michigan. Kalamazoo, Michigan: The W. F. Upjohn Institute for Employment Research, July 1966.
- Tatsuoka, M. M. Discriminant Analysis: The Study of Group Differences. Selected Topics in Advanced Statistics, 1970, No. 6.
- Turner, R. V. and DeFord, E. F. Follow-up study with differing preschool experiences. Paper presented at the annual meeting of the American Educational Research Association, March 1970.
- Van De Riet, V. and Van De Riet, H. An evaluation of the effects of a unique sequential learning program on culturally deprived preschool children--Final Report, United States Office of Economic Opportunity, 1965.
- Vane, J. R. The Vane Kindergarten Test. Journal of Clinical Psychology, 1968, 24(2), 121-154.
- Walsh, J. F.; D'Angelo, R.; and Lomangino, L. Performance of Negro and Puerto Rican Head Start children on the Vane Kindergarten Test. Psychology in the Schools, 1971, 8(4), 357-358.
- Webb, W. W. and Pate, J. E. Predicting failure in the primary grades. Educational and Psychological Measurement, 1970, 30(2), 459-462.
- Weininger, O. Ready or not: Some psychological aspects of readiness in relation to learning effectiveness. Education, 1972, 93(2), 141-146.
- Willis, D. J. Perceptual and cognitive performance of children as functions of socioeconomic class. Unpublished Doctoral Dissertation, The University of Oklahoma, 1970.
- Winer, B. J. Statistical Principals in Experimental Designs, (2nd Ed.). New York: McGraw-Hill, 1971.
- Yahraes, H. Reducing the effects of restricted environments. In O. Milton and R. G. Wahler (Eds.), Behavior Disorders: Perspectives and Trends. Philadelphia: J. B. Lippincott Company, 1973.

APPENDICES

APPENDIX A

DIFFERENCES BETWEEN MEANS FOR ANALYSIS BY S-METHOD ON THE METROPOLITAN READINESS TESTS

Tests	Group 1	Group 2	Group 3
<u>Word Meaning</u>			
Group 1	--	1.45*	.65
Group 2		--	-.80
Group 3			--
<u>Listening</u>			
Group 1	--	1.39*	.81
Group 2		--	-.58
Group 3			--
<u>Alphabet</u>			
Group 1	--	1.03	2.06*
Group 2		--	1.03
Group 3			--
<u>Numbers</u>			
Group 1	--	1.79*	2.36*
Group 2		--	.58
Group 3			--
<u>Total</u>			
Group 1	--	6.93*	7.51*
Group 2		--	.58
Group 3			--

*significant difference

APPENDIX B

DIFFERENCES BETWEEN MEANS FOR ANALYSIS BY S-METHOD ON THE STANFORD ACHIEVEMENT TESTS

Tests	Group I	Group 2	Group 3
<u>Reading, SAT I</u>			
Group 1	--	.13	5.86*
Group 2		--	5.73*
Group 3			--
<u>Arithmetic, SAT I</u>			
Group 1	--	.81	3.14*
Group 2		--	2.33
Group 3			--
<u>Reading, SAT II</u>			
Group 1	--	.43	8.46*
Group 2		--	8.03*
Group 3			--
<u>Arithmetic, SAT II</u>			
Group I	--	1.32	4.72*
Group 2		--	3.40*
Group 3			--

*significant difference