The Impact Of Programed Tutoring On The Reading Achievement Of Lower Achieving Second Grade Children From Low-Income Areas

Keith Eugene Hartwig
University of the Pacific

Follow this and additional works at: https://scholarlycommons.pacific.edu/uop_etds

Recommended Citation

This Dissertation is brought to you for free and open access by the Graduate School at Scholarly Commons. It has been accepted for inclusion in University of the Pacific Theses and Dissertations by an authorized administrator of Scholarly Commons. For more information, please contact mgibney@pacific.edu.
THE IMPACT OF PROGRAMED TUTORING ON THE READING ACHIEVEMENT OF LOWER ACHIEVING SECOND GRADE CHILDREN FROM LOW-INCOME AREAS

A Dissertation
Presented to
the Faculty of the School of Education
University of the Pacific

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

by
Keith Eugene Hartwig
May 1972
ACKNOWLEDGMENTS

The author wishes to express his appreciation to Mr. C. Paul Thompson of the Sacramento City Unified School District, and to Dr. Phillip L. Harris of Indiana University. Without their interest, advice, and cooperation, this investigation would not have been possible.

The author is also indebted to the members of his committee, Drs. Kenneth L. Beauchamp, Patrick W. Carlton, Dale M. Heckman, Heath Lowry, and particularly his chairman Dr. William C. Theimer Jr. for their advice and counsel during the preparation of this manuscript. Their sincerity and efficiency in action are deeply appreciated.

Last, but certainly not least, the author would simply like to thank Shirley and Doug for all they have given up over the last two years.
AUTHOR'S NOTE

The existing literature in the area of programmed instruction generally has used the double "m" spelling of the term "programmed". The literature that has evolved in the area of programed tutoring has generally adopted the single "m" spelling of the term "programed". These conventions of the literature have been followed in this dissertation; thus, what may appear to be an inconsistency in spelling is, in fact, a consistency with the conventions of the literature.

Keith E. Hartwig
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION TO THE STUDY</td>
<td>1</td>
</tr>
<tr>
<td>STATEMENT OF THE PROBLEM</td>
<td>2</td>
</tr>
<tr>
<td>PURPOSE OF THE STUDY</td>
<td>3</td>
</tr>
<tr>
<td>RATIONALE FOR THE STUDY</td>
<td>5</td>
</tr>
<tr>
<td>DEFINITION OF TERMS</td>
<td>7</td>
</tr>
<tr>
<td>RESEARCH HYPOTHESES</td>
<td>8</td>
</tr>
<tr>
<td>LIMITATIONS OF THE STUDY</td>
<td>11</td>
</tr>
<tr>
<td>OVERVIEW OF THE STUDY</td>
<td>11</td>
</tr>
<tr>
<td>2. REVIEW OF THE LITERATURE</td>
<td>13</td>
</tr>
<tr>
<td>PROGRAMMED INSTRUCTION</td>
<td>14</td>
</tr>
<tr>
<td>Theory Underlying Programmed Instruction</td>
<td>15</td>
</tr>
<tr>
<td>Differences Between Pressey and Skinner</td>
<td>16</td>
</tr>
<tr>
<td>Research Related to the Underlying Theory</td>
<td>17</td>
</tr>
<tr>
<td>Summary</td>
<td>28</td>
</tr>
<tr>
<td>PROGRAMMED INSTRUCTION AND READING</td>
<td>28</td>
</tr>
<tr>
<td>Programmed versus basal materials</td>
<td>29</td>
</tr>
<tr>
<td>Supplemental Programmed Instruction</td>
<td>34</td>
</tr>
<tr>
<td>Other Research and Opinion</td>
<td>36</td>
</tr>
<tr>
<td>Summary</td>
<td>38</td>
</tr>
<tr>
<td>TUTORIAL PROGRAMS</td>
<td>39</td>
</tr>
<tr>
<td>The Rationale for Tutorial Programs</td>
<td>40</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>High School Students as Tutors</td>
<td>43</td>
</tr>
<tr>
<td>Elementary School Students as Tutors</td>
<td>47</td>
</tr>
<tr>
<td>Adults as Tutors</td>
<td>50</td>
</tr>
<tr>
<td>Summary of Tutorial Programs</td>
<td>54</td>
</tr>
<tr>
<td>PROGRAMED TUTORING</td>
<td>55</td>
</tr>
<tr>
<td>The Developmental Period</td>
<td>55</td>
</tr>
<tr>
<td>The Field Test Period</td>
<td>57</td>
</tr>
<tr>
<td>Operational Programs</td>
<td>59</td>
</tr>
<tr>
<td>Summary of Programed Tutoring</td>
<td>61</td>
</tr>
<tr>
<td>SUMMARY OF THE CHAPTER</td>
<td>62</td>
</tr>
<tr>
<td>3. PROCEDURES OF THE STUDY</td>
<td>63</td>
</tr>
<tr>
<td>SELECTION OF THE SAMPLE</td>
<td>65</td>
</tr>
<tr>
<td>SELECTION AND ADMINISTRATION OF THE INSTRUMENTS</td>
<td>66</td>
</tr>
<tr>
<td>DESCRIPTION OF THE INSTRUMENTS</td>
<td>67</td>
</tr>
<tr>
<td>THE EXPERIMENTAL TREATMENT</td>
<td>71</td>
</tr>
<tr>
<td>THE EXPERIMENTAL DESIGN</td>
<td>74</td>
</tr>
<tr>
<td>THE STATISTICAL PROCEDURES</td>
<td>75</td>
</tr>
<tr>
<td>4. FINDINGS OF THE STUDY</td>
<td>81</td>
</tr>
<tr>
<td>CRITERION REFERENCED READING ACHIEVEMENT</td>
<td>81</td>
</tr>
<tr>
<td>NORMATIVE REFERENCED READING ACHIEVEMENT</td>
<td>89</td>
</tr>
<tr>
<td>PUPIL ATTENDANCE</td>
<td>97</td>
</tr>
<tr>
<td>PUPIL MOBILITY</td>
<td>98</td>
</tr>
<tr>
<td>PUPILS WHO WERE TUTORED IN THE FIRST GRADE</td>
<td>99</td>
</tr>
<tr>
<td>SUMMARY OF THE FINDINGS</td>
<td>101</td>
</tr>
<tr>
<td>5. CONCLUSIONS AND RECOMMENDATIONS</td>
<td>103</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>READING ACHIEVEMENT</td>
<td>103</td>
</tr>
<tr>
<td>PUPIL ATTENDANCE AND MOBILITY</td>
<td>112</td>
</tr>
<tr>
<td>PUPILS TUTORED IN THE FIRST GRADE</td>
<td>113</td>
</tr>
<tr>
<td>RECOMMENDATIONS FOR FURTHER STUDY</td>
<td>114</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>117</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Summary of Program Mortality and its Effect on the Similarity of the Experimental and Control Groups, Criterion Referenced Reading Achievement</td>
<td>82</td>
</tr>
<tr>
<td>2.</td>
<td>Summary of Selection and Classification Test Data by Cell for Pupils Included in the Analysis of Criterion Referenced Reading Achievement</td>
<td>84</td>
</tr>
<tr>
<td>3.</td>
<td>Summary of Cell Sizes, Means, and Variances for Analysis of Criterion Referenced Reading Achievement</td>
<td>86</td>
</tr>
<tr>
<td>4.</td>
<td>Summary of Analysis of Variance for Criterion Referenced Reading Achievement—Harper &amp; Row First Reader &amp; Second Reader Achievement Tests, Composite Raw Score</td>
<td>87</td>
</tr>
<tr>
<td>5.</td>
<td>Distributions of Criterion Referenced Reading Achievement Data for High Perceptual Reasoners and Low Perceptual Reasoners Showing Rank and Treatment Condition and Mann-Whitney U-test Results</td>
<td>88</td>
</tr>
<tr>
<td>6.</td>
<td>Summary of Program Mortality and its Effect on the Similarity of the Experimental and Control Groups, Normative Referenced Reading Achievement</td>
<td>90</td>
</tr>
<tr>
<td>7.</td>
<td>Summary of Selection and Classification Test Data by Cell for Pupils Included in the Analysis of Normative Referenced Reading Achievement</td>
<td>91</td>
</tr>
<tr>
<td>8.</td>
<td>Summary of Cell Sizes, Means, and Variances for Analysis of Normative Referenced Reading Achievement</td>
<td>94</td>
</tr>
<tr>
<td>9.</td>
<td>Summary of Analysis of Variance for Normative Referenced Reading Achievement—Cooperative Primary Reading Test, Form 23A</td>
<td>95</td>
</tr>
</tbody>
</table>
Table

10. Distributions of Normative Referenced Reading Achievement Data for High Perceptual Reasoners and Low Perceptual Reasoners Showing Ranks and Treatment Condition and Mann-Whitney U-test Results .............. 96

11. Summary Test Data for Pupils Tutored in Both the First and Second Grades (Group A) and in the First Grade Only (Group B.) .......... 100
Chapter 1

INTRODUCTION TO THE STUDY

On September 23, 1969, James E. Allen Jr., former United States Commissioner of Education, speaking before the 1969 Annual Convention of the National Association of State Boards of Education in Los Angeles stated:

We should immediately set for ourselves the goal of assuring that by the end of the 1970's the right to read shall be a reality for all—that no one should be leaving our schools without the skill and the desire necessary to read to the full limits of his capability (Report on Education Research, 1969, p. 3).

If this "right to read" goal is to be achieved, it is the responsibility of today's educators to seek out and develop viable reading programs for the youth of this country.

Development in this context does not mean the attractive packaging and commercial promotion of any particular program or method. Rather, it means the careful construction of a program that is based upon sound educational principles and is logically congruent from the purpose through the activities. This logical congruence must be verified empirically, and only after empirical congruence has been established is the program ready for use. Even then, use is limited to individuals or groups comparable to those with which the program was verified.
If one subscribes to the notion of differential learning styles for different children, it seems obvious that a number of programs and/or techniques must be developed to achieve the "right to read" goal. As this development proceeds, it should be guided by the following principle: "Evaluation is best looked at as a form of educational intelligence for the guidance of curriculum construction and pedagogy (Bruner, 1966, p. 163)."

The last few years have brought an influx of categorical federal aid to education. During the 1969-70 school year California was allocated $96,870,756 under Title I of the Elementary and Secondary Education Act (California State Department of Education, 1971). One of the mandated components for projects funded under the Elementary and Secondary Education Act was language development, including reading. The increase in federal funding has led to numerous attempts to improve reading achievement, particularly for children from the disadvantaged areas. However, Ruth Holloway, director of the Right to Read Program in the U. S. Office of Education, in a recent interview indicated that there were some seven million elementary and secondary school children who were reading at least two grades below their abilities (Right to Read, 1971).

Statement of the Problem

Substantial numbers of children from low-income
areas do not learn to read sufficiently well to enjoy success in school. There is no published empirical evidence regarding the use of programed tutoring in reading at the second grade level. The impact of programed tutoring on the reading achievement of second grade children must be demonstrated and not inferred from research at the first grade level. This is true, not only in terms of criterion referenced instruments related to the tutoring materials, but in terms of normative referenced instruments designed to sample reading skills. Does programed tutoring increase reading achievement, affect children differentially with varying perceptual reasoning abilities, and relate to pupil attendance and mobility?

Purpose of the Study

Programed tutoring, a technique developed and field tested by the Psychology Department of Indiana University under the leadership of Dr. Douglas G. Ellson, has shown promise as an effective supplement to the teaching of reading for first grade children from low-income areas (Ellson, Harris, & Barber, 1968). On the basis of this research, other research by Ellson and his associates (1968, 1969), and research conducted by the Sacramento City Unified School District (1969, 1970), programed tutoring was extended into the second grade in the Sacramento City Unified School District during the 1970-71 school year.

Since the research cited was limited to first
grade children, this extension of the program into the second grade represented an experimental effort. In keeping with the earlier comment from Bruner, this experiment must be subjected to evaluation.

The research conducted by Ellson and his associates (1968, 1969, 1970) has shown that programed tutoring was most effective with lower achieving children, but the research conducted by the Sacramento City Unified School District (1969, 1970) has shown that even among lower achieving children, all did not benefit equally. Factors relevant to such differential benefits have not been previously investigated. The nature of the programed tutoring technique, a series of small steps presented in a fixed sequence, suggested to the investigator a low degree of compatibility with the learning style of the child with high perceptual reasoning ability, but a high degree of compatibility with the learning style of the child with low perceptual reasoning ability. This dimension has not been previously investigated.

Casual observations by personnel in schools that have been involved in programed tutoring activities have suggested relationships between the tutoring and pupil attendance and between the tutoring and pupil mobility. The existence of such relationships has not been previously investigated. If these relationships do exist, they may be indicators of changing attitudes toward school and education on the part of children and/or their parents.
This study has been conducted to investigate the impact of programed tutoring on the reading achievement of lower achieving second grade children from low-income areas; to investigate the differential benefits relative to perceptual reasoning ability; and to investigate the existence of relationships between programed tutoring and pupil attendance and between programed tutoring and pupil mobility.

Rationale for the Study

No skill in education is more fundamental than reading. It remains the chief means by which anyone can continue his education independently long after his school days have passed (Educational Policies Commission, 1960, p. 10).

These words provided the basic rationale for this study, as well as all studies concerned with the teaching of reading.

Techniques must be devised that will aid the lower achieving children in the elementary schools to develop their reading potential to the fullest. Chall states that, "No program can do all things for all children, and no program can be all things for all teachers (1967, p. 310)." This statement suggests the need for the development of a variety of techniques and the investigation of the effectiveness of these techniques.

tutoring may be an effective supplemental technique for teaching reading to lower achieving first grade children. This research consistently showed significant differences favoring the tutored children on criterion referenced instruments. There were, however, inconsistencies in this regard with normative referenced instruments. These inconsistencies coupled with the lack of research regarding the use of programed tutoring with second grade children served as the major justifications for this study.

If programed tutoring can be shown to be an effective supplement to the classroom teaching of reading, many children will be given opportunities for success that they may not have otherwise enjoyed. These benefits would not be limited to the children and their parents, but would be shared with society as well. Programed tutoring also has the potential of providing the school with previously untapped instructional resources. Parents, volunteers, and other pupils, as well as salaried paraprofessionals are all potential candidates for service in the tutoring role. Service by individuals from these groups would not only increase their contributions to society, but would serve to bring the schools and the community closer together.

This study was further justified and was of educational importance in that programed tutoring can be replicated and transported to second grade children throughout the state and nation. If the efficacy of this
technique can be demonstrated, it will contribute to the achievement of the "right to read" goal expressed by Allen.

**Definition of Terms**

The following definitions of terms were used throughout this study:

**Programed tutoring.** Programed tutoring is a form of individualized instruction given fifteen minutes daily as a supplement to the classroom teaching of reading. It is designed to be carried out effectively by paraprofessionals of limited educational background. The tutor's teaching activities are tightly prescribed by (a) detailed instructions (programs) which they follow to the letter, (b) teaching materials, and (c) success and failures of the children they tutor (Ellson, 1970). For this study, the teaching materials accompanied the Harper & Row basic reader series (1968).

**Lower achieving children.** Children who scored below the school median on the Harper & Row Second Year Readiness Test at the beginning of the school year.

**Low-income areas.** Areas that qualified for a "target area" designation under the guidelines for Title I of the Elementary and Secondary Education Act (ESEA) of 1965.

**Criterion referenced reading achievement.** Learning reflected by the composite total raw score on the Harper & Row first and second reader achievement tests at the end of
the school year.

Normative referenced reading achievement. Learning reflected by the total raw score on the Cooperative Primary Reading Test, Form 23A, at the end of the school year.

High perceptual reasoning ability. Intellectual capacity reflected by a raw score of 23 or higher on The Coloured Progressive Matrices, Sets A, Ab, and B.

Average perceptual reasoning ability. Intellectual capacity reflected by a raw score that is less than 23, but greater than 16 on The Coloured Progressive Matrices, Sets A, Ab, and B.

Low perceptual reasoning ability. Intellectual capacity reflected by a raw score of 16 or lower on The Coloured Progressive Matrices, Sets A, Ab, and B.

Pupil attendance rate. The actual number of days attended by a pupil between the first and last days of a program divided by the total number of school days between the first and last days of the program.

Pupil mobility. A pupil is termed mobile if he transfers from and does not return to a particular school during the school year.

Research Hypotheses

The question stated earlier in this chapter is now restated in the form of the research hypotheses tested in
this study. These hypotheses are restated in the null form in Chapter 4.

$H_1$. Lower achieving second grade children from low-income areas will demonstrate significantly greater criterion referenced reading achievement after one year of programed tutoring in the second grade than similar second grade children who have not been tutored in the second grade.

$H_2$. There will be a difference in criterion referenced reading achievement among lower achieving second grade children of varying perceptual reasoning abilities from low-income areas, with high perceptual reasoners demonstrating significantly greater achievement than average and low perceptual reasoners, and average perceptual reasoners demonstrating greater achievement than low perceptual reasoners.

$H_3$. Lower achieving second grade children from low-income areas who possess different perceptual reasoning abilities will achieve differentially under the programed tutoring treatment: high perceptual reasoners who are not tutored will rank higher in criterion referenced reading achievement than high perceptual reasoners who are tutored, and low perceptual reasoners who are not tutored will rank lower in criterion referenced reading achievement than low perceptual reasoners who are tutored.
H₄. Lower achieving second grade children from low-income areas will demonstrate significantly greater normative referenced reading achievement after one year of programmed tutoring in the second grade than similar second grade children who have not been tutored in the second grade.

H₅. There will be a difference in normative referenced reading achievement among lower achieving second grade children of varying perceptual reasoning abilities from low-income areas, with high perceptual reasoners demonstrating significantly greater achievement than average and low perceptual reasoners, and average perceptual reasoners demonstrating greater achievement than low perceptual reasoners.

H₆. Lower achieving second grade children from low-income areas who possess different perceptual reasoning abilities will achieve differentially under the programmed tutoring treatment: high perceptual reasoners who are not tutored will rank higher in normative referenced reading achievement than high perceptual reasoners who are tutored, and low perceptual reasoners who are not tutored will rank lower in normative referenced reading achievement than low perceptual reasoners who are tutored.

H₇. There will be a significant positive
biserial correlation between pupil attendance rates and treatment for lower achieving second grade children from low-income areas.

\[ H_0 \] There will be a significant positive tetrachoric correlation between pupil mobility and treatment for lower achieving second grade children from low-income areas.

Limitations of the Study

This study was limited to lower achieving second grade children from low-income areas who were tutored in the second grade by salaried paraprofessional tutors. The findings of this study may be generalized only to second grade populations in schools similar to those in the low-income areas of the Sacramento City Unified School District.

Overview of the Study

In this first chapter, the problem and purpose of the study have been stated, the rationale for the study has been presented, terms have been defined, the research hypotheses have been formulated, and the limitations of the study have been indicated. Chapter 2 presents a review of the literature related to this study. This review includes research relevant to programmed instruction in reading for primary level children (grades 1 - 3), tutorial programs in reading for primary level children,
and programed tutoring as a supplement to reading instruction.

Chapter 3 describes the procedures followed in conducting the study. This description includes selection of the sample, selection and administration of the instruments, description of the instruments, the experimental treatment, the experimental design, and the statistical analyses. The results of the study are presented in Chapter 4. Each of the hypotheses is restated in the null form, and the results of the statistical tests are presented in both narrative and tabular forms. The final chapter, Chapter 5, is devoted to interpretation and discussion of the results, conclusions, and recommendations for further study.
Chapter 2

REVIEW OF THE LITERATURE

One need not look far into the literature to find that there has been a substantial amount of research in the area of reading instruction. However, research dealing with programed tutoring as a supplement to reading instruction is less conspicuous. This lack of research in the area of programed tutoring is likely the result of two factors: the technique of programed tutoring is relatively new, and has drawn limited attention; and most of the research in the area of reading instruction has been concerned with comparing one method or approach of reading instruction with another (Chall, 1967), and has not been concerned with supplemental techniques such as programed tutoring.

This chapter presents a review of the research related to the use of programed tutoring as a supplement to reading instruction for lower achieving second grade pupils from low-income areas. The chapter is organized into three major sections: (a) programmed instruction, (b) tutorial programs, and (c) programed tutoring. The first section considers the theoretical basis for programed instruction and the impact of this method of instruction in the area of reading in the elementary
school. The second section deals with the underlying rationale for tutorial programs, and their impact on reading instruction in the elementary school. The final section considers the existing research on the technique of programmed tutoring. In this final section, programmed tutoring is traced from the developmental period, through the field test period, and into operational programs.

**Programmed Instruction**

Programmed instruction is one of a number of terms used synonymously for instructional methods that are capable of instructing effectively through presentation of an automated sequence of instructional segments without direct intervention by a teacher. Sidney Pressey is generally credited as being the first to give serious thought to the use of teaching devices in the classroom (Carr, 1962).

Pressey's work in the late 1920's and early 1930's dealt primarily with automatic testing devices, but he saw great potential for teaching-testing devices. His work did not gain popularity in the educational world, and Pressey attributed this to two major factors: the educational world was not yet ready for any such innovation; and this early work was done during the Great Depression and funds were not available for innovation, particularly one that might create further unemployment among teachers when thousands of teachers were already unemployed (Pressey, 1964).

The real impetus for the movement in programmed
instruction came from the work of B. F. Skinner.

Historically, the term "program" as applied to a sequence of instruction presented by a teaching machine, derives from the 1954 and 1958 papers of Skinner, whose influence has, directly or indirectly, guided the mainstream of developments in programmed instruction during the later 1950's and early 1960's (Lumsdaine, 1964, p. 382).

Theory underlying programmed instruction. Skinner's work was based upon a number of learning principles which had been demonstrated in the laboratory. These principles have been summarized in non-technical terms by Carr (1962) as follow: (a) learning takes place most rapidly if the student is actively engaged with the subject matter, (b) learning is most effective if the student develops the skills and knowledge in a form which will readily generalize to the "real life" situation for which they are intended (constructed responses), (c) learning takes place most rapidly if immediate "knowledge of results" is given for each response, (d) learning takes place most rapidly if the subject matter is organized in a hierarchic form, (e) learning is enhanced if the learner receives frequent "knowledge of results" by keeping him working at the assigned task, and (f) since learning takes place in individuals, the learning situation should be designed so that each student may proceed at his own pace.

These principles have been and continue to be a source of controversy. The research that has been conducted relative to programmed instruction over the last 15 years has dealt primarily with these principles, along with
comparative studies relating programmed instruction to "conventional" methods of instruction. Before considering this research, it is interesting to note that there are several basic differences in the positions taken by Pressey and Skinner regarding programmed instruction.

Differences between Pressey and Skinner. The basic differences in the positions taken by Pressey and Skinner relative to programmed instruction have been summarized by Fry (1960). The first major difference relates to the type of question used in the program. Skinner prefers the type of question that requires a constructed response, while Pressey prefers the multiple-choice type of item. Closely related to this matter is the amount of error desirable. Skinner prefers a minimal amount of error in student responses, or better yet, no error at all. Pressey is more tolerant, although he feels that the number of correct responses should greatly exceed the number of incorrect responses.

In terms of usage of programmed instruction, Skinner supports the replacement of classroom instruction with programmed instruction, while Pressey sees programmed instruction in more of a supplemental role. Relative to the intelligence of the student, Skinner views the number and order of steps in the program as the critical features. His approach to programmed instruction is linear in the sense that there is a fixed order for the steps in the program although the size of the steps may vary as a function
of the intelligence of the learner. Pressey places no stress on the size or order of the learning steps. This flexibility allowed by Pressey led to the development of branched or scrambled programs that provide alternative sequences as a function of the learner's response.

Differences of opinion such as these have paved the way for much of the research relative to programmed instruction. The next section considers the research related to the theoretical basis of programmed instruction.

Research related to the underlying theory. In 1964, Schramm (1964) reviewed the research that had been reported between the time of Skinner's article of 1954, "The Science of Learning and the Art of Teaching," and February of 1963. His review located 190 reports of original research. The large majority of these reports (165) dealt with the principles of learning that guided Skinner's approach to programmed instruction.

Schramm's review indicated that the matter of the sequencing of steps had not been settled. Similar conclusions were reached regarding the size of the steps in the programs and the amount of error desirable. He also reported that the great majority of studies found no significant difference between the amount of learning from constructed and selected responses. Schramm reported that the majority of the studies supported the idea that immediate knowledge of results contributes to learning, but
the evidence on self-pacing had not demonstrated the advantages expected.

The experiments reviewed had not been too successful in identifying the incentive in programmed instruction that makes response-confirmation an act of reinforcement. In short, Schramm's review indicated that there were still questions regarding the learning principles underlying programmed instruction. Probably the most significant finding from his review was that students do learn from programmed instruction.

They learn from linear programs, from branching programs built on the Skinnerian model, from scrambled books of the Crowder type, from Pressey review tests with immediate knowledge of results, from programs on machines or programs in texts (p. 3).

From the research available to Schramm, however, he concluded that the question of how well students learn from programs as compared to other types of instruction could not be answered so confidently. This matter, as it relates to reading instruction, will be considered later in this chapter.

The research subsequent to Schramm's review continued to pursue the same types of questions relative to the learning principles underlying programmed instruction. This is evident from some of the more recent research on programmed instruction.

Rosenstock, Moore, and Smith (1965) studied four schedules of knowledge of results with sixth grade pupils
using a constructed response, linear program on set theory. The pupils were randomly assigned to one of four knowledge of results conditions: 100% or answer provided for every frame; 20% fixed with answers provided for every fifth frame; 20% variable with answers provided for one-fifth of the frames, randomly selected; and 0% or no answers provided. The results of this study indicated that the amount achieved on a posttest was not affected by the knowledge of results condition. The groups were tested for retention two weeks later, and there were still no differences among the four groups.

Knowledge of results is considered to be a form of intrinsic reinforcement according to the principles advocated by Skinner (Carr, 1962). Extrinsic reinforcement is reinforcement that is not built into the instructional materials, such as monetary rewards for performance. Sullivan, Baker, and Schutz (1967) investigated the effects of intrinsic and extrinsic reinforcement when both conditions were employed in the same instructional program. The study involved Air Force Reserve Officers Training Corps cadets using a program on the military justice system. The cadets were randomly assigned to one of four conditions: feedback and monetary reward contingent on performance, feedback and assured monetary reward, no feedback and monetary reward contingent on performance, and no feedback and assured monetary reward.

No significant differences were found among the four
groups on a criterion test. Further analysis of the data did, however, indicate that the feedback and no feedback groups developed different strategies to learn the material. The feedback group depended more on the instructional value of the feedback than the instructional material, while the no feedback group expended more time and energy perusing the instructional material. Relative to the monetary rewards, it was concluded that the amounts of money offered were not of sufficient strength to affect the cadets' performance.

In a subsequent experiment using cadets with the same instructional material (Sullivan, Schutz, & Baker, 1971), delayed feedback was substituted for no feedback, release from drill time was substituted for monetary reward, and the release from drill time was made contingent upon en route performance as well as terminal performance. Two types of en route mastery tests were used, unit and cumulative.

Immediate feedback was found to be more effective than delayed feedback on the criterion test, but the type of en route mastery test and extrinsic reinforcement produced no significant differences in criterion test performance. Comparison with the criterion test results of the previous study (Sullivan et al., 1967) indicated that release from drill time was a stronger reinforcer than the small monetary rewards offered in the 1967 experiment.
Davis, Marzocco, and Denny (1970) investigated the feedback versus no feedback question with college students using an instructional program covering two units from an introductory psychology course. Students were randomly assigned to the feedback and no feedback conditions. A comparison of the pretest-posttest differences revealed no significant difference in the two feedback conditions.

Another aspect of this study was the investigation of relationships between the pretest-posttest difference scores for the two feedback conditions and student abilities in English, reading, verbal ability, and arithmetic. No significant correlations were found with either feedback condition.

Murphy (1970) and Pfau (1970) conducted studies with deaf children to consider the effects of different schedules of knowledge of results with programmed instruction. Murphy used a sight association program to teach the eight seas of the moon to deaf children in grades 7-12. Using four levels of knowledge of results ranging from complete to no knowledge of results, he found no significant differences among the four groups in recognizing the new words. Murphy did find, as might be expected, a lower error rate per frame for the no knowledge of results group as they had only one opportunity to respond to each of the program frames.

Pfau's findings using a sight association program dealing with unfamiliar animals were essentially the same
as those of Murphy. Pfau's subjects were deaf learners 11-16 years of age. He also found that the group receiving no knowledge of results took less time to complete the program than those under the reinforcement conditions.

From further analyses of the learners and the program, Pfau suggested several implications: (a) if immediate knowledge of results is provided, the increment from one step to the next in the program should be sufficient to challenge but not frustrate the learner; (b) conversely, if no knowledge of results is provided, the increment should be small enough to give the learner confidence as he proceeds through the program; (c) if the material has no hierarchical structure, knowledge of results may not be important; (d) if programs are over-cued, little learning takes place; and (e) if response confirmation is redundant, it may do little to alter learning.

Two of the assumptions underlying programmed instruction, knowledge of results and logical item sequence, were studied by Jacobs and Kulkarni (1966) using junior high school mathematics and senior high school chemistry programs. Each of the programs was the constructed response, Skinnerian type of program. Students were randomly assigned to one of three versions of their respective programs: the regular version of the program, sequenced with knowledge of results; a modified version, sequenced with no knowledge of results; or a second modified version, inverted sequence of some items.
Comparison of the posttest scores in the junior high school program showed no significant differences among the three groups. However, in the senior high school program students using the regular version were found to be inferior to those using the two modified versions in terms of the posttest scores. There were indications that the students using the modified versions were required to work harder than the students using the regular version. The conflicting findings in this study bring two additional variables into focus, the subject matter and the student populations.

Maier and Jacobs (1966) employed a linear, self-instructional program to teach Spanish reading and writing to sixth grade pupils. Two versions of the program were used. The first version was a carefully designed and tested program based on Skinnerian principles. In the second version, the frames were scrambled by a group of experts on programmed instruction on an intuitive basis to enhance student interest in the program.

Variables under consideration in this study included achievement, interest in the subject area, and attitudes toward programmed instruction. Classes of sixth grade pupils were randomly assigned to the two versions of the program. The findings of this investigation revealed no significant differences between the two groups in terms of achievement, interest in the subject matter, or attitudes toward programmed instruction.
The matter of logical versus scrambled sequencing of items in a program might be considered to be a function of the hierarchal structure of the subject matter. Payne, Krathwohl, and Gordon (1967) investigated this dependence using three programs dealing with basic concepts in educational measurement with college students. The three programs were judged to be at different points on a hierarchal structure continuum. Students were randomly assigned to one of eight groups who worked through all three programs. The eight groups represented the eight permutations of logical (L) and randomly scrambled (S) sequences of the three programs (LLL, LLS, LSS, ..., SSS).

Immediate and delayed posttests covering all three programs indicated the eight groups did not differ in either immediate acquisition or retention of the measurement concepts. It was anticipated that there would be a relationship between student ability and performance, with the more able students better able to handle the scrambled sequences in the programs. This relationship failed to materialize.

Using a mathematics program that had shown frame dependency, Brown (1970) investigated the effects of scrambling this program with high ability high school students in trigonometry classes. The variables under study were time to complete the program, errors made during instruction, scores on en route and criterion tasks, and attitudes regarding the logical and scrambled versions of
the program.

The findings of this investigation indicated that
the logically sequenced group took less time to complete
the program, made fewer errors during the instruction, and
scored higher on the criterion test. There were no reli-
able differences between the logical and scrambled groups
on the en route tests or attitudes toward the two versions
of the program.

Cartwright (1971) considered the matter of sequenc-
ing of steps with mentally retarded adolescents using a
program designed to teach fractions to fifth grade students.
The reading level of the program was modified for the popu-
lation included in the experiment, and the program was
expanded to fill in gaps in the logical arrangement. The
scrambled version of the program modified the order of
items within sections, but the order of the sections was
not altered. Criterion, retention, and transfer tests were
used to investigate the performance of the retarded stu-
dents on the two versions of the program.

Analysis of the data obtained in this study indi-
cated that the students under the two sequencing conditions
did not differ on the tests of immediate learning, reten-
tion or transfer. From this study, Cartwright speculated
that an unsystematic, unordered program places more de-
mand on the student's general intellectual abilities, while
a systematic, ordered program places more demand on the
student's specific abilities.
A recent study on the size of steps in programmed instruction (Furukawa, 1970) questioned the use of small steps to obtain maximal achievement. Furukawa used a program designed to teach the spelling and meaning of Hawaiian words with female college students. He conducted this study to investigate optimal step size and its relationship to the individual's short-term memory ability. The "chunking" concept was used, where a chunk was defined as a word to be recalled to respond to a question. Measurements were taken during the instruction, immediately following the instruction, and two days after completion of the instruction. For this experiment, 1-, 2-, 7-, 14-, and 21-chunk step sizes were considered.

The findings of the study indicated that those subjects with high short-term memory scores generally performed better than those with low short-term memory scores, regardless of chunk size in the program. In terms of chunk size, 7 and 14 chunks were found to be most effective, leading Furukawa to the conclusion that for this type of learning task, programmed instruction steps need not consist of one or two sentences, but may consist of pages of text material containing 7 to 14 chunks of information related to the objectives of the program.

Eisman (1970) investigated the effects of overt and covert responding to programmed instruction using naval personnel with programs relative to rote learning, problem solving, learning relationships, and learning foreign
phrases. The naval personnel were randomly assigned to one of two response conditions, overt or covert. The overt responders wrote their answers when directed to do so in the program, and the covert responders were instructed to think their answers. The results of this study indicated that overt responding was advantageous for rote learning, but there was no difference in overt and covert responding relative to problem solving.

Employing an algebra program with college students in a remedial course, Davis, Marzocco, and Denny (1970) investigated overt versus covert responding and the matter of constructed versus multiple-choice responding. The subjects were given an opportunity to experience all possible treatment combinations, and then approximately one-third were given their choice of treatment combination, thereby introducing another variable, choice versus no choice.

Using three measurements, section midterm examination, section final examination, and department final examination, the treatment preference variable was found to have no significant effect. The choice and no choice subjects were then pooled for analysis of the two response variables. On the basis of the same three measurements, the analyses indicated that there were no significant differences in performance for the overt-covert response variable or for the constructed-multiple-choice response variable.
Summary. One generalization that is apparent from the research reviewed regarding the underlying principles of programmed instruction is that it is difficult, if not impossible to generalize. Programming variables appear to be interrelated with learner variables, population variables, subject matter variables, the cognitive levels within a subject matter, and perhaps other variables thus prohibiting broad generalizations regarding programming variables alone.

Programmed Instruction and Reading

Turning to the matter of programmed instruction as it bears directly on the problem considered in this investigation, reading instruction for young children, the research findings are generally similar to those discussed previously for the principles of learning, inconclusive. Until relatively recently, little research had been reported comparing programmed instruction and "conventional" methods for the teaching of reading to young children. The review conducted by Schramm (1964) included no such comparative studies. On this matter, Fry (1969) has commented:

If one were to search the literature to answer our question, "How effective is programmed instruction in the teaching of reading?", he would be hard put to find much serious research bearing directly on this question (p. 194).

Fry singled out Ruddell's study published in 1965 as the biggest and best controlled study of the teaching of reading by programmed instruction. Ellson (1969a)
supported Fry in this judgment.

Programmed versus basal materials. Ruddell's study was conducted during the 1964-65 school year with 24 first grade classrooms representing a wide range of socioeconomic levels (Ruddell, 1966). The classrooms were stratified on the basis of the socioeconomic levels of the neighborhoods using the 1960 census report. The classrooms from each stratum (high, middle, and low) were randomly assigned to four treatment conditions. The treatment conditions were reading programs.

Program B was the Sheldon Basic Reading Series, 1957; and Program B+ was the same as Program B supplemented with materials developed by the investigator and designed to stress language structure related to meaning. Program P was McGraw-Hill's Programmed Reading, 1963; and Program P+ was the same as Program P supplemented in the same way as Program B+. All of the groups were taught by classroom teachers with 60 minutes devoted to the reading program each day.

In terms of standardized test raw score means, Ruddell found Programs P and P+ superior to Programs B and B+ in word reading, word study skills, and regular word identification as hypothesized. The findings relative to irregular word identification were inconclusive as were those relative to sentence and paragraph meaning. The findings did, however, suggest to Ruddell a trend favoring
the supplementary programs (B+ and P+) on the sentence meaning variable.

During the 1965-66 school year this study was extended into the second grade on a longitudinal basis (Ruddell, 1967). At the end of the second grade, Program P+ was found to be superior to Program B+ in terms of word reading, regular word identification, and irregular word identification; but Programs P and B did not differ on any of the criterion tests. The findings were again inconclusive relative to paragraph meaning, and no differences were found in sentence meaning.

Another 1964 study (Harris, 1964) compared the McGraw-Hill Programmed Reading materials with a developmental reading program and a basal phonetic program. The experimental subjects were matched groups of second grade children reading below grade level. Matching criteria included chronological age, mental age, IQ, and reading level. The groups employing the programmed materials and the developmental program were taught by the investigator, and the basal phonetic group was taught by a classroom teacher. The period of instruction covered 60 days, one hour daily.

The findings of this study, in terms of gain scores controlled for pretest scores, revealed no significant differences among the three groups on the Gates Reading Test, but the programmed and developmental groups were favored on the California Reading Test. Harris also found
girls achieving significantly higher gains than boys in reading comprehension with the California Reading Test and the Gray Oral Reading Paragraphs Test. Language mental age was found to be a factor in determining comprehension gains on the California Reading Test, and age was a factor in determining gains in oral reading. All other findings were found to be non-significant.

Wollenberg (1968) used a matched pair design to compare reading achievement for culturally deprived first and second grade students instructed with Sullivan's Programmed Reading Series and those instructed with the Scott Foresman basal reading materials. The study also investigated differential achievement between the high achievers at each grade level under the two treatment conditions, and between the low achievers at each grade level under the two treatment conditions. Matching criteria for the matched pairs included sex, mental age, and reading achievement scores. All students were instructed for 90 minutes daily for 120 days, and were tested at the conclusion of the study with the Gates Primary Reading Tests. Wollenberg found no significant difference in reading achievement at either grade level for the groups instructed with the programmed and basal materials. In addition, there was no significant difference between the high achievers under the two treatment conditions, and there was no significant difference between the low achievers under the two treatment conditions.
A similar study comparing Sullivan's Programmed Reading Series and the Scott Foresman materials used with culturally deprived first grade students was reported by Hill (1968). The students in the study were matched pairs randomly selected from groups equated on sex, mental age, and reading achievement scores. Instruction under each of the reading programs was provided for 90 minutes daily for 140 days. This study also used the Gates Primary Reading Tests as the criterion measure at the end of the experimental period. Hill's findings differed from those reported by Wollenberg in that the group using the Sullivan materials scored significantly higher than the group using the Scott Foresman materials in word recognition and in sentence reading. There was no significant difference between the two groups in terms of paragraph reading.

Burkett and Clegg (1968) compared the effectiveness of McGraw-Hill's Programmed Reading and the Betts Basal Readers in a remedial reading program for "mildly" retarded readers at grades one, two, and three. Mild reading retardation was defined as a reading age of 0.5 to 1.6 years below the mental age. Silent reading behavior, oral reading behavior, and spelling proficiency were the variables under consideration in the study. Instruction was provided for 45 minutes daily for four months in a remedial reading room.

The findings of this study indicated no significant differences between the programmed and basal groups on any
of the dependent variables, silent reading behavior, oral reading behavior, and spelling proficiency. Burkott and Clegg concluded that, "Programmed Reading can take its place along with the basal reader approach in the repertoir of methods of teaching remedial reading (p. 748)."

Results of a three year longitudinal study of the Sullivan reading program in an inner-city school were reported by the School District of Philadelphia (Scheiner, 1969). This study was conducted with one class in one school, and the report recognized the limitations of the study. It is included in this review as a part of the existing body of research comparing programmed and traditional reading approaches.

This study covered the period from first through third grade, and assessments were made at the end of each year. Standardized tests were employed at the end of the second and third years. The test data indicated no significant difference between the programmed and traditional groups at the end of the second year, but a highly significant difference at the end of the third year favoring the programmed group. Different standardized tests were used for each of these two years.

One study was reviewed that sought to justify the use of programmed instruction for the initial teaching of reading on a benefit/cost basis (Burkett, 1970). This pilot study compared two reading programs using programmed materials, Borg-Warner's System 80 and McGraw-Hill's
Programmed Reading, and a traditional basal reading program. The investigation involved only five classrooms, one under each of the programmed conditions, and three under the basal condition. Program costs were based upon professional time expended and materials, and benefits were gauged by gains in reading achievement.

Using the traditional program as a base benefit/cost factor of 1.00, Burkett found the benefit/cost factor of Programmed Reading to be 1.82 and that of System 80 to be 1.62. Applying the same benefit/cost model to poverty sub-groups in the same sample, benefit/cost factors of 1.53 and 1.51 were found for Programmed Reading and System 80 respectively. Burkett reported that the benefit factors for the programmed instruction programs were greater than that of the traditional program, and the cost factors for the programmed instruction programs were less than that of the traditional program.

Supplemental programmed instruction. Hammill and Mattleman (1969) reported a study comparing three reading programs for second and third grade inner-city children. The three programs included programmed instruction exclusively, programmed instruction in conjunction with basal readers, and basal readers only. The pupils were low achievers in reading as determined by pretest scores, and were matched on the basis of achievement.

The findings of this study indicated no significant
differences among the three groups in terms of reading achievement at either grade level. In discussing the findings, Hammill and Mattleman pointed out the difficulty in comparing new instructional approaches with old approaches in terms of teacher preparation and background, as well as different teacher roles for approaches such as programmed instruction.

There are variables other than the instructional approaches to be considered. They also pointed out the weaknesses of standardized achievement tests as criterion measures for different approaches and materials, particularly in reference to vocabulary. Finally, Hammill and Mattleman raised the question of assumed equality when groups are matched on such variables as chronological age, IQ, and reading ability.

Another study that investigated programmed instruction as a supplement to a basal reading series for first grade children produced different results (Feldhusen, Lamb, & Feldhusen, 1970). In this study, the experimental group used McGraw-Hill's Programmed Reading as a supplement to the Ginn series, and the control group used only the Ginn series. The investigators noted that this study was limited in that the principal investigator taught the experimental classes, one class for each of three years; the amount of instructional time was not constant; and the experimental and control groups were not randomly selected. In addition to comparing the two treatment conditions, the
investigators considered sex differences and IQ differences. The findings reported by Feldhusen et al. showed that the experimental group scored significantly higher than the control group in word knowledge, word discrimination, and comprehension. There were no sex differences, and the high IQ group scored significantly higher than the low IQ group. In correlating 16 pupil variables with reading achievement, IQ, father's occupation and education, mother's education, and readiness test scores were found to correlate significantly with reading achievement. However, the greatest of these correlations, IQ, was only .40. In discussing the findings of the study, they noted:

There seems to be a continuing need for research and for improved testing procedures and techniques for assessing the relationship between IQ and reading achievement (p. 453)."

Other research and opinion. An interesting study regarding sex differences and early reading relative to programmed instruction was reported by McNeil (1964). The study did not compare programmed methods with conventional methods of instruction, but investigated differential effects on boys and girls under programmed and conventional methods. He used a program designed to teach 40 words to kindergarten children with schools representative of American upper-middle and lower-middle class communities. The children completed the program in individual cubicles with response and confirmation panels.

Boys were found to score significantly higher than
girls on the criterion test following program completion. Following this aspect of the experiment, the children received 40 minutes of instruction daily for four months from female classroom teachers using basal reading materials. A word recognition criterion test similar in format to the one following the programmed instruction was administered upon completion of this period of instruction. Following classroom instruction by the female teachers, the girls scored significantly higher than the boys, with 67% of the boys dropping in rank, and only 27% of the girls dropping in rank.

Further investigation of the children's perceptions of receiving negative teacher comments, and the teachers' assessments of the children's readiness and motivation for reading indicated that the teachers treated the girls more favorably than the boys. McNeil concluded that auto-instructional procedures afforded boys and girls equal treatment, and future study of the features of auto-instruction may be useful in developing teaching procedures more appropriate for boys than those in use. He also noted that the reduction in peer interaction resulting from the auto-instructional procedure used may have brought about the better male performance under auto-instruction.

Ellson (1969a) from his work and that of others sees a pattern relative to programmed instruction and the teaching of reading. He observed that programmed instruction seems to be more effective at the lower end of the
ability scale. Carr (1962), however, reported from his review of the literature that the findings indicated that the amount of profit from programmed instruction is independent of such factors as intelligence and aptitude. Wollenberg's study (1968) indicated achievement levels were not a factor in comparing programmed and basal reader materials in the teaching of reading to culturally disadvantaged first and second grade children.

The question of intelligence or ability relative to the effectiveness of programmed instruction as a method of teaching reading seemingly remains unanswered from the existing research. If one considers intelligence to be a factorial construct, perhaps specific factors of intelligence are relevant. No research was located that considered the factor of perceptual reasoning ability in programmed instruction in reading, thus this factor was included in this investigation.

A number of authors, including Ausubel (1965), Cheyney (1967), and Reissman (1962) have supported the use of programmed instruction with disadvantaged children as being appropriate for their learning style. Howards (1969), on the other hand, reported that most programmed material has proved to be of little value in his experience with disadvantaged persons.

Summary. The above observations along with the re-research discussed earlier in this section strongly suggest that we need to know more about the learner and programmed
instruction. "Researchers would do well to continue to search for characteristics of the learner which are correlated with amount of improvement (Carr, 1962, p. 76)."

Programmed instruction and automated teaching devices have their advocates and opponents, and one of the concerns among the opponents is the matter of human relations. While not represented here as an opponent of programmed instruction, Smith (1969) stated the concern quite well.

It is in the human relations area that the future teacher of reading has her supreme role to play. There are the human relationships between teacher and child for which no automatic device can substitute; the growing together in understanding of teacher and pupil; the encouragement and sympathy of a warm, friendly teacher; the satisfaction of a smile or nod of approval; the soft touch of a hand on the shoulder when one successfully completes a difficult learning task; and the personal stimulation of a teacher who believes in the pupil and expects him to do his best. These personal relationships are far more potent in a learning situation and more lasting in memory than skills or subject matter. Such interactive relationships of a human teacher with a living pupil must always supplement inhuman, impersonal, automations in teaching (p. 485).

One instructional technique that seems to provide abundant opportunities for such human relationships is the use of tutors in the school setting. The next section of this review considers tutorial programs.

**Tutorial Programs**

The concept of tutoring, one individual teaching another, as an instructional practice in all likelihood dates back to the beginning of education, formal and
informal. The monitorial programs developed by Andrew Bell and Joseph Lancaster in the late 1700's and early 1800's are among the first recorded tutorial programs in a formal school setting. These programs used the older, more knowledgeable students to "monitor" or tutor the younger, less knowledgeable students (Salmon, 1932). Lancaster brought the monitorial concept to America where it enjoyed some popularity until about 1850.

In modern times, however, tutoring has not been a widespread instructional practice relative to formal education until quite recently. The basic and most obvious reasons for the limited use of tutors in the schools have been the lack of resources, both human and financial, and the self-perpetuating nature of the educational enterprise. The last few years have seen a substantial movement toward tutorial programs, particularly for children in the low-income areas. This movement has been a part of the recent thrust toward the use of volunteers and salaried paraprofessionals in the educational system.

The rationale for tutorial programs. Bowman and Klopf (1969) have identified five convergent forces that have been catalytic in changing the structure of the educational system.

The gap between expanded needs for school services and the availability of professional personnel to meet these needs reached critical proportions in the late sixties;

New dimensions in educational concepts and technology required a more complex role for teachers;
Heightened awareness of the special learning needs of young children, and a developing insight into the communication blocks that often exist between middle-class professionals and disadvantaged children called for closer linkage of school and community;

The plight of the undereducated person, unable to compete in an increasingly automated society pointed to the need for a new entry level to careers of human service with opportunity for upward mobility on the job;

Finally, and most importantly, new resources became available to school systems through OEO, MDTA, Title I of ESEA, the Nelson-Scheuer Amendment to the Poverty Act, and the Javits-Kennedy Act for Impacted Areas, all of which provided Federal funds for the employment of low-income persons who lacked the traditional certification for education (p. 6). These forces have led to the employment of many paraprofessionals and the use of many volunteers in the public schools of America.

The Institute for Educational Development (1970) in conducting an in-depth study of paraprofessionals for the New York City Schools surveyed the professional literature to ascertain the generally-held purposes of paraprofessional programs. This review identified five target populations served by such programs.

Pupils will be given more individual attention in the classroom. Moreover, pupils will be placed into contact with capable, respected members of their own ethnic groups so that pupils can model their behavior after the behavior of these respected persons. As a result, it is hoped that attitudes of pupils toward school will improve and their achievement will increase.

Teachers will become better able to manage teaching conditions and pupil behavior in their own classrooms, and will thereby affect pupil achievement. Moreover, teachers will relate better to parents and community members in inner-city settings and by doing so will indirectly influence pupil achievement.
Principals will fulfill their long-standing need to increase the size of the school staff and the quality of school services in inner-city schools. In addition, principals will strengthen and extend their relations with parents and with the surrounding communities.

Parents will establish more effective linkages with the schools because they will be better able to relate to neighborhood paraprofessionals than to other school personnel, many of whom live elsewhere and come from a different ethnic and cultural background. In addition, parents will profit from the assistance of paraprofessionals who give them counsel while serving as semi-professional social workers.

Paraprofessionals will profit from the income they realize through employment; will improve their self-respect through carrying out important tasks, will increase their education through training associated with paraprofessional work, and will begin a pattern of upward career mobility (pp. 7-8).

These purposes are essentially the same as those indicated by Bowman and Klopf (1969) and Janowitz (1965). In addition to the purposes for paraprofessional programs noted above, and following the premise that children are the primary target populations of the schools, Gordon's comments (1970) regarding the educability of populations where there is deprivation of developmental and educational opportunity seem appropriate.

In order to be maximally meaningful to a child, education must be relevant in three areas: 1) it must relate to him as an affective being through its materials, experiences, and people with which he can identify; his motivation to learn will be more easily tapped when the learning task leads to goals which he perceives himself as valuing; (2) the content and form of the learning experience must be suited to his cognitive style and temperamental characteristics, and must complement his stage of cognitive development; this implies a sensitive determination of the curriculum to be presented as well as the manner in which it is offered to the child; and (3) it must have social or utilitarian relevance; i.e. it must offer those
skills and competencies which will expand the realm of functional choice available to the child (p. 264).

These remarks by Gordon, as well as the purposes of paraprofessional programs stated earlier, seem to be appropriate to and provide the basic rationale for tutorial programs in the American educational system today. In particular, tutorial programs for the disadvantaged in the area of reading instruction seem to be justified.

Tutorial programs for reading instruction have been devised using paraprofessionals as tutors. In addition, tutoring programs have been devised using students, both high school and upper elementary, as tutors. Considering the problem addressed in this study, what effects have tutoring programs had on the reading achievement of lower achieving elementary school children?

While there is a substantial body of literature relative to tutorial programs, Erickson (1971) noted that few tutoring programs have been examined objectively, and their popularity is too often based upon subjective evaluation. The situation is similar to that reported by Riessman and Gartner (1969) regarding the use of paraprofessionals in the schools.

The bulk of the current literature on the use of paraprofessionals in the schools focuses upon their activities in the classroom, their selection, training and compensation, and inter-action between teacher and aide. Less documented is the relationship of the aide to pupils performance (p. 7).

High school students as tutors. One study frequently
cited is that conducted by Cloward (1967) in which high school students tutored fourth and fifth grade students whose reading achievement was below grade level. The tutoring was conducted after the school day with each tutor seeing his student one or two afternoons a week, after which the tutor escorted his tutee home. Each of the tutoring centers was supervised by a teacher. The tutors were salaried and participated in preservice and inservice training that focused on the goals and organization of the program, tutor duties, tutee characteristics, the curriculum, techniques, and materials. As the program developed,

The typical tutoring session consisted of 30 minutes spent on homework, 30 minutes on reading, 15 to 30 minutes on games and recreation, and 15 minutes for refreshments, roll-taking and other non-tutorial duties (p. 15).

Students were randomly assigned to the tutored and control groups, and the reading achievement of both groups was assessed at the beginning of the program and five months later. Other assessments were made in the areas of school marks and behavior, attitudes and aspirations, and the tutor-pupil relationship.

The experiment also investigated effects on the tutors. Eligible tutor applicants were randomly assigned to the tutor and control groups, and the reading achievement of each group was assessed at the beginning of the program and seven months later. Other assessments included school achievement, attitudes, and aspirations.

The findings of the study relative to tutees
indicated that tutorial assistance resulted in significant reading improvement when provided for at least four hours per week (two sessions) for a period of 26 weeks. The group tutored two hours per week (one session) showed greater improvement than the control group, but this difference was not found to be statistically significant. No differences were found between the tutored and control groups in terms of school marks, school behavior, attitudes, or aspirations. Regarding the tutor-pupil relationship, Cloward reported an indication that sex-ethnic matching of tutors and tutees affected tutee reading achievement positively for Negroes.

Relative to the tutors, the findings of the study indicated that tutors demonstrated significantly more improvement than the control group on three subtests dealing with reading comprehension and study skills. No significant differences were found between the tutors and their control group in terms of school achievement, attitudes, or aspirations.

In discussing the findings of the study, Cloward noted that the gains in reading achievement made by the tutors and tutees may eventually be translated into increased achievement in other areas as they progress through school. From the study he concluded:

Tutors do not need 12 years of formal education and extensive training in reading pedagogy. Nor need they be highly successful in their own work. The average high school student can learn to be an effective tutor (p. 24).
Landrum and Martin (1970) reported on a six-weeks summer tutorial program in which high school students who were reading two or more years below grade level and were dropout prone tutored fourth through sixth grade pupils who were behind in reading. This program was evaluated on the basis of performance objectives with the following criteria: (a) three months grade placement gain for tutees, (b) six months grade placement gain for tutors, (c) 50% reduction in absenteeism for tutors during the subsequent year, and (d) 95% of the tutors completing the subsequent year of school. The tutors were salaried and trained, and were supervised by a teacher-supervisor. Pupils were tutored for two hours daily for the six-weeks summer session.

Data relative to reading achievement were reported for the tutors and tutees for the summers of 1967 and 1968. In 1967, the tutors demonstrated eight months gain, and the tutees gained 4.6 months in reading achievement. In 1968, even greater gains were reported with the tutors gaining 8.5 months and the tutees gaining 4.8 months. All gains in reading achievement were in excess of the expected performance criteria. No data were reported regarding tutor absenteeism or their continuing in school.

An investigation which used ninth grade girls as reading tutors for primary children and provided some control over the personal contact variable was reported by Smith (1971). In this study the tutors used a structured reading program for one group of tutees and a placebo
program for another group of tutees. The placebo program involved reading related activities such as Frostig worksheets and reading to the child, but did not provide direct reading instruction. The tutoring was conducted for 11 weeks with the tutees completing an average of 42 thirty minute sessions.

Six criterion measures were used, three of which were directly referenced to the structured reading program, and three of which were not referenced to either the structured program or the placebo program. The results of the study showed no significant differences between the two groups of tutees on any of the criterion measures.

The mutual benefits of boys tutoring boys in reading was the primary focus of a study by Trickson (1971). In this investigation, seventh grade boys tutored third grade boys twice weekly for 30 minutes each session over a period of five months. Tutoring activities included ordinary conversation, oral reading, and language games. Variables under consideration included reading scores, grades, behavior, interests and attitudes, social acceptance, and attendance for both tutors and tutees. The results of the study indicated significant improvements in reading scores for both tutors and tutees when compared to control groups. No significant differences were found relative to any of the other criterion variables.

Elementary school students as tutors. A number of tutorial programs have been devised and implemented in
which older elementary school children tutor younger children. One such program (Rogers, 1970) investigated the effects of under-achieving sixth grade pupils tutoring under-achieving third grade pupils in reading. The tutors received three weeks of training, and then tutored for eight weeks during their regular reading period. Two control groups were used for the tutors: one control group participated in the training but did not tutor, and the other control group received no special treatment.

The tutees in the experiment received eight weeks of tutoring during their regular reading period, and a control group received no tutoring. The findings relative to the tutors indicated no significant differences in reading achievement among the tutors and the two control groups. However, the tutees made significantly greater gains in reading achievement than their control group. On the basis of subjective analysis and observations, Rogers reported the development of friendships, improved behavior in some tutors and tutees, and general enthusiasm regarding the program from teachers, parents, tutors and tutees.

Another project using fifth and sixth grade tutors was reported by Niedermeyer and Ellis (1971). In this study, kindergarten pupils were the recipients of the tutoring in the area of reading instruction. The investigators stressed the importance of training the tutors and using structured materials. One component of the study focused on the training variable. Tutors in this program
were trained to provide assistance to the kindergarten children using highly structured practice exercises related to the tutee's performance in the kindergarten reading program. The pupils were thus tutored on an "as needed" basis, with the teachers designating the appropriate practice exercises for each child.

The findings of this study indicated that the kindergarten classes that had the assistance of the tutors scored significantly higher than classes without such assistance on a criterion test. Tutor performance was also compared to the performance of untrained fifth and sixth grade pupils using the same practice exercises with kindergarten children. The results of this aspect of the study indicated that trained tutors displayed significantly more instructional behaviors than untrained tutors.

The reciprocal effects of sixth graders tutoring third graders in reading was the basis of a study by Paoni (1971). This experiment sought to establish mutual benefits for tutors and tutees in reading comprehension, vocabulary, and attitudes toward reading. The third grade students were tutored three days a week for 30 minutes each day for a period of four months.

The results of the study indicated that the third grade tutees demonstrated significantly greater gains in reading comprehension than a control group, and both tutors and tutees showed significantly greater gains in attitudes toward reading than their control groups. There were no
significant differences in reading vocabulary gains for either the tutors or tutees, nor were the tutor's reading comprehension gains different from those of their control group.

Somewhat different findings were reported by Snapp (1971) in an experiment in which fifth and sixth graders tutored first, second, and third graders in reading. All of the pupils in this study were from schools serving a disadvantaged population. Training and supervision of the tutors were considered to be of prime importance. Another variable considered in the study was that of reinforcement versus no reinforcement during the tutoring.

The findings of the study indicated that both the reinforced and non-reinforced tutees made significantly greater gains in word recognition than a control group, but there was no difference between the two groups of tutees. No differences were found between the tutors and their control group.

Adults as tutors. Turning to tutorial programs utilizing adults in the tutoring role, Klosterman (1970) investigated the effects of college students majoring in elementary education as tutors in a diagnostic and structured reading program. Fourth grade pupils from low socioeconomic areas were randomly assigned to individual tutoring, small group tutoring, or a control group. The investigator provided five hours of training for the tutors
and advised and supervised them throughout the treatment period.

Pupils were tutored for 30 minutes daily, four days per week for six months. The time was taken from the regular classroom reading period. The tutoring program was based upon tutor diagnosis of the child's reading, and was structured to include the teaching of concepts, vocabulary, word recognition skills, comprehension, and interpretation. Materials included workbooks, charts, library books, pictures, and audio-visual aids. Reading vocabulary, comprehension, and total reading achievement were the criterion variables.

The results of the study indicated that the pupils tutored individually made significantly greater gains than the control group on all three criterion measures, and those tutored in small groups made significantly greater gains than the control group in reading comprehension and total reading achievement. The differences between pupils tutored individually and in small groups were not found to be significant.

Schoeller and Pearson (1970) reported on a volunteer after-school reading tutoring program for fourth grade pupils reading one or more years below grade level. The tutoring was conducted in tutoring centers under the supervision of a reading specialist. Tutors were provided ten hours of inservice training related to understanding the nature of reading disabilities, understanding the pupil,
learning how reading takes place, and procedures to help pupils read better. Pupils were tutored for an average of one and one-half hours per week for an average of 3.1 months. The program actually operated over a period of 4.6 months, but pupils who entered late were included in the study.

The evaluation of the program compared gains in eight areas with an expected gain over the 3.1 month period. Since the pupils were achieving one or more years below grade level prior to the program, and the average IQ was 86.5, it was felt that any gain in excess of the 3.1 month treatment period would be a great improvement. The mean gains in the eight criterion areas ranged from a high of 9.2 months in letter sounds to a low of 3.6 months in oral reading. Thus, the gains in all eight areas exceeded the average 3.1 month tutoring period. Based upon the subjective evaluation of the program, Schoeller and Pearson reported that pupils' attitudes toward school, reading, and themselves improved.

Retention effects as well as immediate benefits were considered in the evaluation of a tutoring program for underachievers in reading and writing (Shaver & Nuhn, 1971). This investigation also dealt with a comparison of the benefits of individual and small group (three pupils) tutoring. Tutoring was provided for one hour daily throughout the school year to fourth, seventh, and tenth grade students who were achieving below their predicted potential. The
eligible students were randomly assigned to the two tutoring conditions and a control group. The tutors were adults selected on the basis of their reading and writing abilities, reading interests and activities, and an estimate of their ability to work with students. The tutors were given two weeks of training prior to the beginning of the program with emphases on understanding the underachiever, diagnosing reading and writing deficiencies, and providing assistance in an accepting atmosphere.

The results of the study indicated no significant differences between the pupils tutored individually and those tutored in small groups, with one exception. The tenth grade pupils tutored individually scored higher in reading achievement than those tutored in small groups. The comparisons for the pooled tutored groups and the control group indicated that the tutored groups scored significantly higher in both reading and writing across all three grade levels after one year of tutoring. These significant differences were maintained two years later for the seventh and tenth grade pupils, but not for the fourth grade pupils.

The evaluation of a second year of this program substantiated the immediate benefits reported for the first year of the program. From the study, Shaver and Nuhn concluded:

It seems evident that tutoring had a positive impact and that tutoring can take place effectively in more economical arrangements than
a 1-to-1 student-tutor ratio (p. 112).

One final study considering adults as tutors is an evaluation of the "Bridging-the-Gap" program in San Francisco (Falik & Wexler, 1971). This was strictly a descriptive report, and is cited here for the summary and conclusions which seem to capture a number of the salient points regarding tutorial programs.

A tutorial program represents the interplay of many complex variables, many of which can only be speculated upon by a descriptive study of this kind. Nonetheless, the tutorial program must be evaluated, and its strengths and weaknesses isolated. Our evaluation suggests that the program we studied was limited by the short duration of tutorial contact, the lack of training and ongoing assistance to basically untrained tutors, and the inability to get a "good fix" on what the tutee's real educational needs were. This leads to emphasis on the social-emotional dimensions of tutoring, and less focus upon academic achievement gains. Such a thrust is not without dividends, however, as the tutees seem to show improvement in many aspects of school and educational attitude, self-concept among them.

It might also be said that the weaknesses noted have with them implications for change--mainly the adding of resources and perspectives that would enable more emphasis on academic skill acquisition. Without such changes, this tutorial program--and others like it--are, in fact, programs to build interpersonal relationships, with gains in achievement due as much to serendipity as to design (pp. 376-377).

**Summary of tutorial programs.** The research regarding tutorial programs, while limited, seems to indicate benefits for both tutors and tutees in both the affective and cognitive domains. However, such benefits are not conclusively established at this time. This state of affairs relative to reading achievement may be due to
limited structure or lack of direction in the tutoring sessions and/or limited continuity and coordination with the regular classroom reading program. In discussing the pitfalls of employing nonprofessionals in the schools, Saltzman (1965) noted the following:

While it is possible to demonstrate that school districts can greatly expand their work force with little net additions to their budget, this argument cannot stand separate and apart from the issue of the educational program which may result. The place of each subprofessional needs to be carefully designed ... to insure that he fills an appropriate, useful role which is geared to fit in with the overall program of the school (p. 52).

Programed tutoring, the combining of components of programmed instruction and tutoring, seems to offer the potential structure and coordination with the classroom without dehumanizing the instructional process. The final section of this review deals with the existing research regarding programed tutoring.

**Programed Tutoring**

Programed tutoring is a form of individualized instruction ordinarily given 15 minutes daily as a supplement to classroom teaching of beginning reading. It is designed to be carried out effectively by paraprofessionals of limited educational background. Aides with less than a high school education can be trained to tutor effectively in 9-15 hours of group instruction supplemented by on-the-job supervision. Their teaching activities are tightly prescribed (a) by detailed instructions (programs) which they follow to the letter, (b) by teaching materials and, (c) not least important, by the moment-to-moment success and failures of the children they tutor (Ellson, 1970, p. 1).

*The developmental period.* The development of
programed tutoring began with a series of ten experiments carried out by Indiana University under grants from the United States Public Health Service (Ellson, Barber, Engle, & Kampwerth, 1965). These experiments involved children in a state school for retarded children, children in public school systems in Indiana, and associated laboratory settings. The ten experiments progressed from an attempt to teach sight-reading vocabulary to retarded children using pictures as prompt stimuli to a preliminary field test of programed tutoring as a supplement to classroom teaching using a content program defined by the Ginn Basal Reader Series.

In summarizing the results of these ten experiments, Ellson et al. noted the following: (a) retarded children showed relatively rapid acquisition of a reading vocabulary; (b) retarded children using a simple tutoring program could teach reading vocabulary effectively; (c) a sight-reading program that taught reading vocabulary in a sentence context through tutoring to slow readers, retarded and normal children had practical effectiveness; (d) programed tutoring in combination with standard classroom teaching was more effective than classroom teaching alone, and probably more effective than programed tutoring alone; (e) approximately daily alternation of programed tutoring and classroom instruction was more effective than less frequent alternation; (f) a "therapeutic" effect was noted; that is programed tutoring reinstated effects of earlier
learning; (g) children demonstrated abilities in the pro-
gramed tutoring situation that they did not demonstrate in
the classroom; (h) favorable effects can extend to atti-
tudes and behavior outside the classroom; (i) the optimal
duration of tutoring sessions appeared to be about 15
minutes; (j) in the form of programed tutoring, effective
individual teaching could be done efficiently by persons
with far less training than that required of a professional
teacher; and (k) programed tutoring seemed to function best
as a supplement to classroom teaching.

Some of these conclusions may certainly be ques-
tionable in light of the experiments that were conducted,
and no attempt was made to conceal this.

Some are experiments only in the primitive sense
of tryouts; in others, a carefully planned de-
sign was abandoned in midstream to permit informal
investigation of obviously important effects that
had not been anticipated in the planning (Ellson
et al., 1965, p. 79).

The field test period. The ten experiments did
yield sufficient evidence and motivation for Ellson and
his associates to refine the technique and proceed to the
field test and pilot stage of programed tutoring. The
tenth experiment of those conducted during the developmental
period, a preliminary field test of programed tutoring,
paved the way for a carefully controlled field test in-
volving 240 first grade children in 20 inner-city schools
(Ellson, Harris, & Barber, 1968).

The field test was designed to answer the following
specific questions: (a) since all of the tutored children in the earlier experiment were from one classroom, were conditions specific to that classroom confounded with the experimental variable? (b) was the superior performance of the experimental group due to tutoring or to one or more of these classroom conditions? (c) assuming that programed tutoring was effective, was the effect due to the programed tutoring or the individual attention that each child received? and (d) could similar results have been obtained with less tutoring?

Four groups, each initially containing 60 first grade pupils, were tutored throughout the school year. Two groups were given programed tutoring as a supplement to their classroom instruction, with one group receiving 15 minutes of tutoring daily, and the other group receiving 30 minutes of tutoring daily in two sessions. The other two groups were given "directed tutoring" as a supplement to their classroom instruction, with one group receiving 15 minutes of directed tutoring daily, and the other group receiving 30 minutes of directed tutoring daily in two sessions. Directed tutoring was a carefully selected set of activities supported with materials and planned so that the procedures could be carried out by persons with no more educational background, training, and supervision than that required by the programed tutors.

In summarizing the findings of this study, Ellson et al. noted the following: (a) directed tutoring had no
measured effects on reading achievement, (b) programed tutoring produced significant improvement in reading achievement test scores when given twice daily, but not when given once daily for the same period, (c) improvement in achievement was markedly greater for children who have the greatest difficulty in learning to read in the classroom, and (d) a number of defects in the programs were identified.

While the field test did demonstrate the effectiveness of programed tutoring in producing significant improvement in reading achievement, this improvement was confined to criterion referenced instruments. The results in terms of normative referenced instruments were not found to be significant.

Operational programs. Following this field test and utilizing the consistent finding that lower achieving children seemed to benefit most from programed tutoring, two more experiments were conducted involving children from the lower one-third of inner-city school populations (Ellson, Barber, & Harris, 1969a, 1969b). Both of these studies indicated that tutored children performed significantly better than non-tutored children. In these programs children receiving only one session of tutoring performed significantly better than non-tutored children, but not as well as children receiving two sessions of tutoring daily. Again, it should be pointed out that these studies were confined to first grade children, and the measurements of
reading achievement were made with criterion referenced instruments.

Programed tutoring, after a period of approximately nine years, was emerging as an effective supplement to the classroom teaching of reading for lower achieving first grade children from inner-city school populations. In 1969, the American Institute for Research in the Behavioral Sciences identified the Programed Tutorial Reading Project in Indianapolis as one of the 31 most successful compensatory education programs in the country (U. S. Department of Health, Education and Welfare, 1970).

During the 1968-69 school year, 1711 children in 39 schools received programed tutoring (Ellson, 1969c). Again, these were first grade children who ranked in the lower one-third of their school populations. The findings from this study confirmed the earlier results, and tutored children continued to show significant improvement in reading achievement. In this study, significant gains were shown on both criterion and normative referenced instruments. Ellson pointed out in this study that significantly fewer tutored pupils were retained in the first grade than non-tutored pupils.

Further studies conducted during the 1968-69 and 1969-70 school years (Ellson & Harris, 1970a, 1970b; Sacramento City Unified School District, 1969, 1970) continued to support the findings of the earlier research. Since the early work with the Ginn series, programed tutoring materials
and techniques have been extended to basal series other than the Ginn series, and the results have been equally effective. The research continues to support the earlier finding that lower achieving children benefit most from programed tutoring, and when higher achieving children were included, the results were less dramatic (Ellson & Harris, 1970a).

The findings regarding the effectiveness of programed tutoring at the first grade level received further support in terms of criterion referenced reading achievement in a study reported by McCleary (1971). This study also supports the previously mentioned finding of lower first grade retention rates for tutored children.

Summary of programed tutoring. All of the studies discussed above have been limited to first grade children with the exception of some of the early developmental experiments. Programs have been developed for use at the second grade level, and the question posed by Duker (1969) regarding individualized instruction seems appropriate here, "Is individualized reading equally effective at all grade levels (p. 448)?" The same question can certainly be posed for programed tutoring, and the existing research cannot provide an answer. The investigation reported in this paper should provide some evidence toward answering that question.
Summary of the Chapter

The research reviewed in this chapter indicates that while both programmed instruction and tutorial programs have proven beneficial to reading achievement in some cases, neither technique has clearly demonstrated superiority over traditional techniques. The evidence indicates that learner variables should be considered in devising instructional programs in reading, and that effects in areas other than reading achievement should be investigated.

Programed tutoring, which combines features of programmed instruction with a tutorial approach, has been successful in improving the reading achievement of lower achieving first grade children relative to criterion referenced instruments. The findings relative to normative referenced instruments, however, have not been conclusive. Programed tutoring remains as an untested program at the second grade level, and its effects in areas other than reading achievement have not been investigated. This study pursued these unexplored dimensions of the programed tutoring technique as a supplement to the classroom reading program.

The procedures employed in this investigation are described in the next chapter. This description includes the selection of the sample, the measures used, the experimental treatment, the experimental design, and the statistical analyses.
Chapter 3

PROCEDURES OF THE STUDY

The procedures employed in conducting this investigation are presented in detail in this chapter. These procedures are discussed under six main headings: (a) selection of the sample, (b) selection and administration of the instruments, (c) description of the instruments, (d) the experimental treatment, (e) the experimental design, and (f) the statistical procedures.

Selection of the Sample

The sample for this investigation was selected from the second grade classes of six elementary schools in low-income areas of the Sacramento City Unified School District. Five of these schools were public elementary schools (kindergarten - grade six), and the sixth was a parochial elementary school (grades 1 - 6). These schools were designated "target area" schools in the Sacramento City Unified School District, and were participating in a comprehensive program of compensatory education funded by special State and/or Federal resources.

The target area designations were based upon the positions of these schools in a composite ranking of the following characteristics: percentage of pupils receiving Aid to Families with Dependent Children (AFDC), percentage...
of pupils qualifying for free lunches, percentage of ethnic minority pupils, racial isolation, and standardized reading achievement test results.

The programed tutoring technique investigated in this study was funded under Title I of the Elementary and Secondary Education Act of 1965, and was included as a part of the comprehensive program of compensatory education. Funding and district regulations placed restrictions on the sampling procedures in that sufficient tutoring time was allocated to each of the six schools to serve one-third of the second grade population. These restrictions necessitated selection of the experimental and control groups on a school-by-school basis rather than from the total second grade population in the six schools. These regulations also placed a restriction on the sizes of the experimental and control groups. In order to limit the treatment to lower achieving children, and provide treatment to one-third of the total population, the control group was restricted to one-half the size of the experimental group at each school.

Observing the above restrictions, pupils were selected for the experimental and control groups from the lower achieving second grade children at each school on a school-by-school basis as follows:

1. The second grade children at each school were listed in rank order on the basis of the Harper & Row Second Year Readiness Test total raw scores obtained in
September of 1970.

2. The median raw score for each school was computed, and those pupils scoring below the median were designated as the lower achieving children. A test of the equality of the six medians (Siegel, 1956) revealed no significant differences in the six medians.

3. The lower achieving children at each school were then listed in alphabetical order and numbered.

4. A coin-flip was used to determine whether the experimental or control pupils would be selected first. Heads was designated experimental, and tails was designated control.

5. As a result of the coin-flip, the experimental pupils were chosen first using a random start in a table of random numbers (Arkin & Colton, 1950). Two-thirds of the lower achieving pupils at each school were selected for the experimental group. The remaining pupils at each school were assigned to the control group.

These selection procedures produced a total experimental group of 108 pupils and a total control group of 55 pupils. Further investigation of these pupils revealed that of the 108 experimental pupils, 18 had received programmed tutoring in the first grade; and of the 55 control pupils, nine had been tutored in the first grade. To avoid contamination, these 27 pupils were dropped from consideration in testing the hypotheses stated in Chapter 1. These 27 pupils were, however, considered descriptively in
Chapter 4.

Selection and Administration of the Instruments

One test, the Harper & Row Second Year Readiness Test (Harper & Row, 1968), was administered for the purpose of selecting the experimental and control groups. The test was administered to all second grade pupils in the six schools on a classroom group basis by the pupils' classroom teachers during the last week of September of 1970.

Three tests were administered to measure the effects of the programmed tutoring technique on the reading achievement of lower achieving second grade pupils. The Harper & Row First Reader Achievement Test (Harper & Row, 1968) and the Harper & Row Second Reader Achievement Test (Harper & Row, 1968) were employed to measure criterion referenced reading achievement. The total raw scores from these two tests were added together and used as a composite to provide adequate range for all pupils in the experimental and control groups. The Cooperative Primary Reading Test, Form 23A (Educational Testing Service, 1967) was employed to measure normative referenced reading achievement. This test was the California state-mandated reading achievement test for second grade pupils. These three tests were administered on a classroom group basis by the pupils' classroom teachers during the first two weeks of May of 1971.
One test, The Coloured Progressive Matrices, Sets A, Ab, B (Raven, 1965), was administered to stratify the second grade lower achieving children on the basis of perceptual reasoning ability. This test was administered on an individual basis to all experimental and control pupils by the investigator or his assistant (trained by the investigator) during the last two weeks of May of 1971. The investigator observed the assistant on numerous occasions during the testing procedure to insure quality control of the results.

Description of the Instruments

The Harper & Row Second Year Readiness Test used to select the experimental and control pupils is designed for use at the beginning of the second grade. The stated purpose of the test is to identify pupils who are least likely to succeed in the Harper & Row second year reading program. The test consists of six subtests, each of which yields a raw score, and the sum of these raw scores yields a total raw score. The six subtests include: word recognition; auditory-visual perception (initial consonants, clusters, and digraphs); auditory-visual perception (final consonants, clusters, and digraphs); auditory-visual perception (vowels); relationships; and comprehension. The total test contains 185 items. Reliability for the test was .98 using the Kuder-Richardson Formula 20. The standard error of measurement for the test was 5.88 (Harper & Row, 1968).
The Harper & Row First Reader Achievement Test used as a part of the measure of criterion referenced reading achievement is designed for use when pupils have completed the lessons in the Harper & Row first reader. The stated purpose of the test is to measure how well pupils have acquired the understandings and skills taught in the first reader. The test consists of five subtests, each of which yields a raw score, and the sum of these raw scores yields a total raw score. The five subtests include: word recognition; auditory-visual perception (initial consonants, clusters, and digraphs); auditory-visual perception (final consonants, clusters, and digraphs); auditory-visual perception (vowels); and comprehension. The total test contains 164 items. Reliability for the test was .98 using the Kuder-Richardson Formula 20. The standard error of measurement for the test was 3.80 (Harper & Row, 1968).

The Harper & Row Second Reader Achievement Test used as a part of the measure of criterion referenced reading achievement is designed for use when pupils have completed the lessons in the Harper & Row second reader. The stated purpose of the test is to measure how well the pupils have acquired the understandings and skills taught in the second reader. The test consists of seven subtests, each of which yields a raw score, and the sum of these raw scores yields a total raw score. The seven subtests include: word recognition; auditory-visual perception (initial consonants, clusters, and digraphs); auditory-visual
perception (final consonants, clusters, and digraphs); auditory-visual perception (short vowel sounds); auditory-visual perception (long vowel sounds); word structure (syllables); and comprehension. The total test contains 213 items. Reliability for the test was .98 using the Kuder-Richardson Formula 20. The standard error of measurement for the test was 5.43 (Harper & Row, 1968).

The Cooperative Primary Reading Test, Form 23A, used as the measure of normative referenced reading achievement is designed for use at grades two and three. The test is a general reading achievement test with a vocabulary level geared to that of standard primary reading programs. The test is not tied to any particular instructional materials or published vocabulary lists. One of the stated purposes of the test is to provide teachers with measures of children's concepts and skills that relate closely to their work in the classroom. The test is not divided into subtests and yields only a total raw score. The total test contains 50 items. Reliability for second grade administration of the test was .89 using the Kuder-Richardson Formula 20. The standard error of measurement was 3.09 (Educational Testing Service, 1967).

The Coloured Progressive Matrices, Sets A, Ab, B, administered to stratify the population in terms of perceptual reasoning ability are designed for use with young children and old people, for anthropological studies, and for clinical work. This is not a test of general
intelligence, but a perceptual test to assess a person's capacity for intellectual activity, irrespective of his acquired knowledge. The three sets of 12 problems each are arranged to assess the chief cognitive processes usually within the capability of children under 11 years of age. Each of the sets yields a raw score, and the sum of these raw scores yields a total raw score. The test-retest reliability for the test with "normal schoolchildren" was .89 with a three month interval, and .86 with a six month interval (Raven, 1965).

Nelson and Edelstein (1963) reported a correlation of .69 between the Progressive Matrices and the California Test of Mental Maturity, and suggested the use of the Progressive Matrices as an additional method of assessing the intelligence of children with language and/or cultural handicaps. Martin and Wiechers (1954) reported correlations between the Coloured Progressive Matrices and the Wechsler Intelligence Scale for Children of .91, .84, and .83 for Full Scale, Verbal, and Performance respectively. Slightly lower correlations for the Progressive Matrices and the Wechsler Intelligence Scale for Children were reported by Barratt (1956). He reported correlations of .75, .69, and .70 between the Progressive Matrices and the Wechsler Total, Verbal, and Performance respectively. Barratt also indicated that the Progressive Matrices correlates highest with those tests of the Wechsler Intelligence Scale for Children involving spatial reasoning,
verbal reasoning of a more or less abstract nature, and acquired knowledge.

The Experimental Treatment

Programed tutoring was a technique designed to supplement rather than supplant the classroom teaching of reading, thus it was operated on a "pull-out" basis. The pupils in the experimental group were taken out of their classrooms for 20 minutes each day for the tutoring sessions. Fifteen minutes of each session were devoted to tutoring, and the remaining five minutes were required going to and coming from the tutoring session. The time of tutoring was rotated from day-to-day to avoid interruption of the same classroom activity each day. However, the pupils were never taken out of the classroom reading period. The treatment period extended from October 5, 1970, through April 30, 1971. Variations in individual school programs produced minor variations in the actual number of tutoring sessions available, and across the six schools, the number of available tutoring sessions ranged from 127 to 136.

No formal randomization procedure was employed to assign the pupils to the tutors; however, the assignments were made on a chance basis. Each pupil remained with his assigned tutor throughout the treatment period.

The tutors were salaried, non-certificated personnel employed by the Sacramento City Unified School
District. The high school diploma and poverty status were conditions of employment. Applicants indigenous to the school communities were given preference by District policy. The tutors selected for this experimental program were experienced, having served as programed tutors in the first grade programed tutoring program during the previous school year. A total of 22 tutors were involved in the experimental program.

The tutors were trained in the use of the programed tutoring technique by the Sacramento City Unified School District with the assistance of personnel from Indiana University. A full-time tutoring supervisor was employed by the District to guide and assist all salaried tutors in the District. The tutors in the experimental program had access to this service.

The programed tutoring technique is fully described in the experimental edition of the Harper & Row Programed Tutorial II Kit (Ellson, Barber, Harris, & Adams, 1970). The kit contains the Tutor's Guide, the Tutor's Question Book, and the Word Analysis Book. The kit was used as a supplement to the Harper & Row, Strand I second reader, All Through the Year, and was designed to teach sight reading, comprehension, and word analysis skills. These skills were divided into subskills which were taught through nine programs. The reader is referred to the kit for a detailed description of the programs and the programed tutoring technique.
Programed tutoring requires active learning on the part of the child. He is actively engaged in reading and reacting to what he reads. Each child progresses at his own rate based upon his successes and failures, thus the technique provides each child with a special pattern of teaching.

Each program began by presenting the child with a reading problem or task. If the child could not solve the problem, it was progressively simplified through more information, hints, or additional context until the child "discovered" a solution. If the child failed to discover a solution or "blocked" after ten trials, he was taken to the next lesson which included materials from the lesson on which he had blocked. This procedure provided for re-teaching of the unlearned material, and allowed the child to progress at the same time. The tutor never provided the child a complete solution to a problem.

The child's successes were emphasized by praise and encouragement, while his failures were ignored in the sense that the tutor did not call attention to errors with any obvious action. If a child made an error, the tutor simply took him to the next procedure in the program which was designed to elicit an appropriate response. The tutor's words and actions were pre-determined, depending upon the child's responses.

Programed tutoring was systematic teaching with each lesson building upon previous lessons with mastery as
the goal. Teaching time was concentrated where it was needed, and time was not spent in teaching what the child already knew.

**The Experimental Design**

This investigation represented a partially randomized, posttest control group design. There was randomization within, but not across the six schools. There was a pretest, but its function was for selection of the sample rather than as a base line for gain scores. Posttest differences were used to test the hypotheses relating to reading achievement.

This design was chosen to maximize both internal and external validity (Campbell & Stanley, 1963) within the restrictions imposed on the sampling procedures. The randomization provided the necessary internal controls for history, maturation, selection, testing, instrumentation, regression, and mortality. Pretest results were used to investigate any possible effects of experimental mortality. The basic fault of this design in terms of external validity, pretest effects, was not a matter of concern, as the pretest was required for selection of pupils for the programed tutoring treatment. Thus, pretesting would be a part of implementation of this treatment, and any possible reactive pretest effects would not limit generalization of the findings of the study.

The schools chosen to participate in this
experiment were not chosen from among volunteers. The treatment was unobtrusively included as a part of the comprehensive compensatory education program provided in these schools. This procedure served to control any selection bias that might threaten the external validity of the study. The treatment, by design, was intended for lower achieving pupils, and the selection of the sample from lower achieving pupils did not present a problem in generalizing the results of the study.

The "pull-out" nature of the treatment was a common aspect of special programs and services in the elementary schools, and as such would not create any reactive effects. External validity would not be jeopardized by this procedure.

Despite the restrictions placed on the sampling procedures, the design for this investigation provided adequate controls for internal and external validity. The findings of the study should be generalizable within the limitations stated in Chapter 1.

The Statistical Procedures

Each of the hypotheses stated in Chapter 1 was restated in the null form and tested by appropriate statistical tests. Two-tailed tests were applied in all cases, and the level of significance for rejecting the null hypotheses was set at .05. This level of conservatism was judged appropriate by the investigator for this initial
test of the use of programed tutoring in reading at the second grade level. Ultimately, educational significance, as well as statistical significance must guide decisions regarding educational practices.

The nature of the data collected in this investigation required the use of four different statistical tests to test the null hypotheses. The following hypotheses were tested by means of a 2x3 factorial analysis of variance.

$H_1$. Lower achieving second grade children from low-income areas will demonstrate significantly greater criterion referenced reading achievement after one year of programed tutoring in the second grade than similar second grade children who have not been tutored in the second grade.

$H_2$. There will be a difference in criterion referenced reading achievement among lower achieving second grade children of varying perceptual reasoning abilities from low-income areas with high perceptual reasoners demonstrating significantly greater achievement than average and low perceptual reasoners, and average perceptual reasoners demonstrating greater achievement than low perceptual reasoners.

$H_3$. Lower achieving second grade children from low-income areas will demonstrate significantly greater normative referenced reading achievement after one year of programed tutoring in the second grade than similar second grade children who have not been tutored in the second grade.
second grade.

\[ H_5 \] There will be a difference in normative referenced reading achievement among lower achieving second grade children of varying perceptual reasoning abilities from low-income areas with high perceptual reasoners demonstrating significantly greater achievement than average and low perceptual reasoners, and average perceptual reasoners demonstrating greater achievement than low perceptual reasoners.

Separate analyses were carried out for criterion referenced reading achievement (\( H_1 \) and \( H_2 \)) and normative referenced reading achievement (\( H_4 \) and \( H_5 \)). The paradigm for these analyses follows:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptual Reasoning Ability</td>
<td>High</td>
<td>Dependent</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>Variable</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td></td>
</tr>
</tbody>
</table>

Analysis of covariance was considered for testing the hypotheses relating to reading achievement, using the pretest (selection test) as the covariate. However, the lack of strong linear relationships between the proposed covariate and the dependent variables ruled out this method of analysis. The correlations between the proposed covariate and criterion and normative referenced reading achievement were .26 and .15 respectively.
The two remaining hypotheses relative to reading achievement are stated below:

H2. Lower achieving second grade children from low-income areas who possess different perceptual reasoning abilities will achieve differentially under the programed tutoring treatment with high perceptual reasoners who are not tutored ranking higher in criterion referenced reading achievement than high perceptual reasoners who are tutored, and low perceptual reasoners who are not tutored ranking lower in criterion referenced reading achievement than low perceptual reasoners who are tutored.

H6. Lower achieving second grade children from low-income areas who possess different perceptual reasoning abilities will achieve differentially under the programed tutoring treatment with high perceptual reasoners who are not tutored ranking higher in normative referenced reading achievement than high perceptual reasoners who are tutored, and low perceptual reasoners who are not tutored ranking lower in normative referenced reading achievement than low perceptual reasoners who are tutored.

These hypotheses were tested by means of separate applications of the Mann-Whitney U-test. This procedure was employed to test differential effects of the programed tutoring treatment for the high and low perceptual reasoning groups. The Mann-Whitney U-test was selected to avoid difficulty with the assumptions required for parametric techniques.
The seventh hypothesis was tested by means of the biserial correlation technique.

\( H_7 \). There will be a significant positive biserial correlation between pupil attendance rates and treatment for lower achieving second grade children from low-income areas.

The biserial method of correlation was required to test the relationship between attendance and treatment as the first variable, attendance, was continuous; and the second variable, treatment, was forced into a dichotomy. For the treatment variable, tutoring was assigned the value one, and no tutoring was assigned the value zero.

The eighth hypothesis was tested by means of the tetrachoric correlation technique.

\( H_8 \). There will be a positive significant tetrachoric correlation between pupil mobility and treatment for lower achieving second grade children from low-income areas.

The tetrachoric method of correlation was required to test the relationship between mobility and treatment as both the variables were continuous and forced into dichotomies. For the mobility variable, mobility was assigned a value of zero, and no mobility was assigned a value of one. In the case of the treatment variable, tutoring was assigned a value of one, and no tutoring was assigned a value of zero.

In addition to testing the hypotheses stated in
Chapter 1, data were collected for the 27 pupils who were identified as having been tutored in the first grade. These pupils were excluded from the above analyses to avoid contamination of the second grade tutoring program with any previous tutoring effects. The data collected for these 27 pupils included test scores on all of the instruments used in the investigation, attendance data, and mobility data. These data were treated descriptively and summarized for inclusion in Chapter 4.

The results of the statistical analyses described above are presented in the following chapter. Brief interpretations follow each of the sets of data presented.
Chapter 4

FINDINGS OF THE STUDY

This chapter presents the findings of the investigation. It is organized into five sections: (a) criterion referenced reading achievement, (b) normative referenced reading achievement, (c) pupil attendance, (d) pupil mobility, and (e) pupils who were tutored in the first grade. In sections a – d, the relevant research hypotheses are stated in the null form, and the results of the statistical tests employed to test these hypotheses are reported. Section e presents descriptive data relative to the 27 pupils in the original sample who were tutored in the first grade during the previous year, and were excluded from the statistical analyses.

Criterion Referenced Reading Achievement

Of the 108 pupils selected originally for the programmed tutoring treatment in the second grade, data regarding criterion referenced reading achievement were available for 72 pupils. Of the 55 pupils selected originally for the control group, criterion referenced reading achievement data were available for 36 pupils. These losses of subjects from the experimental and control groups are accounted for in Table 1, which also includes the
effect of this mortality on the similarity of the experimental and control groups.

Table 1
Summary of Program Mortality and its Effect on the Similarity of the Experimental and Control Groups
Criterion Referenced Reading Achievement

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of Pupils</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
</tr>
<tr>
<td>Original Sample</td>
<td>108</td>
</tr>
<tr>
<td>Excluded--First Grade Tutoring</td>
<td>18</td>
</tr>
<tr>
<td>Transferred from the Schools</td>
<td>14</td>
</tr>
<tr>
<td>Incomplete Test Data</td>
<td>4</td>
</tr>
<tr>
<td>Final Sample</td>
<td>72</td>
</tr>
</tbody>
</table>

| Statistic                                      | Harper & Row Second Year Readiness Test |
|                                               | Experimental | Control |
| Original Sample                               |             |         |
| Number of Pupils                              | 108         | 55      |
| Mean Raw Score                                | 77.68       | 78.58   |
| Variance                                      | 334.22      | 311.81  |
| Final Sample                                  |             |         |
| Number of Pupils                              | 72          | 36      |
| Mean Raw Score                                | 77.56       | 78.61   |
| Variance                                      | 384.49      | 321.96  |
The data presented in Table 1 indicate that between the time of selecting the sample and the collection of the posttest data, the original sample was reduced by approximately one-third. This reduction was consistent across the experimental and control groups. The data in Table 1 also indicate that despite the loss of approximately one-third of the original sample, the similarity of the experimental and control groups was generally not affected in terms of the mean raw scores and variances on the selection test (pretest).

Table 2 presents a summary of the selection and classification test data for each of the cells in the analysis of variance model employed in the analysis of the criterion referenced reading achievement. These data demonstrate the similarity of the experimental and control groups for each classification of perceptual reasoning ability. The distributions of scores for the experimental and control groups on The Coloured Progressive Matrices Tests show a wide range of perceptual reasoning abilities among the pupils considered in the study. The control groups contained both the high and low extremes in terms of perceptual reasoning ability, but these individuals were not at the extremes on any other measure. Thus, while these extreme scores contributed markedly to the variance in perceptual reasoning, similar effects were not carried over to other measures, and the individuals were retained in the sample.
<table>
<thead>
<tr>
<th>Classification</th>
<th>Number of Pupils</th>
<th>Statistic</th>
<th>Cell Sizes, Means, and Variances</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>34</td>
<td>33</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Exp.</td>
<td>Con.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of Pupils</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Raw Score</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>22</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Exp.</td>
<td>Con.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of Pupils</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Mean Raw Score</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Low</td>
<td>16</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Exp.</td>
<td>Con.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of Pupils</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean Raw Score</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Three hypotheses were stated in Chapter 1 regarding criterion referenced reading achievement. These hypotheses were:

$H_1$. There will be no significant difference in the criterion referenced reading achievement of lower achieving second grade children from low-income areas after one year of programmed tutoring in the second grade and similar second grade children who were not tutored in the second grade.

$H_2$. There will be no significant difference in the criterion referenced reading achievement of high, average, and low perceptual reasoners among lower achieving second grade children from low-income areas.

$H_3$. There will be no significant difference in the criterion referenced reading achievement ranks of tutored and non-tutored high perceptual reasoners, and no significant difference in the criterion referenced reading achievement ranks of tutored and non-tutored low perceptual reasoners.

Table 3 presents summary data relative to the analysis model used to test $H_1$ and $H_2$. Application of the $F_{\text{max}}$ test (Winer, 1971) to these data indicates that the assumption of homogeneity of cell variance has not been violated. The data in Table 3 further indicate the need for an analysis of variance procedure that is appropriate for unequal, non-proportional cell sizes.

A least squares analysis of variance procedure was
selected from Winer (1971) and applied to the criterion referenced reading achievement data. The results of the analysis of variance for criterion referenced reading achievement are summarized in Table 4. The data reported in Table 4 support rejection of the first hypothesis, but fail to reject the second hypothesis.

Table 3
Summary of Cell Sizes, Means, and Variances for Analysis of Criterion Referenced Reading Achievement

<table>
<thead>
<tr>
<th>Perceptual Reasoning Ability</th>
<th>Number of Pupils</th>
<th>Mean Raw Score</th>
<th>Variance</th>
<th>Harper &amp; Row First Reader &amp; Second Reader Achievement Tests—Composite Raw Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Experimental</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Pupils</td>
<td>18</td>
<td>300.33</td>
<td>892.21</td>
<td></td>
</tr>
<tr>
<td>Mean Raw Score</td>
<td></td>
<td></td>
<td></td>
<td>271.86</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td></td>
<td>2626.59</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Pupils</td>
<td>43</td>
<td>303.58</td>
<td>1540.33</td>
<td></td>
</tr>
<tr>
<td>Mean Raw Score</td>
<td></td>
<td></td>
<td></td>
<td>287.35</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td></td>
<td>2838.98</td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Pupils</td>
<td>11</td>
<td>299.73</td>
<td>921.95</td>
<td></td>
</tr>
<tr>
<td>Mean Raw Score</td>
<td></td>
<td></td>
<td></td>
<td>278.08</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td></td>
<td>1333.93</td>
</tr>
</tbody>
</table>
Table 4

Summary of Analysis of Variance for Criterion Referenced Reading Achievement—Harper & Row First Reader & Second Reader Achievement Tests, Composite Raw Score

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>9435.41</td>
<td>1</td>
<td>9435.41</td>
<td>5.73*</td>
</tr>
<tr>
<td>Perceptual Reasoning (B)</td>
<td>1034.45</td>
<td>2</td>
<td>517.23</td>
<td>.31</td>
</tr>
<tr>
<td>AXB</td>
<td>548.67</td>
<td>2</td>
<td>274.34</td>
<td>.17</td>
</tr>
<tr>
<td>Error</td>
<td>168,004.31</td>
<td>102</td>
<td>1647.10</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

The third hypothesis relative to criterion referenced reading achievement (H₃) was tested by means of the Mann-Whitney U-test (Haber & Runyon, 1969). Table 5 presents the experimental-control composite distributions of criterion referenced reading achievement scores and ranks for the high perceptual reasoners and the same for the low perceptual reasoners, along with the results of the Mann-Whitney U-test. Application of the Mann-Whitney U-test to the distributions of ranks yielded U values too high to reject the third hypothesis.

Of the three research hypotheses stated in Chapter 1 relative to criterion referenced reading achievement, only the first was confirmed. Lower achieving second grade children from low-income areas did demonstrate significantly greater criterion referenced reading achievement after one year of programed tutoring in the second grade than similar pupils who were not tutored.
Table 5

Distributions of Criterion Referenced Reading Achievement Data for High Perceptual Reasoners and Low Perceptual Reasoners Showing Ranks and Treatment Condition* and Mann-Whitney U-test Results

<table>
<thead>
<tr>
<th>High Perceptual Reasoners</th>
<th>Low Perceptual Reasoners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>Rank</td>
</tr>
<tr>
<td>194</td>
<td>1</td>
</tr>
<tr>
<td>238</td>
<td>2</td>
</tr>
<tr>
<td>243</td>
<td>3</td>
</tr>
<tr>
<td>245</td>
<td>4</td>
</tr>
<tr>
<td>258</td>
<td>5</td>
</tr>
<tr>
<td>259</td>
<td>6</td>
</tr>
<tr>
<td>264</td>
<td>7</td>
</tr>
<tr>
<td>275</td>
<td>8</td>
</tr>
<tr>
<td>277</td>
<td>9</td>
</tr>
<tr>
<td>278</td>
<td>10</td>
</tr>
<tr>
<td>280</td>
<td>11</td>
</tr>
<tr>
<td>283</td>
<td>12</td>
</tr>
<tr>
<td>287</td>
<td>13</td>
</tr>
<tr>
<td>295</td>
<td>14</td>
</tr>
<tr>
<td>315</td>
<td>15</td>
</tr>
<tr>
<td>318</td>
<td>16</td>
</tr>
<tr>
<td>321</td>
<td>17</td>
</tr>
<tr>
<td>324</td>
<td>18</td>
</tr>
<tr>
<td>326</td>
<td>19</td>
</tr>
<tr>
<td>329</td>
<td>20</td>
</tr>
<tr>
<td>331</td>
<td>21</td>
</tr>
<tr>
<td>335</td>
<td>22</td>
</tr>
<tr>
<td>338</td>
<td>23</td>
</tr>
<tr>
<td>344</td>
<td>24</td>
</tr>
<tr>
<td>352</td>
<td>25</td>
</tr>
</tbody>
</table>

U = 42  
U = 51.5  
n_E = 18  
n_E = 11  
n_C = 7  
n_C = 12

*E = Experimental (Tutored)  
C = Control (Not Tutored)
Of the 108 pupils selected originally for the programmed tutoring treatment in the second grade, data regarding normative referenced reading achievement were available for 63 pupils. Of the 55 pupils selected originally for the control group, normative referenced reading achievement data were available for 30 pupils. These losses of subjects from the experimental and control groups are accounted for in Table 6, which also includes the effect of this mortality on the similarity of the experimental and control groups.

The data presented in Table 6 indicate that between the time of selecting the sample and the collection of the normative referenced posttest data, the original sample was reduced by approximately 40%. This reduction was consistent across the experimental and control groups. The data reported in Table 6 also indicate that despite the loss of approximately 40% of the original sample, the similarity of the experimental and control groups was generally not affected in terms of the mean raw scores and variances on the selection test (pretest).

Table 7 presents a summary of the selection and classification test data for each of the cells in the analysis of variance model employed in the analysis of normative referenced reading achievement. These data demonstrate the similarity of the experimental and control groups for each classification of perceptual reasoning ability. The
Table 6
Summary of Program Mortality and its Effect on the Similarity of the Experimental and Control Groups Normative Referenced Reading Achievement

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of Pupils</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
</tr>
<tr>
<td>Original Sample</td>
<td>108</td>
</tr>
<tr>
<td>Excluded--First Grade Tutoring</td>
<td>18</td>
</tr>
<tr>
<td>Transferred from the Schools</td>
<td>14</td>
</tr>
<tr>
<td>Incomplete Test Data</td>
<td>13</td>
</tr>
<tr>
<td>Final Sample</td>
<td>63</td>
</tr>
</tbody>
</table>

| Statistic                                  | Experimental     | Control |
|--------------------------------------------|------------------|
| Original Sample                            |                  |
| Number of Pupils                           | 108              | 55      |
| Mean Raw Score                             | 77.68            | 78.58   |
| Variance                                   | 334.22           | 311.81  |
| Final Sample                               |                  |
| Number of Pupils                           | 63               | 30      |
| Mean Raw Score                             | 77.60            | 78.07   |
| Variance                                   | 386.21           | 298.89  |
Table 7
Summary of Selection and Classification Test Data by Cell for Pupils Included in the Analysis of Normative Referenced Reading Achievement

<table>
<thead>
<tr>
<th>Distribution of Scores</th>
<th>Cell Sizes, Means, and Variances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The Coloured Progressive Matrices Test</td>
</tr>
<tr>
<td>High</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>33</td>
</tr>
<tr>
<td>32</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>28</td>
</tr>
<tr>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>
distributions of scores for the experimental and control groups on The Coloured Progressive Matrices Tests show a wide range of perceptual reasoning abilities among the pupils considered in the study. The control groups contained both the high and low extremes in terms of perceptual reasoning ability, but these individuals were not at the extremes on any other measure. Thus while these extreme scores contributed markedly to the variance in perceptual reasoning, similar effects were not carried over to other measures, and the individuals were retained in the sample.

Three hypotheses were stated in Chapter 1 regarding normative referenced reading achievement. These hypotheses were:

\( H_4 \). There will be no significant difference in the normative referenced reading achievement of lower achieving second grade children from low-income areas after one year of programed tutoring in the second grade and similar second grade children who were not tutored in the second grade.

\( H_5 \). There will be no significant difference in the normative referenced reading achievement of high, average, and low perceptual reasoners among lower achieving second grade children from low-income areas.

\( H_6 \). There will be no significant difference in the normative referenced reading achievement ranks of tutored and non-tutored high perceptual reasoners, and no
significant difference in the normative referenced reading achievement ranks of tutored and non-tutored low perceptual reasoners.

Table 8 presents summary data relative to the analysis of variance model used to test \( H_4 \) and \( H_5 \). Application of the \( F_{\text{max}} \) test (Winer, 1971) to these data indicates that the assumption of homogeneity of cell variance has not been violated. The data in Table 8 further indicate the need for an analysis of variance procedure that is appropriate for unequal, non-proportional cell sizes.

A least squares analysis of variance procedure was selected from Winer (1971) and applied to the normative referenced reading achievement data. The results of the analysis of variance for normative referenced reading achievement are summarized in Table 9. The data presented in Table 9 fail to reject either the fourth or fifth hypotheses.

The third hypothesis relative to normative referenced reading achievement (\( H_6 \)) was tested by means of the Mann-Whitney U-test (Haber & Runyon, 1969). Table 10 presents the experimental-control composite distributions of normative referenced reading achievement scores and ranks for the high perceptual reasoners and the same for the low perceptual reasoners, along with the results of the Mann-Whitney U-test. Application of the Mann-Whitney U-test to the distributions of ranks yielded U values too high to reject the sixth hypothesis.
Table 8
Summary of Cell Sizes, Means, and Variances for Analysis of Normative Referenced Reading Achievement

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Pupils</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>Mean Raw Score</td>
<td>22.19</td>
<td>19.00</td>
</tr>
<tr>
<td>Variance</td>
<td>36.66</td>
<td>18.80</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Pupils</td>
<td>37</td>
<td>14</td>
</tr>
<tr>
<td>Mean Raw Score</td>
<td>23.16</td>
<td>23.29</td>
</tr>
<tr>
<td>Variance</td>
<td>24.34</td>
<td>36.87</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Pupils</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Mean Raw Score</td>
<td>22.80</td>
<td>22.55</td>
</tr>
<tr>
<td>Variance</td>
<td>40.56</td>
<td>65.27</td>
</tr>
</tbody>
</table>
Table 9
Summary of Analysis of Variance for Normative Referenced Reading Achievement--Cooperative Primary Reading Test, Form 23A

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>7.60</td>
<td>1</td>
<td>7.60</td>
<td>.21</td>
</tr>
<tr>
<td>Perceptual Reasoning (B)</td>
<td>47.76</td>
<td>2</td>
<td>23.88</td>
<td>.65</td>
</tr>
<tr>
<td>AXB</td>
<td>31.59</td>
<td>2</td>
<td>15.80</td>
<td>.43</td>
</tr>
<tr>
<td>Error</td>
<td>3218.66</td>
<td>87</td>
<td>37.00</td>
<td></td>
</tr>
</tbody>
</table>
Table 10
Distributions of Normative Referenced Reading Achievement Data for High Perceptual Reasoners and Low Perceptual Reasoners Showing Ranks and Treatment Condition* and Mann-Whitney U-test Results

<table>
<thead>
<tr>
<th>Score</th>
<th>Rank</th>
<th>Condition</th>
<th>Score</th>
<th>Rank</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1</td>
<td>E</td>
<td>14</td>
<td>1.5</td>
<td>E</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>C</td>
<td>14</td>
<td>1.5</td>
<td>E</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>E</td>
<td>16</td>
<td>3.5</td>
<td>E</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>E</td>
<td>16</td>
<td>3.5</td>
<td>E</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>C</td>
<td>17</td>
<td>5</td>
<td>C</td>
</tr>
<tr>
<td>18</td>
<td>6</td>
<td>E</td>
<td>18</td>
<td>6</td>
<td>C</td>
</tr>
<tr>
<td>18</td>
<td>6</td>
<td>E</td>
<td>19</td>
<td>7.5</td>
<td>C</td>
</tr>
<tr>
<td>19</td>
<td>7.5</td>
<td>E</td>
<td>19</td>
<td>7.5</td>
<td>C</td>
</tr>
<tr>
<td>19</td>
<td>8.5</td>
<td>C</td>
<td>20</td>
<td>9</td>
<td>E</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>E</td>
<td>21</td>
<td>10</td>
<td>E</td>
</tr>
<tr>
<td>21</td>
<td>11</td>
<td>E</td>
<td>22</td>
<td>11.5</td>
<td>E</td>
</tr>
<tr>
<td>22</td>
<td>13</td>
<td>E</td>
<td>22</td>
<td>11.5</td>
<td>E</td>
</tr>
<tr>
<td>22</td>
<td>13</td>
<td>E</td>
<td>23</td>
<td>14</td>
<td>C</td>
</tr>
<tr>
<td>23</td>
<td>14</td>
<td>E</td>
<td>23</td>
<td>14</td>
<td>C</td>
</tr>
<tr>
<td>24</td>
<td>15</td>
<td>E</td>
<td>23</td>
<td>14</td>
<td>C</td>
</tr>
<tr>
<td>24</td>
<td>16.5</td>
<td>E</td>
<td>24</td>
<td>16.5</td>
<td>E</td>
</tr>
<tr>
<td>24</td>
<td>16.5</td>
<td>E</td>
<td>25</td>
<td>18</td>
<td>E</td>
</tr>
<tr>
<td>27</td>
<td>18</td>
<td>C</td>
<td>36</td>
<td>19</td>
<td>C</td>
</tr>
<tr>
<td>30</td>
<td>19</td>
<td>E</td>
<td>39</td>
<td>20</td>
<td>E</td>
</tr>
<tr>
<td>31</td>
<td>20</td>
<td>E</td>
<td>41</td>
<td>21</td>
<td>C</td>
</tr>
<tr>
<td>36</td>
<td>21</td>
<td>E</td>
<td>41</td>
<td>21</td>
<td>C</td>
</tr>
</tbody>
</table>

\[ U = 23.5 \quad U = 44.5 \]

\[ n_E = 16 \quad n_E = 10 \]
\[ n_C = 5 \quad n_C = 11 \]

*E = Experimental (Tutored)
C = Control (Not Tutored)
In reviewing the results of the statistical analyses regarding normative referenced reading achievement, no significant differences were identified. Thus, of the fourth, fifth, and sixth hypotheses stated in Chapter 1 relative to normative referenced reading achievement, none was confirmed.

Pupil Attendance

Of the 108 pupils selected originally for the programmed tutoring treatment in the second grade, data regarding pupil attendance were available for 73 pupils. Of the 55 pupils selected originally for the control group, attendance data were available for 37 pupils. These data were in the form of rates of attendance, that is percentages of days attended.

One hypothesis was stated in Chapter 1 regarding pupil attendance. This hypothesis was:

\( H_7 \): There will be no significant biserial correlation between pupil attendance rates and treatment for lower achieving second grade pupils from low-income areas.

The biserial correlation was chosen because the attendance variable was continuous, and the treatment variable was forced into a dichotomy (tutoring = 1 and no tutoring = 0). Application of the biserial correlation procedure to the attendance and treatment data resulted in a biserial correlation of .11. This correlation was not sufficiently high enough to reject \( H_7 \), thus the
research hypothesis stated in Chapter 1 regarding the relationship between pupil attendance and treatment was not confirmed (Dick & Hagerty, 1971).

Pupil Mobility

Of the 108 pupils selected originally for the programmed tutoring treatment in the second grade, and the 55 pupils selected originally for the control group, 18 experimental and 9 control pupils were excluded from the statistical analyses due to their having been tutored in the first grade. Pupil mobility data were analyzed for the remaining 90 experimental and 46 control pupils.

One hypothesis was stated in Chapter 1 regarding pupil mobility. This hypothesis was:

$H_8$: There will be no significant tetrachoric correlation between pupil mobility and treatment for lower achieving second grade pupils from low-income areas.

The tetrachoric correlation was chosen because both the pupil mobility and treatment variables were forced into dichotomies. For pupil mobility, mobility was assigned the value zero, and no mobility was assigned the value one; for treatment, tutoring was assigned the value one, and no tutoring was assigned the value zero.

Application of the tetrachoric correlation procedure to the pupil mobility and treatment data resulted in a tetrachoric correlation of .11. This correlation was not sufficiently high to reject $H_8$, thus the hypothesis
stated in Chapter 1 regarding pupil mobility and treatment was not confirmed (Dick & Hagerty, 1971).

Pupils Who Were Tutored in the First Grade

After the original experimental and control samples had been selected, it was determined that 18 of the experimental and nine of the control pupils had been tutored in the first grade. These pupils were maintained in the programed tutoring program and participated in all of the testing, but were excluded from the analyses relative to the hypotheses stated in Chapter 1. These pupils were not maintained as separate groups representing "two years of tutoring" and "tutoring in the first grade only" as they were not representative of all pupils tutored in the first grade. This was established by the fact that they qualified as low achievers in the second grade, while other first grade tutored pupils did not qualify as low achievers in the second grade.

Descriptive data are reported for these two groups of pupils as a matter of information with the hope that there may be implications for further study. For the purpose of this description, the pupils tutored in both the first and second grades are designated Group A; and the pupils tutored in the first grade, but not in the second grade are designated Group B.

In October of 1970, Group A consisted of 18 pupils and Group B consisted of nine pupils. In May of 1971,
Group A contained 13 pupils for a 27.8% mobility rate, and Group B contained five pupils for a 44.4% mobility rate. Group A had an average attendance rate of 93.1% compared to an average attendance rate of 87.2% for Group B.

Table 11 presents pretest, criterion referenced, and normative referenced reading achievement test data for the pupils in Groups A and B.

### Table 11

Summary Test Data for Pupils Tutored in Both the First and Second Grades (Group A) and in the First Grade Only (Group B)

<table>
<thead>
<tr>
<th>Test/Statistic</th>
<th>Group A N = 13</th>
<th>Group B N = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harper &amp; Row Second Year Readiness Test - September, 1970</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Raw Score</td>
<td>82.5</td>
<td>80.8</td>
</tr>
<tr>
<td>Variance</td>
<td>382.5</td>
<td>80.6</td>
</tr>
<tr>
<td>Harper &amp; Row First Reader &amp; Second Reader Achievement Tests (Composite) May, 1971</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Raw Score</td>
<td>295.8</td>
<td>286.0</td>
</tr>
<tr>
<td>Variance</td>
<td>1866.9</td>
<td>284.8</td>
</tr>
<tr>
<td>Cooperative Primary Reading Test May, 1971</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Raw Score</td>
<td>22.1</td>
<td>18.5</td>
</tr>
<tr>
<td>Variance</td>
<td>28.4</td>
<td>10.8</td>
</tr>
</tbody>
</table>
The data collected for pupils tutored in the first grade and tutored or not tutored in the second grade showed some apparent differences from the data collected for pupils not tutored in the first grade and tutored or not tutored in the second grade. There was a greater disparity in attendance rates between the experimental and control groups during the second grade for pupils who had been tutored in the first grade. Pupils who were tutored in the first grade, but not the second demonstrated a higher mobility rate than pupils tutored both years.

Pupils who were tutored both years showed relatively greater normative referenced reading achievement than criterion referenced reading achievement when compared to their control group. This pattern was reversed for pupils tutored in the second grade only, as they showed relatively greater criterion referenced reading achievement than normative referenced reading achievement when compared to their control group.

Summary of the Findings

Of the eight research hypotheses formulated for this investigation in Chapter 1, only the first was confirmed. Lower achieving second grade children from low-income areas did demonstrate greater criterion referenced reading achievement after one year of programed tutoring in the second grade than similar second grade pupils who were not tutored in the second grade.
There were some apparent differences in effects of the programed tutoring treatment on pupils who had been tutored in the first grade and pupils who had not been tutored in the first grade. These findings are based upon observation of the data collected, and have not been confirmed by statistical analyses.

The final chapter of this study, Chapter 5, presents the investigator's interpretation of the findings reported in this chapter. In addition, the investigator offers recommendations for further study based upon the findings of this investigation.
Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

This chapter is organized into four major sections: in the first three sections, conclusions and interpretations are presented relative to the data reported in Chapter 4 regarding (a) reading achievement, (b) pupil attendance, and (c) pupil mobility; on the basis of these conclusions and interpretations, the fourth section presents recommendations for further study. The investigator recognized the danger in drawing final conclusions from one experiment, and urges the reader to observe the same caution in reading the contents of this chapter.

Reading Achievement

The first hypothesis was substantiated, indicating that programmed tutoring does have a positive impact on the criterion referenced reading achievement of lower achieving second grade children from low-income areas. This finding is consistent with those regarding first grade children reported by Ellison (1969b, 1969c), Ellison and associates (1965, 1968, 1969a, 1969b, & 1970), the Sacramento City Unified School District (1969, 1970), and McCleary (1971). However, if one considers the matter of educational significance, this finding is less encouraging.
Although the tutored pupils did score significantly higher than the non-tutored pupils on the instruments employed to assess criterion referenced reading achievement, the absolute difference between the mean raw scores for the tutored and non-tutored pupils was relatively small. In terms of the mean raw scores, the tutored pupils scored less than 8% higher than the non-tutored pupils.

The criterion referenced instruments used for this part of the study were developed by the publisher of the reading series used in the classroom reading program for the purpose of assessing pupil progress in that series (Harper & Row, 1968). In addition, the programed tutoring materials were based upon the same reading series. Thus, even with this close relationship between the classroom reading materials, the supplemental tutoring materials, and the criterion measure, the absolute effects of the tutoring supplement were minimal. Therefore, while it may be concluded that programed tutoring does enhance the criterion referenced reading achievement of lower achieving second grade children from low-income areas, it is doubtful that the benefits justify the expenditure of funds required for salaried paraprofessional tutors.

Closely related to this matter are the findings regarding the fourth hypothesis. This hypothesis was not confirmed, indicating that programed tutoring has no effect on normative referenced reading achievement for lower achieving second grade children from low-income areas.
This finding is in agreement with those regarding first grade children reported by Ellson et al. (1968) and the Sacramento City Unified School District (1970), but in disagreement with those reported by Ellson (1969) and the Sacramento City Unified School District (1969).

While it is recognized that normative referenced reading achievement instruments measure highly generalized skills rather than the specific instructional objectives of a given program (Skager, 1971), it would seem reasonable to expect that an effective reading program would enhance the pupils' facility with those generalized skills. This expectation, of course, is based upon the assumption that there is congruence between the instructional objectives and the generalized skills. The fact that the program tutoring treatment was adapted to the state adopted Harper & Row reading series and the Cooperative Primary Reading Test used in this part of the study suggests that the above assumption was valid.

The most obvious explanation for the discrepancy in the findings relative to criterion and normative referenced reading achievement lies with the question of congruence between the instructional content and the criterion measures. However, other explanations are possible and worthy of consideration. One possible explanation is that of differences in test format. When the instructional materials and criterion measures are devised by the same publisher, the likelihood of children being accustomed to
the test format is much greater than would be expected when the instructional materials and criterion measures are devised by different publishers. In the latter case, test format becomes a variable along with reading achievement. Symonds (1967) has stated, "Other things being equal the more common the experiences called for in a test are to the members of the group taking the test, the more reliable the test (p. 50)."

Another possible explanation for the discrepancy is a time factor. It is possible that one year is not sufficient time for the specific skills measured by the criterion referenced instrument to be translated into the generalized skills measured by the normative referenced instrument. The apparent benefits in criterion referenced reading achievement may become apparent in normative referenced reading achievement after two or more years.

The discrepancy in the findings regarding criterion referenced and normative referenced reading achievement, at this point, supports the argument that programed tutoring had limited value as a reading supplement for the pupils considered in this investigation. These findings, however, suggest that further investigation be considered in this area.

The second, third, fifth, and sixth hypotheses were not substantiated. These findings indicate that: (a) perceptual reasoning ability, as measured by the Coloured Progressive Matrices Tests, does not affect reading
achievement for lower achieving second grade children from low-income areas; and (b) tutored high perceptual reasoners and tutored low perceptual reasoners demonstrate similar reading achievement patterns relative to their control groups. These findings tend to support those reported by Carr (1962) and Wollenberg (1963), suggesting that the amount of profit from programmed instruction is independent of such factors as intelligence and aptitude.

To conclude that perceptual reasoning ability has no effect on reading achievement is disturbing to the investigator, not only from the standpoint of programmed tutoring, but from the standpoint of reading instruction in general. This conclusion suggests that perceptual reasoning ability, as a factor of the capacity for intellectual activity, is insignificant relative to cognitive achievement in reading for lower achieving second grade children from low-income areas, and contradicts the work of other investigators (Jensen, 1967; Nelson & Edelstein, 1963; Green & Rohwer, 1971).

The data presented in Chapter 4 also suggest that under the programed tutoring condition, pupils of varying perceptual reasoning abilities tend to perform quite similarly in reading achievement; but under the non-tutoring condition, they tend to perform differently. This observation raises the possibility that programed tutoring produces a convergence in reading achievement among pupils of varying abilities. While the differences among the high,
average, and low perceptual reasoners in reading achievement were not large enough to be statistically significant, it is interesting to note that the high perceptual reasoners generally had the lowest mean scores.

Certainly the above interpretations are subject to the alternative explanation that the Coloured Progressive Matrices Tests used to classify the pupils as high, average, or low perceptual reasoners do not measure perceptual reasoning ability as purported by the author (Raven, 1965). The term "perceptual reasoning ability" may be incorrect. Green and Robwer (1971) used the Coloured Progressive Matrices as a measure of higher conceptual functioning, requiring abstract figural reasoning; and Martin and Wiechers (1954) indicated the "matrix" tests have their rationale in Spearman's cognitive principles; in order to act intelligently in any situation one requires the necessary information and the intellectual capacity to apprehend the situation and draw inferences from what he perceives.

Green and Ewert (1955) describe the Progressive Matrices as, "a test of fairly complex intellectual reasoning processes (p. 142)." Thus, while perceptual reasoning ability may be improper nomenclature, the descriptions presented above in concert with the correlations of the Progressive Matrices and other intelligence tests presented in Chapter 3 strongly suggest that the Progressive Matrices do measure capacities that should be relevant to cognitive achievement.
The findings of this study relative to reading achievement suggest to the investigator that programmed tutoring in the form employed in this study is of questionable value for populations of the type considered in the study. A number of interrelated factors, in all likelihood, bear on this matter, including individual pupil differences, program variables, assessment variables, and interpersonal relations. The findings seem to be highly supportive of the position taken by Jensen (1967).

Optimal educational results are produced by designing instruction in accord with individual differences, and this means something much more radical than merely having slow and fast tracks in school or simply allowing some students to take more time than others to learn the same amount of subject matter, taught to all students in the same way. The educational plight of the disadvantaged, I am convinced, is the result of our not having taken individual differences seriously enough (p. 47).

Programed tutoring in its present form allows pupils to progress through the same materials at different rates, and the technique provides some latitude in the program steps each pupil follows. Beyond this, however, it does not provide specifically for individual pupil differences in terms of such variables as sex, ability, race, language facility, interests, attitudes, family, and other out of school factors. The tightly prescribed tutor activities seem to be in conflict with the notion of individual pupil differences, in that all pupils are treated alike by the tutors. Cloward (1967), Smith (1971), Erickson (1971), Schoeller and Pearson (1970), and Shaver and Muhn
(1971) in reporting on reading tutoring programs with some degree of success stress the importance of rapport between the tutor and tutee. Tutor training was emphasized, but was referenced more toward understanding the child and his problems than tightly prescribed instructional techniques.

It may be that programed tutoring places too many restrictions on the tutor's actions, thus creating a human machine that the child does not understand. The warm interpersonal relationships considered important in other tutoring programs may be thwarted. Opportunities for the type of relationships described by Smith (1969) in Chapter 2 of this study (page 39) are limited.

Another factor that should be considered in interpreting the limited success of programed tutoring in this study is the cumulative deficit concept. Lower achieving second grade pupils have experienced one more year of frustration and failure than their first grade counterparts. The cumulative effect of this additional year may serve to mediate the effectiveness of programed tutoring when introduced in the second grade.

The quality of programming in the programed tutoring procedure is an important factor in the effectiveness of the technique, and should be considered as a possible explanation for the findings of this investigation. The research reviewed in Chapter 2 relative to the theory underlying programed instruction strongly suggests that programming variables are interrelated with
learner variables, population variables, subject matter variables, and the cognitive levels within a subject matter. Thus, in further development of the programed tutoring technique, these matters must be given attention.

One other factor that is a matter of concern to the investigator and a possible explanation for the findings of the study is the coordination between the supplemental programed tutoring and the classroom reading instruction. While the same reading series served as the basis for both the tutoring and the classroom reading instruction, close coordination between the tutors and classroom teachers was not an integral part of the program. The findings reported by Niedermeyer and Ellis (1971) relative to tutoring "as needed" and closely coordinated with the classroom program, and those reported by Klosterman (1970) using a diagnostic-structured tutoring program give credence to the matter of coordination.

It is clear from the findings of this study that lower achieving second grade children from low-income areas do have abilities that are not tapped by the classroom reading program, and programed tutoring does not capitalize on these abilities. It is evident that some children benefit from programed tutoring while others do not, and the factors relating to such differential benefits should be delineated rather than providing the treatment to all pupils in the population considered in this investigation.
Pupil Attendance and Mobility

The seventh and eighth hypotheses were not confirmed. Thus, from the findings of this study, it may be concluded that there is no relationship between programed tutoring and pupil attendance, and there is no relationship between programed tutoring and pupil mobility for the population under consideration.

These relationships were hypothesized on the basis of casual observations by individuals previously involved with programed tutoring at the first grade level. These hypotheses were advanced as a preliminary step to testing the effect of programed tutoring on attitudinal changes on the part of pupils and/or their parents. The failure to establish these relationships should not be construed to imply that programed tutoring has no effect in the affective domain. Pupil attendance and mobility are gross measures that would not be sensitive to all changes in the affective domain. Also, the primary focus of this investigation was reading achievement, thus the affective domain was given limited direct attention.

A time factor may also be operating with respect to changes in the affective domain. Increased cognitive success may not be translated into the affective domain immediately, and the effects of programed tutoring in the affective areas, if any, may not become apparent within the treatment period of one year.
Pupils Tutored in the First Grade

The data presented in Chapter 4 relative to pupils tutored in the first grade and either tutored or not tutored in the second grade were not subjected to statistical analyses. However, observation of these data reveals a number of differences between those pupils tutored in both the first and second grades and those tutored in only the first grade, as would be expected. All of these differences favor the pupils tutored both years. The two year tutees showed higher attendance rates, lower mobility rates, and greater reading achievement.

In terms of reading achievement, there is a reversal in the pattern shown by the pupils tutored only in the second grade. The two year tutees demonstrated relatively greater normative referenced than criterion referenced reading achievement when compared to their control group, and the pattern for pupils tutored in only the second grade was reversed.

These findings tend to support the time factors discussed earlier in this chapter, suggesting that the generalized skills assessed by normative referenced instruments and changes in the affective domain may develop over periods of time longer than one year. Another possible explanation is that a treatment period of one year is not sufficient duration to establish the benefits of programmed tutoring. Cost factors must be considered, however, in efforts to extend the treatment period.
It should be noted that the above comments are simply observations as the data regarding pupils tutored in the first grade were not subjected to statistical analyses, and real differences were not established. The samples were small, and these pupils were not representative of all pupils tutored in the first grade, but were drawn from those pupils who were least successful in the first grade tutoring program. Thus, any conclusions regarding this group must be considered carefully.

Recommendations for Further Study

While the findings of this study offer little encouragement for continuing programed tutoring as a supplement to reading instruction for lower achieving second grade children from low-income areas, it is not recommended that the procedure be abandoned on the basis of this one investigation. The data suggest a number of questions that need further exploration, and there is no substitute for replication to confirm or deny the findings of any given study.

It is recommended that the tutoring procedures and content of the programs be reviewed and modified in light of the factors discussed here, and the revised programs and procedures be tested with a population similar to that considered in this study. In connection with the second trial of programed tutoring at the second grade level, the evaluation plan should include procedures for the identification
of pupils who benefit differentially from the treatment, and should attempt to isolate the factors responsible for such differential benefits.

It is recommended that data be collected on a number of pupil variables such as sex, race, academic ability, language facility, interests, attitudes, family, and other out of school factors in an effort to confirm any relationships between these variables and programed tutoring. Similar data should be collected regarding the tutors to investigate tutor-pupil relationships.

It is recommended that two alternative procedures be incorporated in the second trial of programed tutoring, and that the effects of these procedures be investigated. These procedures are: (a) close coordination of the programed tutoring supplement and the classroom reading program; and (b) more flexible tutoring procedures which reduce the restrictions on the tutors' actions, allowing the tutors to establish closer, more human relationships with the tutees during the tutoring sessions. Appropriate tutor training should be provided.

It is recommended that the second trial of programed tutoring be designed to control for test format to provide a more realistic comparison between the relative amounts of criterion and normative referenced reading achievement. All pupils should receive practice exercises relative to the format of the normative referenced instrument used.
It is recommended that the potential benefits of programed tutoring in the affective domain be investigated more fully. More sensitive measures than attendance and mobility should be developed and employed to investigate possible effects of programed tutoring in the affective domain.

It is recommended that a larger sample of pupils tutored in the first grade be included in a second grade tutoring program to investigate the possible benefits of two years of tutoring suggested by this study. This investigation should be limited to pupils who qualify for a second year of tutoring. In connection with this recommendation, it is recommended that follow-up studies be conducted to investigate retention and delayed effects of the programed tutoring supplement.

Finally, it is recommended that efforts be continued and intensified to develop reading instructional techniques that will allow each pupil to achieve in a manner commensurate with his abilities. This is a recommendation for focus on the learner and what he brings to the learning situation, rather than a focus on more new programs and materials.
REFERENCES


Barrett, E. S. The relationship of the Progressive Matrices (1938) and the Columbia Mental Maturity Scale to the WISC. Journal of Consulting Psychology, 1956, 20, 294-296.


Ellson, D. G. Title I (ESEA) support for programed tutoring in 1968-69 and 1969-70. Unpublished manuscript, Department of Psychology, Indiana University, 1969c.


Ellson, D. G., Barber, L., & Harris, P. Follow-through on programed tutoring. Unpublished manuscript, Department of Psychology, Indiana University, 1969a.

Ellson, D. G., Barber, L. W., & Harris, P. L. A nationwide evaluation of programed tutoring. Unpublished manuscript, Department of Psychology, Indiana University, 1969b.


Harris, E. B. An experiment in the use of programmed linguistic reading materials to improve the reading ability of a group of second grade pupils who are reading below grade level. (Doctoral dissertation, University of Georgia) Ann Arbor, Mich.: University Microfilms, 1965. No. 65-4492.


Klosterman, R. The effectiveness of a diagnostically structured reading program. Reading Teacher, 1970, 24, 159-162.


Rogers, M. S. A study of an experimental tutorial reading program in which sixth-grade underachievers tutored third-grade children who were experiencing difficulty in reading. (Doctoral dissertation, University of Alabama) Ann Arbor, Mich.: University Microfilms, 1970. No. 70-9381.


Ruddell, R. E. Reading instruction in first grade with varying emphasis on the regularity of grapheme-phoneme correspondences and the relation of language structure to meaning. Reading Teacher, 1966, 19, 653-660.
Ruddell, R. B. Reading instruction in first grade with varying emphasis on the regularity of grapheme-phoneme correspondences and the relation of language structure to meaning--extended into second grade. Reading Teacher, 1967, 20, 730-739.


