

AN EVALUATION OF A PRESCHOOL PROGRAM FOR AT-RISK FOUR-YEAR-OLDS
IN SUFFOLK, VIRGINIA, PUBLIC SCHOOLS

Deran Richard Whitney

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
DOCTOR OF EDUCATION
in
EDUCATIONAL LEADERSHIP AND POLICY STUDIES

Dr. David Parks, Chairman

Dr. Linda T. Morgan

Dr. Richard Salmon

Dr. Travis Twiford

April 19, 2005

Blacksburg, Virginia

Key Words: Preschool, Early Intervention, At-Risk, Student Literacy Achievement, Reading
Performance

AN EVALUATION OF A PRESCHOOL PROGRAM FOR AT-RISK FOUR-YEAR-OLDS
IN SUFFOLK, VIRGINIA, PUBLIC SCHOOLS

Deran Richard Whitney

ABSTRACT

In the new federal legislation, No Child Left Behind, two of the main themes direct attention to accountability and achievement gaps within subgroup populations. Early intervention may address both of these issues. Specifically, this means quality preschool education. While many agree that preschool education offers much to young children, school leaders have a responsibility to present data to support the effectiveness of preschool programs. This is a report of an evaluation of the effects of a state-funded preschool program on students' language acquisition measured with the Phonological Awareness Literacy Screening—Kindergarten at the end of the program; total English, word analysis, understanding elements of literature, and understanding a variety of resource materials measured with Virginia's Standards of Learning test in grade three; and instructional reading level measured with the STAR Reading Assessment in grade four. The performance of students who attended the Early Start Preschool Program in Suffolk Public Schools was compared with the performance of students who were eligible to attend but did not attend the program.

Two-way ANOVAs with analyses of simple main effects following significant interactions were applied to the data from ten samples of preschool children (five samples that attended the Suffolk, Virginia, preschool program and five samples that did not attend) for each year between 1998 and 2002. There were few significant findings in this evaluation, and they were scattered across cohorts and dependent measures in no systematic order. Attending Early Start may have an effect on the literacy learning of children; however, this study did not provide

the evidence that one would like to see in support of those effects. There are some findings, however, that point to some effects of attendance, particularly on rhyme awareness, concept of word, grade three total English, and grade three understanding elements of literature. These are tentative findings, at best, because they were not found consistently across cohorts.

ACKNOWLEDGEMENTS

Thanks to my entire dissertation committee for the support and guidance throughout this process: Dr. Travis Twiford, Dr. Richard Salmon, Dr. Linda T. Morgan and Dr. David Parks. A very special thanks to Dr. David Parks for not only agreeing to serve as my chairman but for maintaining high expectations with much guidance.

Special thanks are extended to Dr. Milton R. Liverman and Dr. Sandra Witcher for their constant encouragement, knowledge, and support. Sincere thanks is extended to my parents, my two brothers and my sister for believing in me and providing ongoing unconditional support.

To my immediate family, my dear wife and best friend, Conjo V. Whitney, you have always made me feel as if I can accomplish anything. Your encouragement and patience are appreciated, and I feel that “we” have earned this together. Chelsea, my baby girl, thank you for allowing Daddy to write, attend class, and for understanding that I have not always been around as much as I wanted to be due to this task. I love you two girls with all my heart and could not have done this without your support and understanding.

TABLE OF CONTENTS

EXECUTIVE SUMMARY OF AN EVALUATION OF A PRESCHOOL PROGRAM

FOR AT-RISK FOUR-YEAR-OLDS IN SUFFOLK, VIRGINIA, PUBLIC SCHOOLS.. 1

FULL REPORT OF AN EVALUATION OF A PRESCHOOL PROGRAM FOR AT-RISK FOUR-YEAR-OLDS IN SUFFOLK, VIRGINIA, PUBLIC SCHOOLS..... 3

Context of the Study 3

Purpose of the Study 6

Evaluation Questions 7

Definitions of Terms 8

Early Start Program Overview..... 10

Theoretical Framework..... 13

Methods Applied to this Evaluation..... 14

Populations 15

Samples..... 16

Design of the Evaluation..... 17

Data Collection..... 17

Instruments..... 18

Phonological Awareness Literacy Screening--Kindergarten Overview. 20

English Standards of Learning Assessments Overview..... 21

STAR Reading Assessment Overview..... 21

Data Analysis..... 22

Findings.....	22
Differences Between Those Who Attended and Those Who Did Not Attend the Early Start Preschool Program on the Dependent Measures.....	23
Differences Between Males and Females on the Dependent Measures.....	25
Interactions Between Attended Early Start /Did Not Attend Early Start and Gender.....	26
Summary.....	36
Conclusions, Discussion, Implications for Practice, and Recommendations for Future Research.....	38
Conclusions.....	39
Discussion.....	39
Implications for Practice.....	42
Recommendations for Future Research.....	43
References.....	45
Appendices.....	50
A. Expanded Description of the Early Start Program.....	50
B. Expanded Description of Phonological Awareness Literacy Screening-- Kindergarten	53
Validity of the Phonological Awareness Literacy Screening	53
Reliability of the Phonological Awareness Literacy Screening	56
Scoring of the Phonological Awareness Literacy Screening.....	57
C. Expanded Description of the Virginia Standards of Learning	58
Validity of the Virginia Standards of Learning Assessments.....	58

Reliability of the Virginia Standards of Learning Assessments.....	59
Scoring of the Virginia Standards of Learning Assessments.....	60
D. Expanded Description of the STAR Reading Assessment	61
Validity of the STAR Reading Assessment	61
Reliability of the STAR Reading Assessment	61
Scoring of the STAR Reading Assessment	62
E. Institutional Review Board Exemption Approval.....	63
F. Tables of Descriptive Characteristics of Cohorts.....	64
G. Literature Review.....	72
Status of Preschools.....	73
Preschool Models.....	75
Landmark Preschool Studies.....	79
H. Analyses With Non-Significant Statistical Results.....	86

LIST OF TABLES

Table		Page
1	Numbers of Students in the Populations and Samples by Gender and Year.....	16
2	Measures of Criterion Variables Used for Each Cohort by Year Administered	19
3	Means and Standard Deviations for Phonological Awareness Literacy Screening--Kindergarten--Rhyme Awareness, 1998 Cohorts by Group (Control v. Treatment) and Gender.....	28
4	Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy Screening--Kindergarten--Rhyme Awareness, 1998 Cohorts.....	28
5	Means and Standard Deviations for the English Standards of Learning Grade Three Total English Test, 1999 Cohorts by Group (Control v. Treatment) and Gender	29
6	Two-Way ANOVA for Gender by Treatment/Control--English Standards of Learning Grade Three--Total English, 1999 Cohorts.....	29
7	Means and Standard Deviations for the English Standards of Learning Grade Three Subtest Understanding Elements of Literature, 1999 Cohorts by Group (Control v. Treatment) and Gender.....	29
8	Two-Way ANOVA for Gender by Treatment/Control--English Standards of Learning Grade Three Subtest Understanding Elements of Literature, 1999 Cohorts.....	30
9	Means and Standard Deviations for the English Standards of Learning Grade Three Subtest Understanding a Variety of Resource Materials, 1999 Cohorts by Group (Control v. Treatment) and Gender.....	30

Table		Page
10	Two-Way ANOVA for Gender by Treatment/Control-- English Standards of Learning Grade Three Subtest Understanding a Variety of Resource Materials, 1999 Cohorts.....	31
11	Means and Standard Deviations for the Phonological Awareness Literacy Screening--Kindergarten--Rhyme Awareness, 2001 Cohorts by Group (Control v. Treatment) and Gender.....	31
12	Two-Way ANOVA for Gender by Treatment/Control-- Phonological Awareness Literacy Screening-- Kindergarten--Rhyme Awareness, 2001 Cohorts.....	32
13	Means and Standard Deviations for the Phonological Awareness Literacy Screening--Kindergarten--Concept of Word, 2001 Cohorts by Group (Control v. Treatment) and Gender.....	32
14	Two-Way ANOVA for Gender by Treatment/Control-- Phonological Awareness Literacy Screening-- Kindergarten--Concept of Word, 2001 Cohorts.....	33
15	Means and Standard Deviations for the Phonological Awareness Literacy Screening--Kindergarten--Rhyme Awareness, 2002 Cohorts by Group (Control v. Treatment) and Gender.....	33
16	Two-Way ANOVA for Gender by Treatment/Control-- Phonological Awareness Literacy Screening-- Kindergarten--Rhyme Awareness, 2002 Cohorts.....	33
17	Means and Standard Deviations for the English Standards of Learning Grade Three Total English Test, Combined 1998-1999 Cohorts by Group (Control v. Treatment) and Gender.....	34
18	Two-Way ANOVA for Gender by Treatment/Control-- English Standards of Learning Grade Three--Total English--Combined 1998-1999 Cohorts.....	34

Table		Page
19	Means and Standard Deviations for the English Standards of Learning Grade Three Subtest Understanding Elements of Literature, Combined 1998-1999 Cohorts by Group (Control v. Treatment) and Gender.....	34
20	Two-Way ANOVA for Gender by Treatment/Control-- English Standards of Learning Grade Three Subtest Understanding Elements of Literature, Combined 1998-1999 Cohorts.....	35
21	Means and Standard Deviations for the English Standards of Learning Grade Three Subtest Understanding a Variety of Resource Materials, Combined 1998-1999 Cohorts by Group (Control v. Treatment) and Gender.....	35
22	Two-Way ANOVA for Gender by Treatment/Control English Standards of Learning Grade Three Subtest Understanding a Variety of Resource Materials, 1998-1999 Combined Cohorts.....	36
23	Summary of Significant Findings.....	37

LIST OF FIGURES

Figure		Page
1	Theoretical Framework: Components of early intervention related to student performance.....	13

EXECUTIVE SUMMARY OF AN EVALUATION OF A PRESCHOOL PROGRAM FOR AT-RISK FOUR-YEAR-OLDS IN SUFFOLK, VIRGINIA, PUBLIC SCHOOLS

There were two purposes for this evaluation. The first purpose was to evaluate the Early Start preschool program in Suffolk Public Schools to determine if the funds designated for the program are benefiting children. The second purpose was to gather data to direct and initiate any revisions or changes needed in the existing Early Start program.

This evaluation took place in the fall of 2004 in Suffolk Public Schools in Suffolk, Virginia. It is a two-group comparison. The performance of those students who attended the Early Start program during the years 1998 through 2002 was compared to the performance of students who were eligible to attend the program during the same years but did not attend. Not attending the Early Start program was a decision each parent made with no consistent reason given; however, some of those who chose not to attend had moved away from Suffolk and later returned during the student's kindergarten year.

The criteria for effectiveness of the program were three standard assessments that are common to students in Suffolk Public Schools. Each of the assessments is required at particular grade levels and measures students' performance. Two of the assessments have several subtests. These assessments are the Phonological Awareness Literacy Screening--Kindergarten and Virginia's English Standards of Learning Assessment. The third assessment is the STAR Reading Assessment.

Two-way ANOVAs with analyses of simple main effects following significant interactions were applied to the data from ten samples of preschool children (five samples that attended the preschool program and five samples that did not attend) for each year between 1998 and 2002. There were few significant findings in this evaluation, and they were scattered across

cohorts and dependent measures in no systematic order. Attending Early Start may have an effect on the literacy learning of children; however, this study did not provide the evidence that one would like to see in support of those effects. There are some findings, however, that point to some effects of attendance, particularly on rhyme awareness, concept of word, grade three total English, and grade three understanding elements of literature. These are tentative findings, at best, because they were not found consistently across cohorts.

Previous research indicates that those who participate in preschool early intervention programs experience more positive academic and social success than those who do not receive early intervention (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Marcon, 2002; Weikart & Schweinhart, 1978). The results of this evaluation do not contribute much to this literature. Preschool intervention of the type used in Suffolk, Virginia, may make a difference in student success, but additional evaluation is needed to make any definitive statements.

FULL REPORT OF AN EVALUATION OF A PRESCHOOL PROGRAM FOR AT-RISK
FOUR-YEAR-OLDS IN SUFFOLK, VIRGINIA, PUBLIC SCHOOLS

Context of the Study

This is the report of an evaluation of a state and locally funded preschool program. The criteria for assessing effectiveness were measures of students' acquisition of language skills at the end of the program and in third and fourth grades. This evaluation is important for two reasons. First, it relates specifically to the federal legislation, No Child Left Behind, signed by President Bush in 2002. The No Child Left Behind Act of 2001 is a revision of the Elementary and Secondary Education Act and contains five performance goals: (1) All students will reach high standards, at a minimum attaining proficiency or better in reading/language arts and mathematics by the 2013-2014 school year; (2) All Limited English Proficient (LEP) students will become proficient in English and reach high academic standards, at a minimum attaining proficiency or better in reading/language arts and mathematics; (3) By 2005-2006, all students will be taught by highly qualified teachers; (4) All students will be educated in learning environments that are safe, drug-free, and conducive to learning; and (5) All students will graduate from high school (U.S. Department of Education, 2002). The first, third, and fifth performance goals speak specifically to the preschool issues--student performance, quality staff, and accountability. As one reviews the No Child Left Behind Act more closely, the student performance goal addresses closing the achievement gap through accountability. In addition, it includes provisions to ensure all students are reading on grade level by the end of third grade. Preschool experiences play a role in these specific areas.

Current preschool programs vary in structure, curriculum, and expectations. There are private programs, public school programs, and federal government initiatives such as Head Start.

Head Start programs, which traditionally have targeted at-risk students, have been copied by many programs serving children who meet particular criteria. Recently, more options appear to be available for all students regardless of their background and economic circumstances of their parents or guardians. In 1999 the president of the American Federation of Teachers, Sandra Feldman, called for a national commitment to provide school for all three and four year olds. Most advocates of public preschool education believe that early education of low income children is an investment that pays off in the long term by reducing the number of children who perform poorly in school (Evers, 2001).

Over the years, preschool education has been credited with improving students' academic abilities as they enter school. It is believed that effective preschool programs maintain a specific structure with highly trained staff (Hohmann & Weikart, 2002).

Economic trends have increased the need for both parents to work in many families. According to the U.S. Department of Labor, in 2003, 53.2% of families with children younger than six years of age had both parents employed (United States Department of Labor, 2004). This has increased the need for quality preschool programs.

Research supports the effectiveness of programs for young children from low income families (Neuman, Roskos, Vukelich, & Clements, 2003). This evidence has been used to support policy initiatives for the development or expansion of preschool programs. These preschool interventions have been designed to build a solid foundation to help better prepare students for elementary school and are commonly focused on family and parental interactions with their children (Warger, 1988).

For about 40 years, local, state, federal, and private sources have funded early education programs specifically for low income children. Many ask, "Should the public school system

expand to educate all three and four year olds?” Those who support what is sometimes termed universal preschool education frequently argue that most parents fail to provide their children with the experiences and the environment necessary to promote healthy development. Andrew Rotherm of the Progressive Policy Institute believes one can not talk about closing the achievement gap without talking about preschool programs (Starr, 2002).

Overall, data suggest that America’s flexible approach to early education of young children is working with limited success. According to the United States Department of Education report “America’s Kindergartens,” U.S. kindergarteners have a strong start. Upon kindergarten entry, 94 percent of the children are proficient at recognizing numbers, shapes, and counting to ten. Ninety-two percent are eager to learn, and all but three percent are in good health (Boyer, 1991).

In January 1994 the Virginia Commission on Equity in Public Education adopted and endorsed four major programs as the core elements in their recommendations to the 1994 General Assembly. The recommendations focused on programs that had been shown to improve educational achievement. One of these programs is preschool education for at-risk four-year olds (Virginia Department of Education, 2001).

The 1995 Virginia General Assembly, through passage of the Omnibus Education Act (HB2542) and the Appropriation Act, reinforced all components of the 1994 package and provided for expansion of the initiative for at-risk four-year-olds (Virginia Department of Education, 2001). State funds are available to provide comprehensive preschool programs to Virginia's at-risk four-year-olds who are not being served by federal programs such as Head Start or Title I. Each local school division that qualifies must submit an application each year for funding. The primary purpose of the grant is to reduce disparities among young children upon

formal school entry and to reduce or eliminate risk factors that lead to early academic failure. To obtain state funding, localities must submit a plan for their program that includes four services: a quality education program, parental involvement, health services, and a planning team.

The Early Start Preschool Program in the Suffolk Public Schools is a preschool program that began in 1995. Although many would claim that the Early Start program is successful, no formal evaluation has been completed regarding its effectiveness. This is an assessment of the effectiveness of the program regarding students' performance in reading literacy.

As I began this evaluation I questioned how many public schools have preschool programs. I discovered that in 2001 35% of the United States public elementary schools offered some type of preschool education (National Center for Education Statistics, 2003). Considering the overall low percentage of programs in public schools, I believe it is important that we consider the quality and effectiveness of preschool programs to demonstrate their worth and value. Research reviewing students' preschool experiences and education reveals positive outcomes (Love, Schochet, & Meckstroth, 1996). Researchers (Reynolds & Temple, 1998; Schweinhart, Weikart, & Larner, 1986) agree that the primary goal of preschool education is to promote a smooth transition to school. They found that the problems of many minority children who experience school difficulties can be traced to experiences they had early in their school careers (Alexander & Entwisle, 1988; Belsky & MacKinnon, 1994).

Purpose of the Study

I had two purposes for this evaluation. The first purpose was to evaluate the program to determine if the funds designated for the Early Start program are benefiting children. The Early Start program has approximately \$900,000.00 included in its budget yearly. If the program is not meeting expectations, there is a responsibility to consider if the funds should be spent

differently. The second purpose was to gather data to direct and initiate any changes in the Early Start program. If children are not progressing as expected, recommendations for changes in the program will be made.

Evaluation Questions

Three main evaluation questions with eleven specific evaluation criteria guided the evaluation.

1. How do children who participate in the Suffolk Public Schools' Early Start program differ from children who do not participate in the Early Start program on the Phonological Awareness Literacy Screening--Kindergarten assessment as measured by:

- a. summed score?
- b. rhyme awareness?
- c. beginning sounds?
- d. alphabet knowledge?
- e. letter sound identification?
- f. concept of word?

2. How do children who participate in the Suffolk Public Schools' Early Start program differ from children who do not participate in the Early Start program on the third grade English Standards of Learning Test as measured by:

- a. total English?
- b. word analysis?
- c. understanding elements of literature?
- d. understanding a variety of resource materials?

3. How do children who participate in the Suffolk Public Schools' Early Start program

differ from children who do not participate in the Early Start program on instructional reading level at the fourth grade?

Definitions of Terms

The terms listed below are defined for the purposes of this evaluation. More extensive descriptions of the tests are in Appendixes B through D.

Rhyme Awareness

Rhyme awareness is the ability to recognize words that sound alike. Out of a set of three pictures, students are asked to identify the one that rhymes with the target picture. There are ten items children must complete (Invernizzi, Meier, Swank, & Juel, 2003). This is a subtask of the Phonological Awareness Literacy Screening--Kindergarten. The raw score (range of 0-10) was used for this evaluation.

Beginning Sound Awareness

Beginning sound awareness is the ability to identify or match the initial sound of a word. This includes identifying the picture that has the same initial sound as the target picture, identifying words orally according to shared beginning sound, and, out of a set of three pictures, identifying the one that has the same beginning sound as the target picture. There are ten items children must complete (Invernizzi, Meier, Swank, & Juel, 2003). This is a subtask of the Phonological Awareness Literacy Screening--Kindergarten. The raw score (range of 0-10) was used for this evaluation.

Alphabet Knowledge

Alphabet knowledge involves the ability to identify letters of the alphabet and to name the 26 lower case letters of the alphabet (Invernizzi, Meier, Swank, & Juel, 2003). This is a subtask of the Phonological Awareness Literacy Screening--Kindergarten. The raw score (range

of 1-26) was used for this evaluation.

Letter Sound Identification

Letter sound identification is the ability to produce the letter sounds of 23 upper-case letters of the alphabet and match consonant and short vowel sounds to appropriate letters (Invernizzi, Meier, Swank, & Juel, 2003). This is a subtask of the Phonological Awareness Literacy Screening--Kindergarten. The raw score (range of 0-23) was used for this evaluation.

Concept of Word

The concept of word is the ability to follow words from left to right and from top to bottom on a printed page; the concept of word task measures children's ability to (a) accurately touch words in a memorized rhyme, (b) use context to identify individual words within a given line of text, and (c) identify words presented in isolation, not in a passage or story (Invernizzi, Meier, Swank, & Juel, 2003). This is a subtask of the Phonological Awareness Literacy Screening—Kindergarten. The raw score (range of 0-12) was used for this evaluation.

Word Analysis

Word analysis is the ability to use word-attack skills to decode words and spell words to read fluently. These skills include the use of phonics (vowel patterns and consonant combinations), structural analysis (roots and affixes), and context (to read words with multiple pronunciations) (Virginia Department of Education, 1999). A raw score of 0–10 can be earned in this reporting category of the 3rd grade English Standards of Learning test. The raw score was used for this evaluation.

Understanding a Variety of Resource Materials

Understanding a variety of resource materials is the ability to apply different strategies to read fiction, poetry, nonfiction, content texts, and other printed materials and to apply meaning

clues, language structure, and phonetic strategies when needed (Virginia Department of Education, 1999). A raw score of 0–10 can be earned in this reporting category of the 3rd grade English Standards of Learning test. The raw score was used for this evaluation.

Total English Scaled Score

The total English scaled score is the score students receive on the 3rd grade English Standards of Learning test. The scaled score ranges from 300-600.

Instructional Reading Level

The instructional reading level is the reading level at which students can recognize words and understand instructional material with some assistance. A criterion-referenced score presented in years and months is an estimate of the most appropriate level of reading material for instruction. Instructional reading level was measured with the STAR Reading Assessment (Renaissance Learning, 2000) in the 4th grade.

Suffolk Public Schools' Early Start Program

This is an at-risk four-year-old preschool program in Suffolk Public Schools made possible through the Virginia Preschool Initiative (Virginia Department of Education, 2001). The program has been in operation since 1995 (see the expanded description in Appendix A).

Early Start Program Overview

The Early Start program serves children who are exposed to what is considered high risk factors and situations that are likely to have an adverse impact on their social, emotional, physical, and cognitive development as defined by the Virginia Department of Education (Virginia Department of Education, 2001). Risk factors that are considered for eligibility purposes include, but are not limited to, economics, single parent families, unemployment or underemployment, grandparents as custodial parents, developmental delays, parents who have

not finished high school, foster children, and other family situations such as incarceration, drugs, violence, and homelessness. Suffolk Public Schools considers each of these risk factors when qualifying students for the Early Start program.

The Early Start program was initiated in the fall of 1995. The program is free to all students admitted to the program and is targeted to children who will be four years old on or before September 30 of each year. This preschool program is made possible through the Virginia Preschool Initiative (Virginia Department of Education, 2001). The Virginia Preschool Initiative stems from legislation that offers funding for preschool programs for four-year-olds who are at risk. Transportation is provided by the Suffolk Public Schools. Breakfast and lunch are provided by the school, and parental participation is required.

The Early Start program is housed at three elementary schools throughout the Suffolk school division. They are Booker T. Washington Elementary, which is located in the downtown area of Suffolk; Mount Zion Elementary, which is located in the central part of Suffolk; and Florence Bowser Elementary, which is located in the northern part of Suffolk. Although only three schools house the Early Start program, the participants feed into all of Suffolk's 12 elementary schools. The school system is a rural district that appears to be moving towards more suburban.

The student to adult ratio is 8:1 with a maximum of 16 children in each Early Start classroom. All teachers have a bachelor's degree in elementary education or early childhood education. Staff development opportunities for staff members are continuous, and teachers are encouraged to maintain professional growth goals.

The Early Start curriculum was based totally on the High/Scope educational approach to learning for the first five years. For the last three years the Early Start program has included the

High/Scope approach combined with an early literacy focus. The literacy focus requires teachers to use such literature and reading skills as comprehension, alphabetic knowledge, letter identification, and phonological awareness throughout reading readiness instruction. Children in High/Scope settings are encouraged to make choices about materials and activities throughout the day. As they make their choices and plans, they explore, ask and answer questions, solve problems, and interact with classmates and adults. In this kind of environment, children naturally engage in "key experiences"; i.e., activities that foster developmentally important skills and abilities (Graves, 2002). An expanded description of the Early Start program is in Appendix A.

Theoretical Framework

The theoretical framework for this evaluation predicts that students who have participated in the Early Start program will benefit academically. The Early Start program is an early intervention program with specific components. It is believed that these components are necessary to the program's success. It is considered that students who attend the Early Start program will perform better than those who do not attend the Early Start program. These components are parental involvement, appropriate curriculum, and quality staff (see Figure 1).

Parental involvement is required in the Early Start program. Parents must attend workshops during the school year. Workshop topics include parenting skills, appropriate nutrition, discipline, helping your child at home, and getting ready for school. Parents are required to allow a family service worker into their home to assist as needed. Teachers make direct contact with parents through phone calls, written communication, and home visits. Parent-teacher conferences are held at least twice a year, and student progress is discussed.

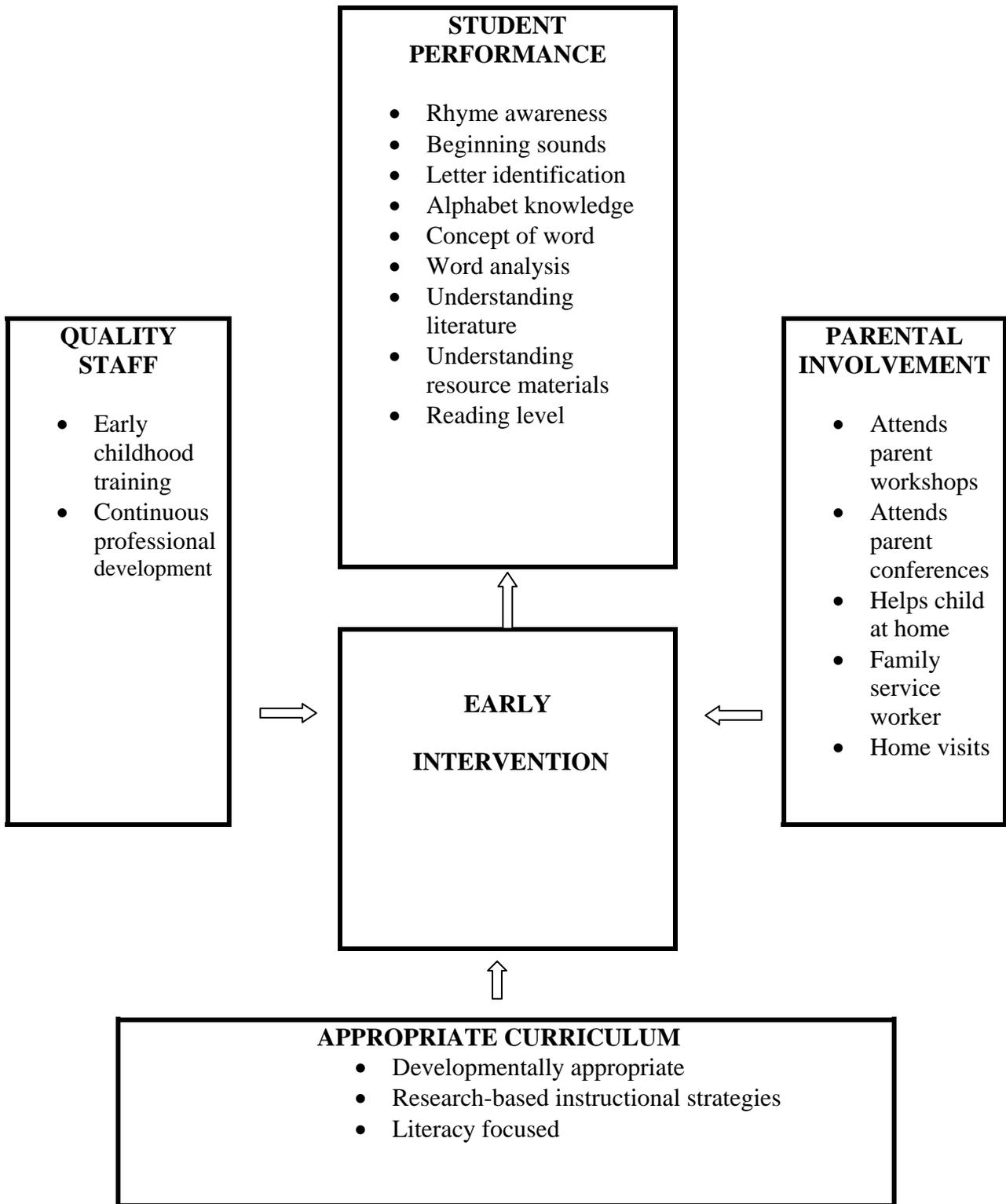


Figure 1. Theoretical Framework: Components of early intervention related to student performance.

I strongly believe an appropriate curriculum is necessary to have an effective preschool program. Kessler (1991) reported that there is much debate over how to best educate young children. However, research does support the positive results of preschool programs (Lazar, Darlington, Murray, & Snipper, 1982).

The Early Start program's curriculum includes the High/Scope educational approach to learning with a focus on early literacy. In a High/Scope setting children are encouraged to make choices about materials and activities throughout the day. As they make their choices and plans, they explore, ask and answer questions, solve problems, and interact with classmates and adults. In this kind of environment, children naturally engage in "key experiences"; i.e., activities that foster developmentally important skills and abilities (Graves, 2002).

Maintaining quality staff is another component of an effective preschool program. Of the 46,000 teachers who taught preschool classes during the 2000-2001 school year, 86% had a bachelor's degree or higher (National Center for Education Statistics, 2003). To have effective preschool programs, quality staff must be employed and developed.

The Early Start program has consistently contained components the Virginia Department of Education (Virginia Department of Education, 2001) and researchers (Hohmann & Weikart, 2002) believe to promote student performance. While many may believe the Early Start program in the Suffolk Public Schools is a quality program, the results of this evaluation provide specific information related to this claim.

Methods Applied to this Evaluation

The evaluation plan is described in this section. A description of the populations, the samples, the design of the evaluation, and the criteria for evaluating effectiveness are included.

Populations

Over the course of five years of Early Start, 800 students have participated in the program. This evaluation used 10 populations from the years 1998, 1999, 2000, 2001, and 2002. Five of the populations participated in the Early Start program, and five of the populations were eligible to participate but did not. Thus, there was one treated and one untreated population for each of the five years. The populations were 99.5 % African-American students. The populations consisted of students from each of the elementary schools in Suffolk, all of which follow the same Suffolk Public Schools curriculum. Specific characteristics of each population are in Table 1.

The five treated populations were identified by obtaining a roster of all the students who participated in the program during the listed years. The rosters were obtained through the student information management system used by Suffolk Public Schools. Students were included in the populations only if it was clearly noted in their cumulative record that they participated in the Early Start program and that they had attended Suffolk Public Schools throughout their elementary years. Although 160 students participated each year, some students left Suffolk Public Schools and were not included in the population. The treated populations included representation from each of the elementary schools in Suffolk.

The five untreated populations were students who were eligible to participate in the Early Start program but did not. They, too, represented each of the elementary schools. Student records were checked to ensure students were eligible to participate, did not participate in a preschool program, and that they had attended Suffolk Public Schools throughout their elementary years.

Table 1
Numbers of Students in the Populations and Samples by Gender and Year

Year	Population			Sample		
	<i>N</i>	Male	Female	<i>N</i>	Male	Female
Treatment (Attended Early Start)						
1998	145	77	68	25	11	14
1999	155	78	77	25	13	12
2000	153	81	72	25	14	11
2001	155	90	65	25	13	12
2002	157	94	63	25	14	11
Control (Did not attend Early Start)						
1998	32	16	16	25	12	13
1999	34	16	18	25	8	17
2000	36	20	16	25	13	12
2001	35	20	15	25	13	12
2002	31	18	13	25	11	14

Samples

Ten random samples of twenty-five students each were used in this evaluation. Five of the samples were selected from the five populations of those who attended the Early Start program during 1998, 1999, 2000, 2001, and 2002. The other five samples were selected from the populations of those who were eligible to attend but did not attend the Early Start program during 1998, 1999, 2000, 2001, and 2002. The samples were 100% African-American students. The samples were students from all elementary schools in Suffolk, each of which follows the same curriculum.

These years and samples were chosen for three main reasons: (1) Resources and time for collecting the data were limited; (2) no curriculum changes were made in the program during these years as opposed to the first two years of the program when the curriculum was not well

established; (3) the samples were small enough to be manageable but large enough to detect differences and make inferences to the populations.

In selecting samples, I viewed the roster of each year, assigned each student a number, and used a random numbers table to select 25 students from each population. This procedure was used for all cohorts, those who attended and those who did not attend Early Start. Students were assigned to the cohort year based on the year they attended Early Start. If a student attended Early Start in 1998, the student was placed in the 1998 treatment cohort. Specific numbers in the samples are in Table 1.

Design of the Evaluation

The evaluation design is discussed in this section. The criteria of effectiveness are described along with the measurement instruments. This evaluation took place in the fall of 2004 in Suffolk Public Schools. It is a two group comparison. The performance of those students who attended the Early Start program during the years of 1998 through 2002 was compared to the performance of those students who did not attend the Early Start program during the same years.

Data Collection

The initial step in collecting data involved seeking permission from the superintendent of Suffolk Public Schools. Once permission was granted, I submitted my proposal to the Virginia Tech Institutional Review Board (IRB) and received an exempt approval (see Appendix E). Next, I obtained a roster of all students who attended the Early Start program by year from 1998 through 2002. This list was compared to the students' record to verify that they actually attended the program. Using the school system's student management system, Pentamation, I obtained the roster which included complete demographics of each student. This information was necessary for grouping students during statistical analyses. The control cohort consisted of those students

who were eligible for the Early Start program but did not attend. The Suffolk Public Schools' Pentamation, the student management system, was used to obtain demographics for the control cohort.

A variety of avenues was used to obtain the data on the dependent variables. The raw scores for word analysis skills, understanding elements of literature, and understanding a variety of resource materials were obtained from the Standards of Learning English test. These scores, along with the scaled score of the English Standards of Learning, were obtained by using the Data Disaggregator. This Standards of Learning database may be used to view students' and schools' performance on each Standards of Learning test administered. It is provided by the Virginia Department of Education.

The summed score and raw scores for rhyme awareness, beginning sound awareness, alphabet knowledge, letter sound identification, and concept of word subtests of the Phonological Awareness Literacy Screening--Kindergarten (PALS-K) were used in this evaluation.. These scores were obtained through the Phonological Awareness Literacy Screening--Kindergarten website. Instructional reading level scores were obtained from the STAR Reading Assessment, and were obtained manually from each student's cumulative record.

Instruments. The criteria for effectiveness for this evaluation were three standard assessments that are common to students in Suffolk Public Schools. Table 2 contains the measures of the criterion variables by year and cohort. Each of the assessments is required at particular grade levels and is used to assess students' performance. Two of the assessments have several subtests. These assessments are the Phonological Awareness Literacy Screening--Kindergarten and the third grade English Standards of Learning test. The Phonological

Table 2
Measures of Criterion Variables Used for Each Cohort by Year Administered

Measure	1998		1999		2000		2001		2002		1998-1999	1998-2002
	ES	CG										
PALS-K summed score	'99	'99	'00	'00	'01	'01	'02	'02	'03	'03		X
PALS-K rhyme awareness	'99	'99	--	--	'01	'01	'02	'02	'03	'03		X
PALS-K beginning sounds	'99	'99	--	--	'01	'01	'02	'02	'03	'03		X
PALS-K alphabet knowledge	'99	'99	--	--	'01	'01	'02	'02	'03	'03		X
PALS-K letter identification	'99	'99	--	--	'01	'01	'02	'02	'03	'03		X
PALS-K concept of word	'99	'99	--	--	'01	'01	'02	'02	'03	'03		X
SOL Eng (3 rd)	'02	'02	'03	'03	--	--	--	--	--	--	X	
SOL word analysis(3 rd)	'02	'02	'03	'03	--	--	--	--	--	--	X	
SOL literature(3 rd)	'02	'02	'03	'03	--	--	--	--	--	--	X	
SOL resources(3 rd)	'02	'02	'03	'03	--	--	--	--	--	--	X	
Instructional reading level (4 th)	'03	'03	--	--	--	--	--	--	--	--	--	--

Note. Dashes mean data unavailable for the cohort. ES = Early Start cohort--attended; CG = Control cohort—did not attend. The listed year for each cohort indicates when the assessment was administered. X=Data were aggregated over the years indicated.

Awareness Literacy Screening--Kindergarten measures students' performance on rhyme awareness, beginning sounds, alphabet knowledge, letter sound identification, and concept of word. The Virginia English Standards of Learning Assessment measures students' performance on word analysis, understanding literature, and understanding a variety of resource materials. The STAR Reading Assessment was used to measure students' instructional reading level. In the next section, a brief overview of each assessment is provided. An expanded description of the Phonological Awareness Literacy Screening--Kindergarten is in Appendix B. An expanded description of the Virginia Standards of Learning Assessments is in Appendix C, and an expanded description of the STAR Reading Assessment is in Appendix D.

Phonological Awareness Literacy Screening—Kindergarten (PALS-K) overview. The Phonological Awareness Literacy Screening is a state-provided screening measure used by 98% of the schools in the Commonwealth of Virginia (Invernizzi, Meier, Swank, & Juel, 2003). There are two Phonological Awareness Literacy Screening instruments: Phonological Awareness Literacy Screening--Kindergarten, for students in kindergarten, and Phonological Awareness Literacy Screening 1-3, for students in grades 1-3. Phonological Awareness Literacy Screening--Kindergarten was used in this evaluation. The tool measures students' knowledge of literacy skills, including phonological awareness, alphabet knowledge, knowledge of letter identification, and concept of word. Each of the subtests was used in this evaluation. The major purpose of Phonological Awareness Literacy Screening--Kindergarten is to identify students who are performing below expected grade level in the areas listed above and who may be in need of additional reading instruction (Invernizzi, Meier, Swank, & Juel, 2003). In the Suffolk Public Schools, teachers are encouraged to use the Phonological Awareness Literacy Screening primarily as a diagnostic tool to guide their instruction. The assessment is administered twice a

year, once in the fall and once in the spring.

Validity and reliability for the Phonological Awareness Literacy Screening--Kindergarten have been tested and established. Over five years of statistical analysis, the test has proven to be steady, reliable, and consistent for many different students (Invernizzi, Meier, Swank, & Juel, 2003). See Appendix B for expanded descriptions of the validity and reliability of the Phonological Awareness Literacy Screening--Kindergarten.

English Standards of Learning Assessments overview. The Virginia Standards of Learning Assessments are tests that measure students' performance on Virginia's objectives for student learning and achievement in grades K–12. These objectives are referred to as the Standards of Learning (SOLs). Content areas are English, mathematics, science, history/social science, technology, the fine arts, foreign language, health, physical education, and driver education. Standards of Learning assessments are administered in the spring of each year to students in grades 3, 5, 8, and certain courses in high school.

The Standards of Learning assessment that was used in this evaluation is the third grade English test. The reading portion of the test is multiple choice and includes reporting categories: word analysis, understanding a variety of resource materials, and understanding elements of literature (Virginia Department of Education, 1999).

The Standards of Learning test development process includes many measures for assessing validity and reliability. The tests meet the requirements for high-stakes tests for the validity and reliability of scores and content (Virginia Department of Education, 1999). Expanded descriptions of the validity and reliability of the SOL tests are in Appendix C.

STAR Reading Assessment overview. The STAR Reading Assessment is a computer adaptive norm-referenced reading test that provides reading scores for grades 1 - 12. The scores

include grade equivalent, percentile ranks, and normal curve equivalents. Teachers in grades 2 - 5 in Suffolk Public Schools are required to administer the STAR reading assessment three times a school year, at the beginning, during the middle, and at the end of the school year. Teachers of first graders are expected to administer STAR to students once they have obtained a minimum of a 100-word reading vocabulary, which is often by second semester. The assessment provides teachers with a quick estimate of the students' reading level that is relative to national norms. The assessment is taken individually by students on a computer. Although the test is mostly norm-referenced, it provides a criterion-referenced score, the student's Instructional Reading Level (IRL), which was used in this evaluation. The ease of test administration and high reliability and validity estimates make this a good criterion measure for this evaluation. An expanded description of the STAR Reading Assessment's validity and reliability is in Appendix D.

Data Analysis

The statistical analysis addressed the three evaluation questions and the specific measurements. A two-way analysis of variance (ANOVA) was used for each dependent variable. The independent variables were attended or did not attend the Early Start program and student gender. The dependent variables were (1) kindergarten phonological awareness summed score (2) kindergarten rhyme awareness raw score, (3) kindergarten beginning sound awareness raw score, (4) kindergarten alphabet knowledge raw score, (5) kindergarten letter identification raw score, (6) kindergarten concept of word raw score, (7) third grade word analysis raw score, (8) third grade understanding a variety of resource materials raw score, (9) third grade understanding elements of literature raw score, (10) third grade English SOL scaled score, and (11) fourth grade instructional reading level.

Findings

Descriptive statistics for each cohort are in Appendix F. A word of caution at the start: These findings must be interpreted in light of the large number of tests that were conducted. One hundred thirty-two statistical tests (two main effects and one interaction effect in each of 44 two-way ANOVAs) were run. Tests for simple main effects were run only if the interaction term was significant in a two-way ANOVA. Nineteen (19/132 = 14.4%) of the tests for main effects and interactions [seven (7/44 = 15.9%) for gender differences, seven (7/44 = 15.9%) for treatment/control differences, and five (5/44 = 11.4%) for interactions between gender and treatment/control] were significant at $\alpha=.05$. Ten (50%) of the twenty simple main effects that were tested were significant at $\alpha=.05$. Overall, this means that those in the program fared better on some measures in some years; however, on many measures (39/44 = 88.6%) in many years program participants fared only equally as well as or worse (2/44 = 4.5%) than those who did not participate. ANOVA tables for non-significant findings are in Appendix H. A summary of results follows:

Differences Between Those Who Attended and Those Who Did Not

Attend the Early Start Preschool Program on the Dependent Measures

Those who attended and those who did not attend Early Start differed on five of the dependent measures for various cohorts (in parentheses): rhyme awareness (2001, see Tables 11 & 12), concept of word (2001, see Tables 13 & 14), third grade total English (1999, see Tables 5 & 6), third grade understanding elements of literature (1999, see Tables 7 and 8; 1998-1999 combined, see Tables 19 & 20), and third grade understanding a variety of resources (1999, see Tables 9 & 10; 1998-1999 combined, see Tables 21 & 22). Those who attended Early Start in the identified cohorts scored significantly higher than those who did not attend Early Start on rhyme

awareness, concept of word, third grade total English, and third grade understanding elements of literature. Those who did not attend Early Start in the identified cohorts scored higher than those who attended Early Start on third grade understanding of a variety of resource materials. There were no differences between those who attended Early Start and those who did not attend Early Start on the following dependent measures for the cohorts indicated (see Appendix H for the data tables):

Phonological Awareness Literacy Screening summed score (1998, 1999, 2000, 2001, 2002, all cohorts combined).

Phonological Awareness Literacy Screening rhyme awareness (1998, 2000, 2002, all cohorts combined)^a

Phonological Awareness Literacy Screening beginning sounds (1998, 2000, 2001, 2002, all cohorts combined)^a

Phonological Awareness Literacy Screening alphabet knowledge ((1998, 2000, 2001, 2002, all cohorts combined)^a

Phonological Awareness Literacy Screening letter identification (1998, 2000, 2001, 2002, all cohorts combined)^a

Phonological Awareness Literacy Screening concept of word (1998, 2000, 2002, all cohorts combined)^a

Standards of Learning assessment of third grade total English (1998, 1998 and 1999 combined)^b

Standards of Learning assessment of word analysis (1998, 1999, 1998 and 1999 combined)^b

Standards of Learning assessment of understanding elements of literature (1998)^b

Standards of Learning assessment of understanding a variety of resource materials
(1998)^b

STAR Reading Assessment, reading level at fourth grade (1998)^c

^aData not available for 1999 cohorts.

^bData not available for 2000, 2001, 2002 cohorts.

^cData not available for 1999, 2000, 2001, 2002 cohorts.

Differences Between Males and Females on the Dependent Measures

Differences between males and females, in the absence of information about the effects of attending Early Start, were of minor interest in this study; however, since the data were available, these findings are presented. Data on interactions between attending and not attending Early Start and gender will be presented later.

Females did better on six of the seven tests on which significant differences were found. They scored higher on third grade total English (1999, see Tables 5 & 6; 1998 and 1999 combined, see Tables 17 & 18), third grade understanding elements of literature (1999, see Tables 7 & 8; 1998 and 1999 combined, see Tables 19 & 20), and third grade understanding a variety of resource materials (1999, see Tables 9 & 10; 1998 and 1999 combined, see Tables 21 & 22). Males scored higher on rhyme awareness than females (2002, see Tables 15 & 16). There were no differences between males and females on the following dependent measures:

Phonological Awareness Literacy Screening summed score (1998, 1999, 2000, 2001, 2002, all cohorts combined).

Phonological Awareness Literacy Screening rhyme awareness (1998, 2000, 2001, all cohorts combined).^a

Phonological Awareness Literacy Screening beginning sounds (1998, 2000, 2001, 2002, all cohorts combined).^a

Phonological Awareness Literacy Screening alphabet knowledge (1998, 2000, 2001, 2002, all cohorts combined).^a

Phonological Awareness Literacy Screening letter identification (1998, 2000, 2001, 2002, all cohorts combined).^a

Phonological Awareness Literacy Screening concept of word (1998, 2000, 2001, 2002, all cohorts combined).^a

Standards of Learning assessment of word analysis (1998, 1998 and 1999 combined).^b

STAR Reading Assessment, reading level at fourth grade (1998).^c

^aData not available for 1999 cohorts.

^bData not available for 2000, 2001, 2002 cohorts.

^cData not available for 1999, 2000, 2001, 2002 cohorts.

Interactions Between Attended Early Start/Did Not Attend Early Start and Gender

Interaction effects are probably the most interesting of all of the findings. Interactions indicate whether one group or another fares better under one of the treatments. In this case, the question is whether males or females fared better if they attended Early Start or not. Statistically significant interactions were found for three of the dependent measures: rhyme awareness (1998, see Tables 3 & 4; 2001, see Tables 11 & 12), third grade understanding elements of literature (1999, see Tables 7 & 8), and third grade understanding a variety of resource materials (1999, see Tables 9 & 10; 1998 and 1999 combined, see Tables 21 & 22). Altogether, there were five (5/44= 11.4%) significant interactions with ten (50%) significant differences found when twenty simple main effects were analyzed. This is a small number of significant results. There were 44 possible interactions with 176 possible tests for simple main effects, thus only 5.7% (10/176 = 5.7%) of the tests were statistically significant. Only the significant differences in the tests of simple main effects are presented.

Females who attended Early Start fared better than males who attended Early Start on two dependent measures: third grade understanding elements of literature (1999, see Tables 7 & 8) and third grade understanding a variety of resource materials (1999, see Tables 9 & 10; 1998 and 1999 combined, see Tables 21 & 22).

Males who did not attend Early Start did as well as females who did not attend Early start on all measures except rhyme awareness in 1998 (see Tables 3 & 4) and 2001 (see Tables 11 & 12). In 1998, males who did not attend Early Start scored higher than females who did not attend Early Start. In 2001, this finding was reversed, with females scoring higher on rhyme awareness than males. Because these two results tend to cancel each other, there really are no differences between males and females who did not attend Early Start on any of the dependent measures.

Males who attended Early Start fared better than males who did not attend Early Start on only one measure—rhyme awareness in 2001 (see Tables 11 & 12). Two cohorts (1999, see Tables 9 & 10; 1998 and 1999 combined, see Tables 21 & 22) of males who did not attend Early Start scored higher on third grade understanding a variety of resource materials than those who attended Early Start. Males who attended Early Start did not differ from males who did not attend Early Start on any of the other dependent measures.

Females who attended Early Start scored higher on rhyme awareness (1998, see Tables 3 & 4) and third grade understanding elements of literature (1999, see Tables 7 & 8) than females who did not attend Early Start. Females who attended Early Start did not differ from females who did not attend early Start on any of the other dependent variables.

Table 3

Means and Standard Deviations for the Phonological Awareness Literacy Screening--Kindergarten--Rhyme Awareness, 1998 Cohorts by Group (Control v. Treatment) and Gender

Gender	Treatment/Control	<i>N</i>	<i>M</i>	<i>SD</i>
Female	Treatment	14	8.57	3.06
	Control	13	5.62	3.75
Male	Treatment	11	7.64	3.53
	Control	12	8.50	2.54

Table 4

Two-Way ANOVA for Gender by Treatment/Control-- Phonological Awareness Literacy Screening--Kindergarten--Rhyme Awareness, 1998 Cohorts

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Gender	11.78	1	11.78	1.12	.30
Control/Treatment	13.57	1	13.57	1.28	.26
Gender x Control/Treatment	45.23	1	45.23	4.28	.04
Error	486.05	46	10.57		
Total	3433.00	50			
Corrected Total	560.18	49			

Simple Main Effects					
Group	(<i>I</i>) Gender	(<i>J</i>) Gender	(<i>I-J</i>)	<i>p</i>	
Control	Female	Male	-2.89	.03	
Treatment	Female	Male	.94	.48	
	(<i>I</i>) Treatment/Control	(<i>J</i>) Treatment/Control			
Female	Treatment	Control	2.96	.02	
Male	Treatment	Control	-.86	.53	

Table 5

Means and Standard Deviations for the English Standards of Learning Grade Three Total English Test, 1999 Cohorts by Group (Control v. Treatment) and Gender

Gender	Treatment/Control	<i>N</i>	<i>M</i>	<i>SD</i>
Female	Treatment	12	471.17	27.58
	Control	17	418.88	36.86
Male	Treatment	13	412.84	47.92
	Control	8	390.25	52.14

Table 6

Two-Way ANOVA for Gender by Treatment/Control--English Standards of Learning Grade Three--Total English, 1999 Cohorts

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Gender	21973.97	1	21973.97	13.28	.00
Control/Treatment	16295.88	1	16285.88	9.86	.00
Gender x Control/Treatment	2561.58	1	2561.58	1.55	.22
Error	76100.62	46	1654.36		
Total	9157044.00	50			
Corrected Total	113890.08	49			

Table 7

Means and Standard Deviations for the English Standards of Learning Grade Three Subtest Understanding Elements of Literature, 1999 Cohorts by Group (Control v. Treatment) and Gender

Gender	Treatment/Control	<i>N</i>	<i>M</i>	<i>SD</i>
Female	Treatment	12	6.50	2.39
	Control	17	3.88	1.32
Male	Treatment	13	4.00	1.35
	Control	8	3.38	1.60

Table 8

Two-Way ANOVA for Gender by Treatment/Control--English Standards of Learning Grade Three Subtest Understanding Elements of Literature, 1999 Cohorts

Source	SS	df	MS	F	p
Gender	26.29	1	26.29	9.30	.00
Control/Treatment	30.56	1	30.56	10.76	.00
Gender x Control /Treatment	11.54	1	11.54	4.06	.05
Error	130.64	46	2.84		
Total	1193.00	50			
Corrected Total	198.42	49			

Simple Main Effects				
Group	(I) Gender	(J) Gender	(I-J)	p
Control	Female	Male	.51	.49
Treatment	Female	Male	2.50	.00
	(I) Treatment/Control	(J) Treatment/Control		
Female	Treatment	Control	2.62	.00
Male	Treatment	Control	.63	.41

Table 9

Means and Standard Deviations for the English Standards of Learning Grade Three Subtest Understanding a Variety of Resource Materials, 1999 Cohorts by Group (Control v. Treatment) and Gender

Gender	Treatment/Control	N	M	SD
Female	Treatment	12	8.33	.78
	Control	17	8.41	.80
Male	Treatment	13	6.23	1.59
	Control	8	8.13	2.03

Table 10

Two-Way ANOVA for Gender by Treatment/Control--English Standards of Learning Grade Three Subtest Understanding a Variety of Resource Materials, 1999 Cohorts

Source	SS	df	MS	F	p
Gender	16.59	1	16.9	10.05	.00
Control/Treatment	11.31	1	11.31	6.85	.01
Gender x Control /Treatment	9.58	1	9.58	5.80	.02
Error	75.96	46	1.65		
Total	3145.00	50			
Corrected Total	118.58	49			

Simple Main Effects				
Group	(I) Gender	(J) Gender	(I-J)	p
Control	Female	Male	.29	.61
Treatment	Female	Male	2.10	.00
	(I) Treatment/Control	(J) Treatment/ Control		
Female	Treatment	Control	-.08	.87
Male	Treatment	Control	-1.90	.00

Table 11

Means and Standard Deviations for the Phonological Awareness Literacy Screening--Kindergarten--Rhyme Awareness, 2001 Cohorts by Group (Control v. Treatment) and Gender

Gender	Treatment/Control	N	M	SD
Female	Treatment	12	9.00	1.76
	Control	13	8.33	2.53
Male	Treatment	13	9.54	.52
	Control	12	6.54	2.30

Table 12

Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy Screening--Kindergarten--Rhyme Awareness, 2001 Cohorts

Source	SS	df	MS	F	p
Gender	4.93	1	4.93	1.32	.26
Control/Treatment	41.95	1	41.95	11.28	.00
Gender x Control/Treatment	16.99	1	16.99	4.57	.04
Error	171.12	46	3.72		
Total	3715.00	50			
Corrected Total	237.22	49			

Simple Main Effects				
Group	(I) Gender	(J) Gender	(I-J)	p
Control	Female	Male	1.80	.03
Treatment	Female	Male	-.54	.49
	(I) Treatment/Control	(J) Treatment/Control		
Female	Treatment	Control	.67	.40
Male	Treatment	Control	3.00	.00

Table 13

Means and Standard Deviations for the Phonological Awareness Literacy Screening--Kindergarten--Concept of Word, 2001 Cohorts by Group (Control v. Treatment) and Gender

Gender	Treatment/Control	N	M	SD
Female	Treatment	12	7.33	2.67
	Control	13	6.92	3.26
Male	Treatment	13	6.70	2.90
	Control	12	5.31	2.53

Table 14

Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy Screening--Kindergarten--Concept of Word, 2001 Cohorts

Source	SS	df	MS	F	p
Gender	62.82	1	62.82	1.81	.18
Control/Treatment	156.26	1	156.26	4.51	.04
Gender x Control/Treatment	2.66	1	2.66	.08	.78
Error	1593.64	46	34.64		
Total	5014.00	50			
Corrected Total	1814.00	49			

Table 15

Means and Standard Deviations for the Phonological Awareness Literacy Screening--Kindergarten--Rhyme Awareness, 2002 Cohorts by Group (Control v. Treatment) and Gender

Gender	Treatment/Control	N	M	SD
Female	Treatment	11	6.55	2.58
	Control	14	7.57	3.20
Male	Treatment	14	8.64	1.45
	Control	11	9.64	1.21

Table 16

Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy Screening--Kindergarten--Rhyme Awareness, 2002 Cohorts

Source	SS	df	MS	F	p
Gender	53.36	1	53.36	10.15	.00
Control/Treatment	12.56	1	12.56	2.39	.13
Gender x Control/Treatment	3.25	1	3.25	.00	.98
Error	241.92	46	5.26		
Total	3583.00	50			
Corrected Total	302.50	49			

Table 17

Means and Standard Deviations for the English Standards of Learning Grade Three Total English Test, Combined 1998-1999 Cohorts by Group (Control v. Treatment) and Gender

Gender	Treatment/Control	<i>N</i>	<i>M</i>	<i>SD</i>
Female	Treatment	26	436.12	49.13
	Control	30	417.93	44.99
Male	Treatment	24	412.54	40.43
	Control	20	399.65	39.43

Table 18

Two-Way ANOVA for Gender by Treatment/Control--English Standards of Learning Grade Three Total English, Combined 1998-1999 Cohorts

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Gender	10718.19	1	10718.19	5.53	.02
Control/Treatment	5907.05	1	5907.05	3.05	.08
Gender x Control/Treatment	171.22	1	171.22	.08	.77
Error	186163.03	96	1939.19		
Total	17650301.00	100			
Corrected Total	202136.59	99			

Table 19

Means and Standard Deviations for the English Standards of Learning Grade Three Subtest Understanding Elements of Literature, Combined 1998-1999 Cohorts by Group (Control v. Treatment) and Gender

Gender	Treatment/Control	<i>N</i>	<i>M</i>	<i>SD</i>
Female	Treatment	26	5.00	2.28
	Control	30	4.00	1.08
Male	Treatment	24	4.04	1.23
	Control	20	3.50	1.47

Table 20

Two-Way ANOVA for Gender by Treatment/Control--English Standards of Learning Grade Three Subtest Understanding Elements of Literature, Combined 1998-1999 Cohorts

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Gender	13.01	1	13.01	5.21	.03
Control/Treatment	14.54	1	14.54	5.82	.02
Gender x Control/Treatment	1.29	1	1.29	.514	.48
Error	239.96	96	2.50		
Total	2007.00	100			
Corrected Total	268.11	99			

Table 21

Means and Standard Deviations for the English Standards of Learning Grade Three Subtest Understanding a Variety of Resource Materials, Combined 1998-1999 Cohorts by Group (Control v. Treatment) and Gender

Gender	Treatment/Control	<i>N</i>	<i>M</i>	<i>SD</i>
Female	Treatment	26	8.12	1.14
	Control	30	8.43	.82
Male	Treatment	24	7.00	1.56
	Control	20	8.35	1.50

Table 22

Two-Way ANOVA for Gender by Treatment/Control English Standards of Learning Grade Three Subtest Understanding a Variety of Resource Materials, Combined 1998-1999 Cohorts

Source	SS	df	MS	F	p
Gender	8.79	1	8.79	5.61	.02
Control/Treatment	17.02	1	17.02	10.85	.00
Gender x Control/Treatment	6.52	1	6.52	4.15	.04
Error	150.57	96	1.57		
Total	6567.00	100			
Corrected Total	6567.00	99			

Simple Main Effects				
Group	(I) Gender	(J) Gender	(I-J)	p
Control	Female	Male	.08	.82
Treatment	Female	Male	1.12	.00
	(I) Treatment/Control	(J) Treatment/Control		
Female	Treatment	Control	-.32	.35
Male	Treatment	Control	-1.35	.00

Summary

A summary of findings is in Table 23. There are few significant findings in this evaluation, and they are scattered across cohorts and dependent measures with relatively no systematic order. Attending Early Start may have an effect on the literacy learning of children; however, this study did not provide the evidence that one would like to see in support of those effects. There are some findings, however, that point to some effects of attendance, particularly in rhyme awareness, concept of word, grade three total English, and grade three understanding elements of literature. These are tentative findings, at best, because they were not found consistently across cohorts.

Table 23
Summary of Findings

Dependent variable	Cohort	Treatment		Gender		Treatment				Gender			
		Attended	Did not attend	Female	Male	Attended		Did not attend		Male		Female	
						Male	Female	Male	Female	Attended	Did not attend	Attended	Did not attend
PALS-K rhyme awareness	1998							X				X	
	2001	X							X	X			
	2002				X								
PALS-K concept of word	2001	X											
SOL 3 rd grade total English	1999	X		X									
	1998-1999			X									
SOL 3 rd grade understanding elements of literature	1999	X		X			X					X	
	1998-1999	X		X									
SOL 3 rd grade understanding a variety of resource materials	1999		X	X			X				X		
	1998-1999		X	X			X				X		

A curious finding occurred on third grade understanding of a variety of resource materials. Those not attending Early Start scored higher than those who attended Early Start. The reasons for this reversal from expectations should be explored by the program planners.

It was not surprising that females scored higher than males on those measures of literacy on which statistically significant differences were found. The odd finding was that males scored higher than females on rhyme awareness in the 2002 cohort. This could be due to error.

Finally, the interaction effects are the most interesting but may be artifacts of the analyses, thus they must be interpreted with caution. Females who attended Early Start did better than males who attended Early Start at the end of third grade on understanding elements of literature and understanding a variety of resource materials. This long-term effect of attending Early Start deserves further study.

All of the above findings must be considered in terms of the total number of possible findings. There were more analyses on which there were no differences between those who attended Early Start and those who did not attend Early Start. Thus, the reader is cautioned to interpret these findings carefully, and it is recommended that judgment about the program, staff, or schools be withheld until further data are collected, analyzed, and interpreted. The potential for making an error in judgment about the Early Start program, either good or bad, is great, given the limited amount of data and the few significant findings in this study.

Conclusions, Discussion, Implications for Practice, and Recommendations for Future Research

In this section, generalizations derived from the findings are presented. Possible reasons why findings support or fail to support the hypotheses are discussed followed by implications for practice in the field of education. Finally, suggestions for future research are offered.

Conclusions

Significant results on twenty-four analyses were found in this evaluation. Expected results were: (1) Early Start participants would score higher than non-participants on the Phonological Awareness Literacy Screening in kindergarten; (2) Early Start participants would score higher than non-participants on the English Standards of Learning test in third grade; (3) Early Start participants would have a higher instructional reading level than non-participants in fourth grade; and (4) Females would score higher than males on all dependent measures.

Some Early Start participants may benefit from the program in some areas. Students, regardless of gender, who attend the Early Start program may benefit on rhyme awareness, concept of word, English performance generally at the end of third grade, and understanding elements of literature at the end of third grade. Further, the results suggest that females, regardless of whether they attend Early Start, may perform better than males at the end of the third grade in English, understanding elements of literature, and understanding a variety of resource materials. Females who attend Early Start may be expected to do better than males who attend Early Start on understanding elements of literature and understanding a variety of resource materials at the end of third grade. From these data, it seems that females have an edge on males in the area of literacy, regardless of whether they attend Early Start or not. There was no effect of attending Early Start on fourth grade reading level.

Discussion

This study presents only meager evidence that students are benefiting from attending the Early Start program. The small number of significant findings and the inconsistency of the findings across cohorts should cause program planners to pause and reflect on the value of the program. Rash judgments about discontinuing the program should be withheld until further data

are collected and more extensive studies of the effects of the program are completed.

The results are not consistent with those in the literature (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Marcon, 2002; Weikart & Schweinhart, 1978). This raises several questions that must be answered. Was this study adequately constructed? Were the measures appropriate? Were the cohorts equivalent, or did those students in the cohorts that did not participate have better literacy skills at the time they were eligible to enroll than those who attended Early Start? Did those children who did not attend Early Start participate in some other preschool program? Was the program implemented as expected? Is the program powerful enough to produce the results expected? Is the program research based; does it have enough research support behind it? Are the teachers highly qualified? Are teachers well trained in the use of materials and in the curriculum? Did the parents participate as expected? There are probably many other questions that could be raised, and they should be raised as program managers and staff work through the results in this evaluation, which can serve as a catalyst for change in the program.

The No Child Left Behind Legislation states all students should be reading on or above grade level by third grade. School districts and their personnel are responsible for helping all students to reach this goal. Early start programs may be one way of providing this assistance. The issues are whether these programs have immediate effects on student learning, and if there are immediate effects, are these effects maintained throughout the elementary grades and into secondary schools and beyond. For many children, the regular fare of schooling isn't enough to help them achieve adequately in any of the core subjects. Reinforcement and special programs, including remediation, tutorial work, homework assistance, and parental education are necessary to maintain the momentum established in the early years.

The question may be asked why students who did not attend Early Start in the 1999 cohort and the 1998-1999 combined cohorts scored significantly higher on understanding a variety of resource materials than those who attended Early Start. This may be due to measurement or statistical error; however, there may be other reasons as well. Understanding and using resources may not be an area covered in the Early Start program, and students who didn't attend Early Start may have had teachers who emphasized the use of these resources in their classes. This is an area that should be carefully reviewed by the program managers.

While no rationale is known for why those students who qualified for the program did not participate; it is something to consider when looking at the results. Perhaps, parents of children who qualified but did not participate came to the conclusion their children already had sufficient literacy skills and did not need a year of preschool. Some may have sent their children to daycare or other settings that focused on school readiness skills. Not knowing this information raises questions about the results. Some interviews with these parents may produce information that would help recruitment for the program.

Another consideration is that each of the assessments used in this study is skill oriented. The Phonological Awareness Literacy Screening--Kindergarten focuses specifically on pre-reading skills that are essential for emergent readers (Invernizzi, Meir, Swank, & Juel, 2003), and the third grade English Standards of Learning test focuses on skills that are connected with specific grade level objectives. Perhaps this explains the inconsistent performance of the treatment group on the dependent measures.

How the teachers implemented the curriculum and the instructional methods they used may have contributed to the findings. Often the espoused curriculum is not implemented, and the context of the program may affect the choices teachers make. The most effective methods,

therefore, may not be implemented. Further, the philosophy and the every day routine in the Early Start program allow students to choose activities of interest. If a student does not choose an activity with an academic focus, he or she may be at a disadvantage when compared to students not in the program who are involved in more academic activities with a caregiver or home daycare provider.

Weikart & Schweinhart (1978) found that the academic difference between those who attended a preschool and those who did not appeared to be more prevalent in years beyond the elementary grades. Perhaps, we can expect to see more positive results as students get older. It is worth noting that more statistically significant results were found in this evaluation when students were in grade three as opposed to kindergarten.

Implications for Practice

This study has implications for educators across Virginia as we continue to confront the demands of higher accountability for students and schools. The cost of the program is roughly \$900,000 each year. Considering the cost of the program, benefits were expected. Because of the results of this evaluation, one could argue the need for a thorough review of the program. The questions posed earlier in this report could help guide such a review.

The measures of effectiveness should be reconsidered. For example, the Phonological Awareness Literacy Screening is a kindergarten test. It is research based and serves as a good diagnostic tool for kindergarten teachers. Recently, a pre-school Phonological Awareness Literacy Screening assessment was introduced. The Early Start program may benefit from such an assessment tool. This assessment may be more aligned with preschool and can provide teachers the necessary focus and students the necessary skills for later success. This is not to suggest that the Early Start program needs to become highly structured with only an academic

focus, but a focus on specific reading readiness skills may allow for more benefits.

The research on the High/Scope approach to learning suggests that students can be taught reading readiness skills through this approach (Hohmann & Weikart, 2002). The High/Scope approach has recently added a more specific literacy component. While the Early Start teachers have received High/Scope training, additional professional development on integrating reading readiness skills and the new literacy component may be in order.

The teachers and administrators of the Early Start program could be held more accountable. Professional development for administrators on what to look for and what to expect in the Early Start classroom for preparing students for kindergarten may be helpful. These teachers and administrators could focus more attention on evaluating what the students are learning in relation to expectations. Alignment of curriculum, instruction, and measures of performance could result in a more efficient learning environment for the children.

Recommendations for Future Research

As a result of this evaluation, several questions are raised that could lead to further research. The Instructional reading level data gathered on students in the fourth grade was limited to 50 students. It would be worth comparing the achievement of students who attended Early Start with those who did not attend Early Start in their later school years. The enduring effects of such programs are what is important, not the immediate acquisition of specific literacy skills. In the Perry Preschool Project, students were studied through adulthood. Measures of success other than academics were used. It would be interesting to compare Early Start participants' social and economic success in later life. A qualitative study that focuses on Early Start teacher quality may allow for program strengthening. An extension of this evaluation may include looking at teacher qualifications, experience, and teaching strategies as they relate to

student performance on literacy measures.

References

- Alexander, K., & Entwisle, D. (1988). *Achievement in the first two years of school: Patterns and processes*. Monographs of the Society for Research in Child Development, 53 (Serial No. 218).
- Belsky, J., & McKinnon, C. (1994). Transition to school: Developmental trajectories and school experiences. *Early Education and Development*, 5 (2), 106-119.
- Boyer, E. (1991). *Ready to learn: A mandate for the nation*. Princeton, N J: The Carnegie Foundation for the Advancement of Teaching.
- Campbell, F., Pungello, E., Miller-Johnson, S., Burchinal, M., & Ramey, C.T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology*, 37, 231-242.
- Keane, M., O'Brien, R., Connell, D., & Close, N. (1996). *A descriptive study of the Head Start health component*. (Summary Report, U.S. Department of Education, 1996).
Retrieved February 7, 2004, from
http://www.acf.hhs.gov/programs/opre/hs/descriptive_stdy/reports/descrip_stdy_vol1/hsh_ealth_vol1.pdf.
- Neuman, S., Roskos, K., Vukelich, C., & Clements, D. (2003). *The state of state pre-kindergarten standards*. Retrieved December, 2004, from
<http://www.ciera.org/library/archive/2003-01/index.htm>.
- Elkind, D. (2001). Much too early. *Education Next*. Retrieved January 7, 2004, from
<http://www.educationnext.org/20012/8elkind.html>.
- Evers, C. (2001). Educators push universal preschool. *CNNfyi.com*. Retrieved March 1, 2004, from

<http://www.cnn.com/2001/fyi/teachers.ednews/07/12/universal.preschool.ap/index.html>.

Graves, M. (2002). *High/Scope's preschool key experiences: Initiative and social relations*.

Ypsilanti, MI: High/Scope Press.

Gronlund, N. (1985). *Measurement and evaluation in teaching*. New York: Macmillan.

Head Start Bureau. (2004). *Compilation of the Head Start Act*. Retrieved June 11, 2005, from

<http://www2.acf.dhhs.gov/programs/hsb/budget/headstartact.htm#def>.

Head Start Bureau. (2005). *Head Start program fact sheet*. Retrieved June 11, 2005, from

<http://www.acf.hhs.gov/programs/hsb/research/2005.htm>.

Hohmann, M., & Weikart, D. (2002). *Young children in action: A manual for preschool educators*. Ypsilanti, MI: High/Scope Press.

Invernizzi, M., Robey, R., & Moon, T. (1999). *Phonological Awareness Literacy Screening 1997 - 1998: Description of sample, first year results, task analyses, and revisions*.

Charlottesville, VA: University of Virginia Printing and Copying Services.

Invernizzi, M., Meier, J., Swank, L., & Juel, C. (2003). *Phonological Awareness Literacy Screening--Kindergarten technical reference*. Charlottesville, VA: University of Virginia Printing and Copying Services.

Kagen, S., & Cohen, N. (1997). *Not by chance: Creating an early care and education*

system for America's children. Retrieved March 12, 2004, from

<http://www.nccic.org/pubs/bychance/execsumm.html>

Kessler, S. (1991). Alternative perspectives on early childhood education. *Early Childhood Research Quarterly*, 6, 183-197.

Lazar, I., Darlington, R., Murray, J., & Snipper, A. (1982). *Lasting effects of early education: A report from the Consortium for Longitudinal Studies*. Monographs of the Society for

- Research in Child Development, 47 (2-3, Serial No. 195).
- Love, J., Schochet, P., & Meckstroth A. (1996) *Are they in any real danger? What research does and doesn't tell us about child care quality and children's well being*. Princeton, NJ: Mathematica Policy Research, Inc.
- Lubeck, S. (1990). *Four- year-olds and public schooling*. Urbana, IL: Clearing House on Elementary and Early Childhood Education. (ERIC Document Reproduction Service No. ED325204)
- Marcon, R. (2002). Moving up the grades: Relationship between preschool model and later school success. *Early Childhood Research and Practice*, 4 (1). Retrieved February 15, 2004, from <http://ecrp.uiuc.edu/v4n1/marcon.html>.
- Mehrens, W., & Lehmann, I. (1987). *Using standardized tests in education*. New York: Longman.
- National Center for Educational Statistics (2003). *Pre-kindergarten in the U.S. public schools, 2000-2001*(NCES 2003-019). Washington, DC: Government Printing Office.
- National Center for Parent-Child Home Program. (2004) *Soaring through success through books and play* [Brochure]. Port Washington, NY: Author.
- National Center for Parent-Child Home Program. (n.d.). *The Parent-Child Home Program*. Retrieved June 11, 2005, from <http://www.parent-child.org/localsites/index.html>.
- National Institute for Early Education Research (2004). *The state of preschool, 2003 state preschool yearbook*. Newark, NJ: Author
- Neuman, S. (2003). From rhetoric to reality: The case of the high quality compensatory pre-

- kindergarten programs. *Phi Delta Kappan*, 85, 286–291.
- Popham, W. (1998). *Classroom assessment, what teachers need to know*. Boston: Allyn and Bacon.
- Renaissance Learning. (2000). *STAR Reading: Understanding reliability and validity*. Wisconsin Rapids, WI: Renaissance Learning, Inc. Retrieved January 14, 2003, from <http://research.renlearn.com/research/pdfs/133.pdf>
- Reynolds, A., & Temple, J. (1998). Extended early childhood intervention and school achievement: Age 13 findings from the Chicago longitudinal study. *Child Development*, 69, 231-246.
- Reynolds, A., Temple, J., Robertson, D., & Mann, E., (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest. *Journal of the American Medical Association*, 285, 2339-2346.
- Schweinhart, L., Weikart, D., & Larner, M. (1986). Consequences of three preschool curriculum models through age 15. *Early Childhood Research Quarterly*, 1 (1), 15-45.
- Schweinhart, L., Barnes, H., & Weikart, D. (1993). *Significant benefits the High/Scope Perry preschool study through age 27*. Monographs of the High/Scope Educational Research Foundation, Monograph 10. Ypsilanti, MI: High/Scope Press.
- Standards for Educational and Psychological Testing, 1999*. (2000). Washington DC: American Educational Research Association.
- Starr, A. (2002). Does universal preschool pay? *Business Week*. Retrieved March 27, 2004, from <http://www.keepmedia.com>.
- St. Pierre, R., Ricciuti, A., & Creps, C. (1999). *Synthesis of state and local even start evaluations*. Cambridge, MA: Abt Associates Inc.

- St. Pierre, R., Swartz, B., Games, S., Murray, S., Deck, D., & Nickel, P. (1995). *National evaluation of the Even Start Family Literacy Program: Final report*. Cambridge, MA: Abt Associates Inc.
- Tao, F., Gamse, B., & Tarr, H. (1998). *Second annual evaluation of the Even Start Family Literacy Program: Final report*. Cambridge, MA: Abt Associates Inc.
- U.S. Department of Education. (2002). *No Child Left Behind Act of 2001*. Retrieved November 20, 2002, from <http://www.ed.gov/legislation/ESEA02/beginning.html>.
- U.S. Department of Health and Human Services. (2001). *About Head Start*. Retrieved December 28, 2001, from <http://www.acf.hhs.gov/programs/hsb/about/index.htm>
- U.S. Department of Labor. (2004). Retrieved April 20, 2005, from <http://www.bls.gov/news.release/famee.t04.htm>
- Virginia Department of Education. (1999). *Standards of learning tests: Validity and reliability Information, spring 1998 administration*. Richmond, VA: Author.
- Virginia Department of Education. (2001). *Virginia preschool initiative guidelines*. Retrieved July 18, 2004, from www.pen.k12.va.us/VDOE/suptsmemos/2004/inf081b.pdf
- Warger, C. (1988). *A resource guide to public school early childhood programs*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Weikart, D., & Schweinhart, L. (1978). *Lasting differences: The High/Scope preschool curriculum comparison study through 23*. High/Scope Education Research Foundation, Monograph 12. Ypsilanti, MI: High/Scope Educational Research Foundation.
- Zigler, E., Piotrkowski, C., & Collins, R. (1994). Health services in Head Start. *Annual Review of Public Health, 15*, 511-534.

APPENDIXES

Appendix A

Expanded Description of the Early Start Program

The Early Start Preschool Program in Suffolk Public Schools is made possible through the Virginia Preschool Initiative (VPI). The Virginia Preschool Initiative Grant was developed to offer opportunities for four years olds considered at risk of educational failure. The goal of Suffolk's Early Start program is to reduce disparities among young children upon formal school entry and to reduce or eliminate those risk factors that lead to early academic failure. Essential components of the Early Start program are parental involvement, appropriate curriculum, and qualified staff.

Parental involvement is mandatory for parents to enroll their child into the program. Parents must attend nine parent workshops during the school year. These workshops include topics such as parenting skills, appropriate nutrition, discipline, helping your child at home, and getting ready for school. Parents are required to allow a family service worker into the home to assist as needed. Teachers make direct contact with parents through phone calls, written communication, and home visits. Parent-teacher conferences are held at least twice a year, and student progress is discussed.

The Early Start curriculum is based on the High/Scope educational approach with a focus on early literacy. In a High/Scope setting, children are encouraged to make choices about materials and activities throughout the day. As they make their choices and plans, they explore, ask and answer questions, solve problems, and interact with classmates and adults. In this kind of environment, children naturally engage in "key experiences"; i.e., activities that foster developmentally important skills and abilities (Graves, 2002).

The literacy focus included in the Early Start program is integrated through the High/Scope key experiences. Some of the literacy components are rhyme awareness, upper and lower case alphabet knowledge, beginning sound knowledge, verbal memory, print knowledge, concept of word, and name writing. Other literacy skills are talking with others about personal, meaningful experiences; describing objects, events and relations; listening to stories and poems; and making up stories and rhymes. Students are given the opportunity to express themselves through writing, drawing, and using inventive and conventional spelling. Reading activities in the Early Start program may include being read to the teacher, reading with the teacher, and being read to by the teacher. Student story dictations are an integral part of the reading activities.

Students are assessed a minimum of three times a year using the Child Observation Record (COR). This assessment is an observational tool for preschool children. It is designed to measure children's progress in early childhood settings. The Child Observation Record has 32 dimensions of learning in six broad categories. The categories are initiative, social relations, creative representation, movement and music, language and literacy, mathematics, and science. Teaching staff gather information by taking notes about students' interactions, created products, and responses to different situations. The COR assessment used in the program has language and literacy descriptors; consequently, Suffolk Public Schools encourages literacy instruction. The COR assessment includes anecdotal and work sample results. Students create as they play and visit interest areas in the classroom. The assessment is designed to show children's growth through their own creations.

The Early Start program is a full-day preschool program of five and one-half hours. A typical day may begin with the teacher meeting with the students on the carpeted area for "circle time." During circle time the teacher allows students to share information in an attempt to

encourage and build language skills. During this time the teacher may review the day of the week, the day's date, and the month of the year. Often, the calendar review is done with a simple song that students learn. This is followed by an additional song that may include low impact movement to make students aware of their body parts and personal space. After circle time, students participate in an activity termed plan, do, and review. This activity, the thrust of the High/Scope approach, involves students in deciding on which interest area they would like to attend. Interest areas are carefully planned by the teacher. Students use and return materials they need to explore and learn more about the world. The daily plan, do, and review routine provides an opportunity for students to share their experiences with the teacher and fellow classmates.

The Early Start program has a certified teacher and a teacher assistant for each class. Four of the ten Early Start teachers are endorsed in early childhood education. All teacher assistants are "highly qualified" as defined by the No Child Left Behind legislation. All of the Early Start staff are expected to participate in such staff development activities as parent involvement, review of High/Scope curriculum and methods, adult interaction, music and movement, and large and small group methods. The screening process for Early Start includes a speech evaluation, a parent interview, and a parent application, along with other developmental assessments which include the Denver II & Fluharty tests.

There were ten Early Start classes in Suffolk Public Schools at the start of this evaluation: three at Booker T. Washington Elementary, four at Florence Bowser Elementary, and three at Mount Zion Elementary. An additional Early Start class was added at the beginning of the 2004-2005 school year.

Appendix B

Expanded Description of the Phonological Awareness Literacy Screening--Kindergarten

Validity of the Phonological Awareness Literacy Screening--K

According to Popham, (1998) validity "refers to the defensibility of inferences made from test scores" (p.123). The PALS-K has been assessed for three types of validity by using pilot studies and by examining Virginia state-wide data over the last five years (Invernizzi, Meier, Swank, & Juel, 2003). The three types of validity assessed are content validity, criterion-related validity, and construct validity.

Gronlund (1985) defined content validity as the degree to which the sample of items and tasks provides a relevant and representative sample of the content addressed. The PALS-K assessment is aligned with what research literature identified as two specific levels of phonological awareness. These two levels are rhyme awareness and phonemic awareness, which are well represented in the rhyme awareness and beginning sound awareness sections of the PALS-K. All 26 letters of the alphabet are included in the assessment of letter alphabet recognition. This holds true for knowledge of letter sounds, with the exception of letters Q and X, which are difficult to pronounce in isolation (Invernizzi, Meier, Swank, & Juel, 2003). "To assess concept of word, contextual finger-point reading with specific line-by-line behavioral criteria was included" (Invernizzi, Meier, Swank, & Juel, 2003, p. 31).

Standards for Educational and Psychological Testing (1999) defined criterion-related validity as determining whether assessment scores are related to one or more outcome criteria. The two types of criterion-related validity that were considered with PALS-K are predictive and concurrent. Predictive validity refers to using an assessment to predict performance in the future. Concurrent validity refers to comparing the performance on the PALS—K to performance on

different criteria assessed at about the same time.

Predictive validity was assessed in two ways for the PALS-K. The first way was comparing PALS-K scores from one assessment in the fall with Stanford 9 scores obtained during the spring of the same year. The kindergarten and first grade version of the Stanford 9 was used because it measured concepts similar to the areas assessed by PALS-K (Invernizzi, Meier, Swank, & Juel, 2003). Seventy-four kindergarten students from a school district were screened with PALS-K during the fall of 1998 and did not receive any type of special intervention. In the spring of the school year the same kindergarten students were administered the Stanford 9. The fall PALS-K summed scores and subtask scores were significantly correlated with the spring Stanford 9 scaled scores ($p < .01$). The correlation between the fall PALS-K summed scores and spring Stanford 9 total reading scaled scores was .70 (Invernizzi, Meier, Swank, & Juel, 2003).

The second way predictive validity was examined was a process that predicted students' performance on future PALS-K and future PALS 1-3. Using the 2000 PALS-K score, regression equations were used with subtask scores of rhyme awareness, beginning sound awareness, alphabet recognition, letter sounds, and spelling as independent variables. This revealed "... adjusted R^2 values of .33 in predicting spring 2001 Summed Scores, .45 in predicting fall 2001 PALS 1-3 Entry Level Scores, and .30 in predicting spring 2002 PALS 1-3 Entry Level Scores" (Invernizzi, Meier, Swank, & Juel, 2003, p. 30).

Predictive validity was assessed in a pilot study for the PALS-K. A sample of 799 students was involved by using discriminant analysis to assess the relationship between PALS scores and third grade Reading Standards of Learning scores. The combination of the PALS scores over the kindergarten, first, second, and third grade years resulted in a discriminant

function that correctly classified 82% of the students according to their pass-fail status on the Standards of Learning reading test (Invernizzi, Meier, Swank, & Juel, 2003).

Another validity measure involving PALS-K is concurrent validity. As stated previously, concurrent validity involves comparing the results of performance on a different criterion assessed at about the same time. The different measure was, again, the Stanford 9 that was similar in concepts being assessed. As Invernizzi, Meier, Swank, & Juel (2003) presented:

The correlation between the end-of-year kindergarten PALS summed score and the Total Reading scaled score of the Stanford-9 was medium to high and significant ($r = .72$, $p < .001$). The correlations between the PALS summed score and the three Stanford-9 subtest scaled scores were medium to high and significant (Sounds and Letters, $r = .79$; Word Reading, $r = .74$; and Sentence Reading, $r = .58$). (p. 31)

The PALS-K was judged to have adequate concurrent validity for this evaluation.

The final validity measure is construct validity. To review the measure of construct validity, the PALS-K theoretical model must be discussed. The PALS-K theoretical model involves assessing the use of children's knowledge of sound and print and mixing the two as they attempt the pronunciation of letter sounds, the ability to use inventive spelling, the ability to match speech to print, and the ability to recognize words out of context. This theoretical model was tested in two ways: (1) Principal components analyses were conducted on PALS data, which was to verify the underlying factor structure, and (2) discriminant analyses on PALS data were conducted to determine the extent to which cohort membership could be accurately predicted from PALS subtask scores (Invernizzi, Robey, & Mobey, 1999). These analyses provided sufficient evidence to conclude that the PALS-K has adequate validity to assess the performance of students on literacy tasks.

Reliability of the Phonological Awareness Literacy Screening

Reliability “refers to the consistency with which a measure assesses whatever it is measuring” (Popham, 1998, p. 120). Reliability presents the error of measurement. Three facets of reliability was assessed for PALS-K: test-retest reliability, internal consistency, and inter-rater reliability. Test-retest reliability, which refers to administering the same test at two different times then calculating the correlation coefficient to reflect the relationship between the subjects’ performances (Popham, 1998), was assessed by using a sample of 473 students in 2000. Teachers administered the PALS-K twice to a random sample of students within a two week period. Pearson correlation coefficients were computed between subtest scores on the two administrations. The test-retest reliability coefficients ranged from .78 to .95 (Invernizzi, Meier, Swank, & Juel, 2003)

Internal consistency was assessed using Cronbach’s alpha (Mehrens & Lehmann, 1987). The alpha coefficients for the PALS-K tasks by gender, social economic status, and ethnicity based on statewide samples from screenings from 1998-2002 ranged from .79 to .89 (Invernizzi, Meier, Swank, & Juel, 2003). PALS-K was judged to have sufficient internal consistency to be used in this evaluation.

Inter-rater reliability was assessed using two different scorers during an administration of the PALS-K. One teacher administered the test, and two teachers scored the tests independently; yet at the same time. Inter-rater reliability coefficients for the subtests ranged from .96 to .99. PALS-K was judged to have sufficient inter-rater reliability to be used in this evaluation.

It is clear that several measures have been taken to determine PALS-K validity and reliability. The tool has been used for several years, and teachers are able to make diagnostic

decisions based on students' results. Having considered this, I am confident that the subtests of PALS-K are suitable for use in assessing student literacy performance in this evaluation of Early Start.

Scoring the Phonological Awareness Literacy Screening

The PALS-K has scores for each subtest: rhyme awareness, beginning sound awareness, alphabet recognition, letter sounds, spelling, and concept of word. These scores are added together as a summed score. The summed score is compared to grade level expectations for fall and for spring. If a student scores below the benchmark score, that student is in need of additional instructional assistance (Invernizzi, Meier, Swank, & Juel, 2003). For this evaluation I will use the raw score of each subtest and the summed score of the assessment.

Appendix C

Expanded Description of the Virginia Standards of Learning Assessments

Validity of the Virginia Standards of Learning Assessments

Each Standards of Learning assessment is constructed according to the Standards of Learning blueprint. The blueprint is a document that details the number and percentage of questions the assessment will have from each content area's reporting category. Questions for each test are selected from the testing contractor's item bank. Draft tests and test items are reviewed by Content Review Committees who are generally educators. These Virginia educators have experience and expertise in the content area and grade level of the test (Virginia Department of Education, 1999). The Division of Assessment & Reporting of the Virginia Department of Education has four statements that must hold true for a question to make it to field test status. These are the statements:

- The question measures the Standards of Learning it was designed to measure.
- The question appropriately measures content or skills that students in Virginia should be expected to learn by the spring of the designated grade level or near the end of the course.
- Difficulty of the question is appropriate.
- The question is free from content that stereotypes, offends, or unfairly penalizes students on the basis of personal characteristics such as gender, ethnicity, religion or socioeconomic status.

If the question meets all of the criteria, it is used as a field test question with Virginia students during a regular administration of the Standards of Learning test. Field test questions appear among the operational questions on each SOL test (Virginia Department of Education, 1999).

As far as statistical information regarding each question, the following statistics are used to decide if questions will appear in operational test forms: traditional item statistics, Rasch item statistics, and differential item functioning. Traditional item statistics assess frequency distribution, item mean, standard deviation, point biserial correlation, biserial correlation, item reliability index, and response distribution for each option for all respondents. Rasch item statistics measure item difficulty and item fit. Differential item functioning involves measuring item difficulty among demographic cohorts and is used to detect potential bias. After the statistical information review, the Content Review Committee applies the four statements for each item once again. If each statement holds true for the question, the question is considered for the operational test.

Reliability of the Virginia Standards of Learning Assessments

During the first administration of the Standards of Learning test there were many questions about its reliability. However, the Standards of Learning tests have been assessed for reliability using the measures indicated below (Virginia Department of Education, 1999):

In developing the Virginia Standards of Learning tests, the developers used Kuder-Richardson Formula #20, or the KR-20, as the statistical measure of test reliability for all Standards of Learning tests except English: Writing, where person separation reliability was used. The Kuder-Richardson is a traditional procedure designed to determine the degree to which the test questions consistently measure the same body of content and skills. KR-20 values range from 0-.99. A different writing reliability measure is used because the test contains both multiple choice items and a student writing response. Test developers aim for a test's KR-20 value to be as high as possible, while knowing that reaching .99 is impossible. KR-20 values ranged from a low of .80 on Grade 5 History

and Social Science to a high of .92 on Grade 8 Mathematics. (p. 11)

The procedures used in developing and measuring validity and reliability of the Standards of Learning tests are consistent with most high-stakes tests. Considering the thorough measures used to assess validity and reliability, the third grade Standards of Learning English test was judged to be a good criterion for measuring students' performance in this evaluation.

Scoring the Virginia Standards of Learning Assessments

Each Standards of Learning test is scored on a scale of 0 - 600. If a student scores 500 or higher, he or she is considered to have passed the test at the advanced proficiency level. Students who score 400 - 499 pass the test, and students who score 0 - 399 are considered to have failed the test (Virginia Department of Education, 1999). Each subtest used in this evaluation has a raw score, which is the number of correct responses. The English Standards of Learning scaled score and the raw scores for word analysis skills, understanding elements of literature, and understanding a variety of resource materials were used in this evaluation.

Appendix D

Expanded Description of the STAR Reading Assessment

Validity of the STAR Reading Assessment

To measure the validity of the STAR Reading Assessment, work was completed that had schools submit data on how their students performed on several other standardized test instruments along with their STAR Reading results. These data included more than 12,000 test scores from such tests as the California Achievement Tests (CAT), the Comprehensive Test of Basic Skills (CTBS), the Iowa Test of Basic Skills (ITBS), the Metropolitan Achievement Test (MAT), the Stanford Achievement Test-9th (SAT-9), and several state tests. The data were placed on a common scale, either Normal Curve Equivalent (NCE) or scaled scores. Next, the scores were used to compute Pearson product-moment correlation coefficients. The correlation coefficients were high (Renaissance Learning, 2000).

Reliability of the STAR Reading Assessment

To assess STAR's reliability, four methods were used: split-half method, the test-retest method, the alternative forms method, and the estimation of generic reliability method (Renaissance Learning, 2000).

Using the split-half method, one STAR assessment was divided into two tests. The score of each test was calculated and the two scores were correlated. The coefficient ranged from .89 - .92 for the grade levels assessed (Renaissance Learning, 2000).

In applying the test-retest method, equivalent versions of the test were administered to 2,095 students. Each test was administered a week apart. Correlation coefficients were wide spread for the subjects across grades; however, the reliability estimate over all 12 grades was .94 (Renaissance Learning, 2000).

In the alternative forms method, 4551 students were given the two forms of the STAR assessment and the correlation was .95. When reviewed by grade levels, it was noted that the coefficients ranged from .82 to .89 with an average of .85, thus showing reliable results (Renaissance Learning, 2000).

The fourth method used to measure reliability for STAR Reading was what was referred to as the estimation of generic reliability. This reliability test involves using the STAR Reading Item Response Theory (IRT) feature which is a part of the STAR Reading Test. The generic reliability estimates ranged from .89 to .92 (Renaissance Learning, 2000).

Scoring of the STAR Reading Assessment

The Instructional Reading Level (IRL) score will be used from the STAR Reading Assessment. Number codes 1.0 – 12.9 represent the IRLs. A student who performs at a 3.8 IRL is considered to be able to recognize words and understand instructional material at the level of third grade and eight months with assistance.

Appendix E

Institutional Review Board Exempt Approval

Virginia

Tech

VIRGINIA POLYTECHNIC INSTITUTE
AND STATE UNIVERSITY

Institutional Review Board

Dr. David M. Moore
RB (Human Subjects)Chair
Assistant Vice President for Research Compliance
CVM Phase II- Duck pond Dr., Blacksburg, V A 24061-0442
Office: 540/231-4991; FAX: 540/231-6033
email: moored@vt.edu

DATE: October 25, 2004

MEMORANDUM

TO: David J. Parks Educational Leadership & Policy St. 0302
Deran Whitney

FROM: David Moore

SUBJECT: **IRB Exempt Approval:** “The Evaluation of a At-Risk Four-Year Old
IRB # 04-529 Program”

I have reviewed your request to the IRE for exemption for the above
referenced project. I concur that the research falls within the exempt status.
Approval is granted effective as of October 25, 2004.

cc: File

Appendix F

Tables of Descriptive Statistics on Cohorts

Table F1

Means and Standard Deviations for the Phonological Awareness Literacy Screening-- Kindergarten, the English Standards of Learning Subtests at Grade Three, and the Instructional Reading Level at Grade Four, 1998 Cohorts by Group (Control v. Treatment) and Gender

Criterion	Control (C)			Treatment (T)			Diff.
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	(<i>T - C</i>)
Males							
PALS-K summed score	12	42.42	17.60	11	43.40	18.80	.98
PALS-K rhyme awareness	12	8.50	2.54	11	7.64	3.53	-.86
PALS-K beginning sounds	12	6.17	3.43	11	7.64	2.91	1.47
PALS-K alphabet knowledge	12	16.80	5.94	11	16.64	8.51	-.16
PALS-K letter identification	12	4.25	5.90	11	3.81	3.57	-.44
PALS-K concept of word	12	4.08	4.19	11	3.64	4.00	-.44
SOL Eng (3 rd)	12	405.92	29.11	11	412.18	31.70	6.26
SOL word analysis(3 rd)	12	7.60	1.31	11	7.36	1.43	-.24
SOL literature(3 rd)	12	3.60	1.44	11	4.09	1.14	.49
SOL resources(3 rd)	12	8.50	1.09	11	8.00	.94	-.50
Instructional reading level (4 th)	12	3.30	.32	11	3.62	.26	.32
Females							
PALS-K summed score	13	44.92	19.11	14	49.57	15.31	4.65
PALS-K rhyme awareness	13	5.61	3.80	14	8.60	3.06	2.99
PALS-K beginning sounds	13	5.61	3.36	14	7.07	2.70	1.46
PALS-K alphabet knowledge	13	18.92	8.27	14	20.86	4.70	1.94
PALS-K letter identification	13	6.50	5.00	14	4.00	3.10	-2.50
PALS-K concept of word	13	3.54	1.81	14	5.00	4.30	1.46
SOL Eng (3 rd)	13	416.70	55.50	14	406.07	44.03	-10.63
SOL word analysis(3 rd)	13	7.50	2.70	14	6.80	1.31	-0.70
SOL literature(3 rd)	13	4.15	.70	14	3.71	1.14	-0.44
SOL resources(3 rd)	13	8.46	.90	14	7.93	1.40	-0.53
Instructional reading level (4 th)	13	3.50	.40	14	3.52	.40	.02

Table F1 (continued)

Criterion	Control (C)			Treatment (T)			Diff. (T - C)
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	
Total Cohort							
PALS-K summed score	25	43.72	18.06	25	46.84	16.85	3.12
PALS-K rhyme awareness	25	7.00	3.50	25	8.16	3.24	1.16
PALS-K beginning sounds	25	5.90	3.33	25	7.32	2.73	1.42
PALS-K alphabet knowledge	25	17.90	7.18	25	19.00	6.84	1.10
PALS-K letter identification	25	5.40	5.44	25	3.92	3.23	-1.48
PALS-K concept of word	25	3.80	3.12	25	4.40	4.13	.60
SOL Eng (3 rd)	25	411.52	44.24	25	408.76	38.43	-2.76
SOL word analysis(3 rd)	25	7.52	2.10	25	7.04	1.37	-0.48
SOL literature(3 rd)	25	3.89	1.13	25	3.89	1.13	.00
SOL resources(3 rd)	25	8.48	.96	25	7.92	1.19	-.56
Instructional reading level (4 th)	25	3.40	.36	15	3.56	.34	.16

Table F2

*Means and Standard Deviations for the Phonological Awareness Literacy Screening--
Kindergarten, and the English Standards of Learning Subtests at Grade Three, 1999 Cohorts by
Group (Control v. Treatment) and Gender*

Criterion	Control (C)			Treatment (T)			Diff. (T - C)
	<u>N</u>	<u>M</u>	<u>SD</u>	<u>N</u>	<u>M</u>	<u>SD</u>	
Males							
PALS-K summed score	8	53.25	21.91	13	37.85	17.80	-15.40
SOL Eng (3 rd)	8	390.25	52.14	13	412.85	47.92	22.60
SOL word analysis(3 rd)	8	7.13	2.80	13	8.08	1.70	.95
SOL literature(3 rd)	8	3.38	1.60	13	4.00	1.40	.62
SOL resources(3 rd)	8	8.13	2.03	13	6.23	1.60	-1.90
Females							
PALS-K summed score	17	42.18	20.01	12	46.00	16.06	3.82
SOL Eng (3 rd)	17	418.90	36.86	12	471.16	26.58	52.26
SOL word analysis(3 rd)	17	7.94	1.68	12	7.33	2.39	-.61
SOL literature(3 rd)	17	3.90	1.32	12	6.50	2.40	2.60
SOL resources(3 rd)	17	8.41	.79	12	8.33	.79	-.08
Total Cohort							
PALS-K summed score	25	45.72	20.85	25	41.76	17.13	-3.96
SOL Eng (3 rd)	25	409.72	43.41	25	440.84	48.54	31.12
SOL word analysis(3 rd)	25	7.68	2.07	25	7.72	2.05	.04
SOL literature(3 rd)	25	3.72	1.40	25	5.20	2.27	1.48
SOL resources(3 rd)	25	8.32	1.28	25	7.24	1.64	-1.08

Note. Data were not available for the PALS-K subtests rhyme awareness, beginning sounds, alphabet knowledge, letter identification, and concept of word for the 1999 cohorts.

Table F3

*Means and Standard Deviations for the Phonological Awareness Literacy Screening--
Kindergarten, 2000 Cohorts by Group (Control v. Treatment) and Gender*

Criterion	Control (C)			Treatment (T)			Diff. (T - C)
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	
Males							
PALS-K summed score	13	52.54	12.64	14	60.21	20.96	7.67
PALS-K rhyme awareness	13	8.08	2.18	14	8.42	2.06	.34
PALS-K beginning sounds	13	6.08	3.60	14	7.00	3.23	.92
PALS-K alphabet knowledge	13	20.54	2.67	14	18.42	7.12	-2.12
PALS-K letter identification	13	8.00	3.83	14	9.57	4.99	1.57
PALS-K concept of word	13	7.46	6.11	14	10.85	3.28	3.39
Females							
PALS-K summed score	12	53.25	21.10	11	57.18	21.02	3.93
PALS-K rhyme awareness	12	8.58	1.78	11	7.18	3.16	-1.40
PALS-K beginning sounds	12	7.50	1.17	11	6.82	3.16	-.68
PALS-K alphabet knowledge	12	15.75	6.72	11	20.27	5.12	4.52
PALS-K letter identification	12	7.75	4.50	11	8.73	5.62	.98
PALS-K concept of word	12	9.25	7.51	11	9.82	5.95	.57
Total Cohort							
PALS-K summed score	25	52.88	16.85	25	58.88	20.60	6.00
PALS-K rhyme awareness	25	8.32	1.97	25	7.88	2.62	-.44
PALS-K beginning sounds	25	6.76	2.76	25	6.92	3.13	.16
PALS-K alphabet knowledge	25	18.24	5.49	25	19.24	6.27	1.00
PALS-K letter identification	25	7.88	4.07	25	9.20	5.18	1.32
PALS-K concept of word	25	8.32	6.74	25	10.40	4.56	2.08

Table F4

*Means and Standard Deviations for the Phonological Awareness Literacy Screening--
Kindergarten, 2001 Cohorts by Group (Control v. Treatment) and Gender*

Criterion	Control (C)			Treatment (T)			Diff. (T - C)
	<u>N</u>	<u>M</u>	<u>SD</u>	<u>N</u>	<u>M</u>	<u>SD</u>	
Males							
PALS-K summed score	13	48.15	16.23	13	48.31	14.35	.16
PALS-K rhyme awareness	13	6.54	2.30	13	9.54	.52	3.00
PALS-K beginning sounds	13	5.31	2.53	13	6.70	3.00	1.39
PALS-K alphabet knowledge	13	20.23	4.13	13	19.38	5.52	-.85
PALS-K letter identification	13	10.00	7.81	13	7.54	7.34	-2.46
PALS-K concept of word	13	5.38	4.52	13	8.46	2.22	3.08
Females							
PALS-K summed score	12	55.83	16.20	12	51.25	24.41	-4.58
PALS-K rhyme awareness	12	8.33	2.53	12	9.00	1.76	.67
PALS-K beginning sounds	12	6.92	3.26	12	7.33	2.67	.41
PALS-K alphabet knowledge	12	20.83	5.51	12	17.92	5.85	-2.91
PALS-K letter identification	12	11.42	5.02	12	11.00	9.69	-.42
PALS-K concept of word	12	7.17	4.97	12	11.17	9.62	4.00
Total Cohort							
PALS-K summed score	25	51.84	16.35	25	49.72	19.45	-2.12
PALS-K rhyme awareness	25	7.40	2.53	25	9.28	1.28	1.88
PALS-K beginning sounds	25	6.08	2.96	25	7.00	2.75	.92
PALS-K alphabet knowledge	25	20.52	4.74	25	18.68	5.61	-1.84
PALS-K letter identification	25	10.68	6.52	25	9.20	8.55	-1.48
PALS-K concept of word	25	6.24	4.73	25	9.76	6.84	3.52

Table F5

*Means and Standard Deviations for the Phonological Awareness Literacy Screening--
Kindergarten, 2002 Cohorts by Group (Control v. Treatment) and Gender*

Criterion	Control (C)			Treatment (T)			Diff. (T - C)
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	
Males							
PALS-K summed score	11	59.81	15.65	14	53.50	24.93	-6.31
PALS-K rhyme awareness	11	9.64	1.20	14	8.64	1.45	-1.00
PALS-K beginning sounds	11	8.18	2.36	14	7.00	2.51	-1.18
PALS-K alphabet knowledge	11	19.81	4.75	14	18.86	7.03	-.95
PALS-K letter identification	11	10.91	6.27	14	10.50	8.26	-.41
PALS-K concept of word	11	2.82	2.23	14	2.57	1.99	-.25
Females							
PALS-K summed score	14	48.35	17.81	11	56.00	9.65	7.65
PALS-K rhyme awareness	14	7.57	3.20	11	6.55	2.58	-1.02
PALS-K beginning sounds	14	7.57	2.00	11	8.09	2.26	.52
PALS-K alphabet knowledge	14	19.14	7.18	11	21.36	3.91	2.22
PALS-K letter identification	14	8.07	4.97	11	13.55	4.18	5.48
PALS-K concept of word	14	2.86	1.83	11	2.18	1.17	-.68
Total Cohort							
PALS-K summed score	25	53.40	17.54	25	54.60	19.42	1.20
PALS-K rhyme awareness	25	8.48	2.69	25	7.72	2.25	-.76
PALS-K beginning sounds	25	7.84	2.13	25	7.48	2.41	-.36
PALS-K alphabet knowledge	25	19.44	6.12	25	19.96	5.89	.52
PALS-K letter identification	25	9.32	5.64	25	11.84	6.83	2.52
PALS-K concept of word	25	2.84	1.97	25	2.40	1.66	-.44

Table F6

Means and Standard Deviations for All Treatment and Control Cohorts on the Phonological Awareness Literacy Screening—Kindergarten (1998-2002), English Standards of Learning Grade Three Total and Subtests (1998-1999), and Fourth Grade Instructional Reading Level (1998) by Gender

Criterion	Control (C)			Treatment (T)			Diff. (T - C)
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	
Males							
PALS-K summed score	57	50.91	17.01	65	49.06	20.77	-1.85
PALS-K rhyme awareness	49	8.12	2.35	52	8.60	2.14	.48
PALS-K beginning sounds	49	6.37	3.13	52	7.06	2.83	.69
PALS-K alphabet knowledge	49	19.37	4.60	52	18.40	6.91	-.97
PALS-K letter identification	49	8.27	6.44	52	8.10	6.72	-.17
PALS-K concept of word	49	5.04	4.74	52	6.50	4.50	1.46
SOL Eng (3 rd)	20	400.00	39.43	24	412.54	40.43	12.54
SOL word analysis(3 rd)	20	7.40	1.98	24	7.75	1.59	.35
SOL literature(3 rd)	20	3.50	1.47	24	4.04	1.23	.54
SOL resources(3 rd)	20	8.35	1.50	24	7.00	1.56	-1.35
Instructional reading level (4 th)	12	3.29	.32	11	3.62	.26	.33
Females							
PALS-K summed score	68	48.34	19.11	60	51.77	17.84	3.43
PALS-K rhyme awareness	51	7.50	3.09	48	7.90	2.80	.40
PALS-K beginning sounds	51	6.90	2.65	48	7.31	2.66	.41
PALS-K alphabet knowledge	51	18.69	7.05	48	20.10	4.97	1.41
PALS-K letter identification	51	8.37	5.10	48	9.02	7.00	.65
PALS-K concept of word	51	5.55	5.15	48	7.00	7.00	1.45
SOL Eng (3 rd)	30	417.93	44.99	26	436.12	49.13	18.19
SOL word analysis(3 rd)	30	7.73	2.15	26	7.04	1.87	-.69
SOL literature(3 rd)	30	4.00	1.08	26	5.00	2.28	1.00
SOL resources(3 rd)	30	8.43	.82	26	8.12	1.14	-.31
Instructional reading level (4 th)	13	3.48	.38	14	3.52	.40	.04

Table F6 (continued)

Criterion	Control (C)			Treatment (T)			Diff. (T - C)
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	
Total Combined Cohorts							
PALS-K summed score	125	49.51	18.16	125	50.36	19.39	.85
PALS-K rhyme awareness	100	7.80	2.76	100	8.26	2.49	.46
PALS-K beginning sounds	100	6.64	2.89	100	7.18	2.73	.54
PALS-K alphabet knowledge	100	19.02	5.95	100	19.22	6.09	.20
PALS-K letter identification	100	8.32	5.75	100	8.54	6.81	.22
PALS-K concept of word	100	5.30	4.93	100	6.74	5.75	1.44
SOL Eng (3 rd)	50	410.62	43.39	50	424.80	46.26	14.18
SOL word analysis(3 rd)	50	7.60	2.07	50	7.38	1.76	-.22
SOL literature(3 rd)	50	3.80	1.26	50	4.54	1.90	.74
SOL resources(3 rd)	50	8.40	1.12	50	7.58	1.46	-.82
Instructional reading level (4 th)	25	3.39	.35	25	3.56	.34	.17

Appendix G

Literature Review

Over the years, preschools have been considered safe and convenient and have offered a variety of curricula and activities. Some include specific academic standards while others use a developmentally appropriate model. Some states have begun to implement universal preschools for four year olds; others are moving in that direction. Many view the preschool as the only way to guarantee that children have an equal opportunity for education (Lubeck, 1990).

This review is a summary of the status of preschools, preschool models, and short and long term effects of preschool education. Twenty documents were reviewed. The documents were informational articles, empirical studies, and evaluation reports. Most of the studies included economically disadvantaged, minority students. The documents were selected with specific criteria in mind. They had to include evidence of subsequent success of preschoolers, and each had to have a specific philosophical approach to educating preschools. No date limits were set; the selected documents were published from the sixties to the present.

The topics of the documents varied, but the effectiveness of preschool programs was included in several. Some were follow-up studies that looked at children's ongoing performance (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Reynolds, Temple, Robertson, & Mann, 2001; Marcon, 2002; Weikart & Schweinhart, 1978). Others looked at overall program performance (Keane, O'Brien, Connell, & Close, 1996). Several compared different preschool models. A more extensive review of three "landmark" studies in preschool education is included. These are the High/Scope Perry Preschool Project, the Abecedarian Project, and the Chicago Child-Parent Centers Study. These were especially relevant to this evaluation and provided information that helped with interpretation of the data.

Status of Preschools

The National Center for Educational Statistics (2003) reported that 35 percent of U.S. schools offer some type of preschool program. These offerings vary by school size and location. Larger school districts are more likely to offer a preschool program. Forty states fund public preschools; however, they only enroll approximately 15% of the country's four-year-olds and three percent of three-year-olds. During a recent rating of ten educational categories, researchers found that Georgia, New Jersey, and Oklahoma provided the best preschool programs (National Institute for Early Education Research, 2004). Ten states--Alaska, Idaho, Indiana, Mississippi, Montana, New Hampshire, North Dakota, South Dakota, Utah, and Wyoming--offered no public preschools. Georgia and Oklahoma have attempted to offer preschool education to all four-year-olds (National Institute for Early Education Research).

The major difference between programs in the United States and programs in other countries is that in the United States almost all programs are targeted exclusively for the poorest children and for disabled children, while in other countries all children are included in regular preschool classes. Children with special needs receive additional services.

In the Commonwealth of Virginia, a few local school divisions have preschool programs for children younger than five years old. Some programs in Virginia are offered under the Virginia Preschool Initiative, which is a program designed to serve at risk students not served by any other program. School divisions must apply for this grant-type funding and be prepared to match the awarded amount. There are Title I preschool programs, Head Start, and vocational child care that may be offered by some schools.

The quality of preschool programs appears to make a difference. Susan B. Neuman (2003), former Assistant Secretary for Elementary and Secondary Education in the U.S.

Department of Education, suggested that four concepts should be considered when designing a high quality preschool program that could assist in closing achievement gaps. First, she suggested sufficient time be provided. Neuman shared how several at-risk preschool programs are often half-day or just a few hours. She believed that time is very important when considering the needs of disadvantaged children. Second, Neuman felt that precise targeting should be applied when considering preschool programs. This refers to selecting or designing programs for students who are most at-risk based on a multifaceted index of risk factors rather than one indicator. Next, Neuman believed a thoughtful focus should be in place. She believed the curriculum should include a primary focus on key foundational literacy skills that may include phonological awareness, narrative competence, and concepts of print. Along with this concept, Neuman felt much teacher training is needed to ensure appropriate implementation and maintenance of specific learning objectives. Finally, Neuman suggested that accountability for results must be an area of focus. She maintained that data must be used when attempting to hold one accountable. Neuman believed ensuring accountability is a first step in closing the achievement gap.

Advantages of public preschools suggested by advocates include the amount of money that is saved. They claim that students who attend quality preschool programs often need fewer services in later years. Schweinhart, Barnes, and Weikart (1993) found that investing \$1 in quality early education saved \$7 by reducing later grade retention and special education placement and increasing high school graduation rates. Quality preschool has been said to increase self-sufficiency and reduce delinquency and welfare use. Kagen and Cohen (1997) found that many of the positive effects of quality preschools contribute to children's increased cognitive abilities, positive classroom learning behaviors, long term school success, and long

term social and economic self-sufficiency.

Many believe the advantages of a preschool program are found in high quality programs. High quality programs are those that empower young children, empower parents, and empower teachers. These are programs that encourage children to initiate their own learning activities and focus on developmentally appropriate practices. Again, these programs involve parents, including home visits, and continuously develop parents' ability to see their children as able, active learners. Empowering teachers includes providing them in-service curriculum training and supportive curriculum supervision (Warger, 1988).

Preschool Models

The debate in preschool education commonly involves the type of program to be offered, what should be taught, and the advantages and disadvantages of preschool programs.

There are three basic types of programs: the traditional preschool, the academic preschool, and a combination of the traditional and academic preschool (Lubeck, 1990).

The traditional preschool was historically a part-day program for middle-class children. Today, the program is characterized by a teacher who may not directly teach students and an environment that is carefully planned but allows children to learn through exploration and discovery. Students are allowed to learn at their own pace, and developmentally appropriate practices are used. Teachers respond to children's self-initiated play in a loosely structured, socially supportive setting (Lubeck, 1990).

The academic preschool programs frequently involve efforts to help low-income children catch up to their middle-class peers. The academic preschool shares many characteristics of public schools, including high, specific academic standards; a specific schedule; teacher-directed instruction; clear goals; and specific teaching of skills. Many early childhood specialists feel that

some students may not be developmentally ready for such a rigorous curriculum (Elkind, 2001).

The combination preschools blend aspects of the traditional and academic preschools, including both structured and unstructured activities (Lubeck, 1990). The Suffolk Public Schools' Early Start Preschool Program is perhaps more of a combination model, yet mostly traditional.

Commonly known national preschool programs are Head Start, Even Start, and the Parent-Child Home Program. Each of these programs has key components in the theoretical framework of this evaluation: parental involvement, quality staff, and appropriate curriculum. The Head Start program is a child development program which serves children from birth to age five. The program's overall goal is to increase the school readiness of young children in low income families (Head Start Bureau, 2004). For a family to enroll their children in a Head Start program, they must meet the income eligibility requirements. A family's income must be below the federal poverty line or be a family receiving public assistance (Head Start Bureau, 2004).

Head Start was developed in 1965 as part of an attempt to address poverty and disparities among children and families. Services were delivered to over 900,000 children five years of age or younger in 2004 (Head Start Bureau, 2005). Head Start is currently administered by the Department of Health and Human Services.

One of the Head Start studies selected for review examined the health components of Head Start programs. It is an assessment of the comprehensive services that are often included and associated with preschool programs. During a recent conversation with a retired preschool director, the director said that the comprehensive components are emphasized when the preschool program begins; however, some of these components tend to be stressed less as the focus changes to academics and teacher-student interaction. I felt a need to review the research

that supports the value of the health component and reveal its findings.

I reviewed a description of Head Start health components by Keane, O'Brien, Connell, and Close (1996). The components examined were the number of years of employment for Head Start health coordinators, health behavior changes of Head Start families, immunization information, medical examinations, dental examinations, mental health screenings, and nutrition screenings of Head Start children.

A total of 1,189 parents at 81 centers across 40 Head Start programs were used. Information gathered as data sources were parents' and children's health records, meal observations at Head Start centers, and interviews with center directors, health coordinators, mental health coordinators, parent involvement coordinators, nutrition coordinators, and budget managers. A total of 219 staff interviews and 177 meal observations were completed. The findings (Keane, O'Brien, Connell, & Close, 1996) were as follows: 42.9% of the health coordinators had 1-4 years of experience, 16.7% had 5-9 years of experience, and 40.5% had more than ten years of experience. Ninety-five percent of the staff held a college degree or higher. Sixty-eight percent of the of parents reported health behavior changes since enrolling in the Head Start program, 68% reported positive changes. Eight-seven percent of the four year olds in the program were fully immunized. Ninety-eight percent of the children received medical examinations within the past year. Ninety-six percent of the children received dental examinations in the past year. Eighty-nine percent of the children received health screenings in the past year. Nutrition screenings for the past year amounted to 96%. Evidence of continued health services was obvious throughout this study. The overall goal of Head Start is to promote social competence among participating children (Zigler, Piotrkowski, & Collins, 1994). It is clear from this study that Head Start has assisted many children and families.

Even Start, another program that focuses on preschool education, is a family literacy program that was first authorized in 1989 as Part B of Chapter 1 of Title I of the Elementary and Secondary Education Act of 1965 (ESEA). This program embraces the whole family. It provides participating families with an integrated program of early childhood education, adult education and basic skills instruction, parenting education, and interactive literacy activities between parents and children. The Even Start program's main purposes are to build on existing community resources, promote children and adult academic achievement, incorporate research-based practices into instructional programs, and help children and adults meet the state's content standards. During the first year of the program, 76 grantees were awarded \$14.5 million dollars. In that year, Even Start served over 2,700 children and 2,900 adults. In the late 1990s, Even Start spent more than \$135 million dollars for 800 projects in all 50 states (St. Pierre, Ricciuti, & Creps, 1999).

Two recent national evaluations of Even Start showed that children learned such school readiness skills as colors, shapes, and sizes (as measured by a .90 standard deviation gain on the Preschool Inventory) significantly faster than would be expected on the basis of normal development. After one year of participation, Even Start children in five projects scored significantly higher on the Preschool Inventory than children in a randomly assigned control cohort. However, those children in the control cohort did catch up with their peers in kindergarten (St. Pierre, Swartz, Games, Murray, Deck, & Nickel, 1995). Tao, Gamse, & Tarr (1998) found that children who remained in Even Start for longer periods of time progressed at a faster-than-expected rate both on the Preschool Inventory and on the Preschool Language Scale. These gains suggest that students who participated in Even Start programs benefited from the preschool setting.

The Parent-Child Home Program (National Center for Parent-Child Home Program, 2004) is a home-based literacy and parenting program designed to serve the needs of families in poverty, with limited education, with language barriers, and with other issues preventing educational success. It is a voluntary, home-based two-year program designed to enhance the cognitive development of low income, at risk two to four year old children. It aims to promote appropriate parent and child interactions. Trained professionals model verbal interaction and educational play. The long term goal is preparation for school and prevention of later school problems. The three main features of the Parent-Child Home Program are in-home instruction, professional development, and parental involvement. The Parent-Child Home Program has 150 sites in the U.S. and other countries (National Center for Parent-Child Home Program, n.d.).

Published research demonstrates that the Parent-Child Home Program prepares children to succeed in school. Research findings have included academic progress, not only for the parents involved but for the children as well: Math and reading results on standardized tests of students who participated in the program were above national norms in elementary school. In addition, participants had higher graduation rates than non participants (National Center for Parent-Child Home Program, 2004).

Landmark Preschool Studies

Three landmark studies of preschool education are presented in detail in this section: The Abecedarian Project, the High/Scope Perry Preschool Project, and the Chicago Child-Parent Centers Study. All three have influenced the development of preschool educational programs in the United States.

In the Abecedarian Project (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001), children from low-income families were studied. Fifty-seven of the children were

assigned to a specific year round, full-day child-care program and participated in learning activities that were to enhance language skills. The Abecedarian Project included a specific curriculum. The treatment that those participants received was considered a high quality child-care setting that was provided from infancy to age five. Each participant had planned individualized educational activities. The activities focused on social, emotional, and cognitive development. Data were collected through standardized mental tests. Participants at the age of 18 months revealed an initial difference between those who were receiving high quality child care versus those who were not. By the end of the child care program, the difference became even more apparent. Those who attended the child care program had significantly higher scores on mental tests. The difference between the two cohorts remained at ages 12 and 15.

A young adult follow up assessment approximately 18 – 20 years later was completed (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001). Participants were located by the project's family coordinator, and a letter was sent to students to obtain follow up data. Sixty-eight were assessed within one month of their 21st birthday. Thirty-two within nine months of their 21st birthday, and four more were assessed a year after their 21st birthday (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001). It was revealed that 40% of the cohort receiving intervention was still in school compared to 20% of the cohort that did not receive intervention. Employment rates were higher for the intervention cohort. Sixty-five percent of the intervention cohort was employed, while 50% of the control cohort was employed.

Overall, it was concluded that those who received early educational intervention had significantly higher cognitive test scores from toddler through the age of 21. It was specifically found that reading achievement scores in primary school through the age of 21 were consistently higher for those individuals who received early intervention. In addition, Campbell, Pungello,

Miller-Johnson, Burchinal, and Ramey found that mathematics performance, although not as great as reading, was higher than those who did not receive early intervention. Those who received early intervention were more likely to be in post-secondary education programs. Participants, on average, were two years older (19) when they had their first child than those who did not receive early intervention. While some may question whether preschool experiences had anything to do with the age they had children, there was a pattern noted, hence it was presented as a correlation.

The Abecedarian Project revealed the power of a quality child care setting. The statistics are very clear that early childhood education in this setting made a difference. Unlike any of the other studies that will be discussed, the intervention began very early, during infancy. Focus was on quality staff members in quality child care settings. Trained individuals can provide much for young children to promote overall success. I believe these results are worth noting; however, when you consider the small sample size, I think we have to be careful with generalizations.

The landmark study, High/Scope Perry Preschool Project (Weikart & Schweinhart, 1978), revealed that adults born in poverty who participated in a high-quality, active learning preschool program at ages three and four have greater commitment to marriage, higher earnings and property wealth, and half as many criminal arrests. This study examined the lives of 123 African Americans born in poverty and at risk of failing school. The 123 children were randomly divided into two cohorts. One cohort received the High/Scope program in preschool while the other cohort did not attend preschool. Data have been collected five times on the effects of the program since 1970: at the end of preschool, at the age of ten, at the age of 15, at the age of 19, and at the age of 27. The Stanford-Binet Intelligence Test was used to assess the intellectual performance of the preschoolers and to ensure the program cohort matched the non-program

cohort. Each participant's IQ ranged from 70 – 85 (Weikart & Schweinhart, 1978). Participants were matched regarding their socioeconomic status as they were placed into the two cohorts. Data were collected by reviewing participants' school, social services, and arrest records. General Literacy tests, school achievement tests, and several intellectual tests were used to collect data.

The following findings were noted. Incidence of crime: 7% of those who participated in the preschool program had been arrested 5 times compared to 35% of those who did not attend preschool. Only 7% of those who were in the preschool cohort had ever been arrested for drug charges while 25% of those who did not participate in the preschool had been arrested for drug charges. Regarding higher earnings and property wealth, 29% of those who attended preschool were four times more likely to earn more money per month than those who did not attend preschool. Almost three times as many (36%) owned their own homes. Fifty-nine percent of those who attended preschool, compared to 80% of those who did not attend preschool, received welfare assistance at some time. Educational attainment revealed a striking difference. Seventy-one percent of those who were in preschool completed high school or received a General Education Development certificate, compared to 54% of those who did not attend the program. At the age of 14, those who attended preschool maintained significantly higher achievement scores, specifically in literacy. Lastly, the findings revealed that 40% of the women who attended the program were married compared to 8% of those who did not attend preschool. Eighty-three percent of those who did not attend were single mothers, whereas only 53% of those who attended were single mothers (Weikart & Schweinhart, 1978).

Another landmark study, the Chicago Child-Parent Centers Study (Reynolds & Temple, 1998), was a follow up study. Its main purpose was to determine the long-term effect of a federal

center-based preschool and school-based intervention program for low income students. Participants were 1,539 children. Nine hundred eighty-nine of the children attended 20 of the Chicago Child-Parent Centers in preschool and kindergarten in high poverty neighborhoods. The remaining 550 participated in an all day kindergarten program. The Chicago-Parent Centers included comprehensive educational and family support services from the ages of 3-9 years in neighborhood schools. The curriculum focused on developing children's reading, math, and communication skills. The study was a quasi-experimental design, and data were collected by family surveys and from educational and justice system records. Three-fourths of the surveys were returned. Reading and mathematics test scores were used as outcome variables. Word analysis, reading comprehension, and math scores from the Iowa Test of Basic Skills were used to assess the effects of program participation. Grade retention was used to determine program effectiveness. Descriptive data on families were collected by surveying and interviewing parents. Kindergarten growth was determined by comparing pre and post test results (Reynolds & Temple, 1998).

Three comparisons were made. One was a comparison of those who participated in the preschool program with those who did not. Another was a comparison of those who participated in the school-age intervention at the center with those who did not participate. The third was a comparison of those with extended program participation (4-6 years) with those with limited program participation (1-4 years).

Researchers looked at the following outcomes: school achievement, special education placement, grade retention, juvenile arrest by age 18, child maltreatment, and high school completion by age 21. Being a quasi-experimental design, researchers compared the cohort with a random sample of eligible children who did not participate in the program but were in an

alternative program (Reynolds, Temple, Robertson, & Mann, 2001). Preschool participants, compared to the comparison group, had a 29% higher rate of high school completion, a 33% lower rate of juvenile arrest, a 42% reduction in arrest for violent offense, a 41% reduction in special education placement, a 40% reduction in the rate of grade retention, and a 51% reduction in child maltreatment (Reynolds, et al., 2001). School-age participation and extended program participation for four to six years were associated with 30% to 40% lower rates of grade retention and special education placement (Reynolds et al., 2001). It was clear that participation in childhood intervention made a difference in children's educational and social outcomes. The findings of this study indicates that public investments in early educational programs can contribute positively to children's later success.

The three studies I reviewed--High/Scope Perry Preschool Project, the Carolina Abecedarian Project, and the Chicago Child-Parent Centers Study--have several similarities. They included low income participants, mostly African American children, and were longitudinal studies that looked at academic performance as well as some social issues. However, there were some distinct differences as well. The Abecedarian Study and the High/Scope studies were experimental designs and used random assignment of children. The Chicago study used a quasi-experimental design based on identification of existing groups of children. The Abecedarian Study did not present rates of crime prevention as did the other two studies. Although the results in all three studies were favorable, one has to consider if the particular preschool program was the reason for the success. I bring attention to this simply because those who were enrolled in these programs were probably enrolled by parents or guardians who supported education more than those parents who did not have their children in the program. Parents who support education may present different values and maintain different

expectations for their children. This could have been a factor contributing to the children's success.

Appendix H
Analyses With Non-Significant Statistical Results

Table H1

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Summed Score, 1998 Cohorts

Source	SS	df	MS	F	p
Gender	235.40	1	235.40	.75	.39
Control/Treatment	97.05	1	97.05	.31	.58
Gender x Control /Treatment	42.46	1	42.46	.14	.71
Error	14361.81	46	312.21		
Total	117274.00	50			
Corrected Total	14760.08	49			

Table H2

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Beginning Sounds, 1998 Cohorts

Source	SS	df	MS	F	p
Gender	3.86	1	3.86	.40	.53
Control/Treatment	26.53	1	26.53	2.76	.10
Gender x Control/Treatment	5.78	1	5.78	.00	.99
Error	442.22	46	9.61		
Total	2650.00	50			
Corrected Total	472.00	49			

Table H3

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Alphabet Knowledge, 1998 Cohorts

Source	SS	df	MS	F	p
Gender	126.73	1	126.73	2.62	.11
Control/Treatment	10.27	1	10.27	.21	.64
Gender x Control/Treatment	13.00	1	13.00	.27	.61
Error	2221.43	46	48.29		
Total	19378.00	50			
Corrected Total	2376.32	49			

Table H4

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Letter Sound Identification, 1998 Cohorts

Source	SS	df	MS	F	p
Gender	17.76	1	17.76	.88	.35
Control/Treatment	25.95	1	25.95	1.29	.26
Gender x Control/Treatment	12.77	1	12.77	.63	.43
Error	929.12	46	20.20		
Total	2073.00	50			
Corrected Total	987.22	49			

Table H5

Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy Screening Concept of Word, 1998 Cohorts

Source	SS	df	MS	F	p
Gender	2.08	1	2.08	.15	.70
Control/Treatment	3.19	1	3.19	.23	.63
Gender x Control/Treatment	11.29	1	11.29	.82	.37
Error	630.69	46	13.71		
Total	1489.00	50			
Corrected Total	648.50	49			

Table H6

Two-Way ANOVA for Gender by Treatment/Control--Standards of Learning Grade Three Total English, 1998 Cohorts

Source	SS	df	MS	F	p
Gender	67.47	1	67.47	.03	.85
Control/Treatment	58.81	1	58.81	.03	.86
Gender x Control/Treatment	883.89	1	883.89	.50	.48
Error	81466.25	46	1771.01		
Total	8493257.00	50			
Corrected Total	82516.02	49			

Table H7

*Two-Way ANOVA for Gender by Treatment/Control--Standards of Learning English Grade
Three Word Analysis, 1998 Cohorts*

Source	SS	df	MS	F	p
Gender	1.52	1	1.52	.47	.50
Control/Treatment	2.49	1	2.49	.77	.39
Gender x Control/Treatment	.65	1	.65	.20	.66
Error	149.05	46	3.24		
Total	2804.00	50			
Corrected Total	154.08	49			

Table H8

*Two-Way ANOVA for Gender by Treatment/Control--Standards of Learning English Grade
Three Understanding Elements of Literature, 1998 Cohorts*

Source	SS	df	MS	F	p
Gender	.12	1	.12	.09	.76
Control/Treatment	1.43	1	1.43	.01	.92
Gender x Control/Treatment	2.78	1	2.78	2.20	.15
Error	58.37	46	1.27		
Total	814.00	50			
Corrected Total	61.28	49			

Table H9

Two-Way ANOVA for Gender by Treatment/Control--Standards of Learning English Grade Three Understanding a Variety of Resource Materials, 1998 Cohorts

Source	SS	df	MS	F	p
Gender	1.12	1	1.12	.00	.98
Control/Treatment	3.92	1	3.92	3.21	.08
Gender x Control/Treatment	1.04	1	1.04	.00	.93
Error	56.07	46	1.22		
Total	3422.00	50			
Corrected Total	60.00	49			

Table H10

Two-Way ANOVA for Gender by Treatment/Control--STAR Grade Four Instructional Reading Level, 1998 Cohorts

Source	SS	df	MS	F	p
Gender	2.87	1	2.87	.24	.63
Control/Treatment	.41	1	.41	3.38	.07
Gender x Control/Treatment	.26	1	.26	2.15	.14
Error	5.57	46	.12		
Total	6.11	50			
Corrected Total	6.23	49			

Table H11

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Summed Score, 1999 Cohorts

Source	SS	df	MS	F	p
Gender	24.78	1	24.78	.07	.79
Control/Treatment	389.75	1	389.75	1.09	.30
Gender x Control/Treatment	1074.44	1	1074.44	3.02	.09
Error	16393.66	46	356.38		
Total	113331.00	50			
Corrected Total	17671.62	49			

Table H12

Two-Way ANOVA for Gender by Treatment/Control--Standards of Learning English Grade Three Word Analysis, 1999 Cohorts

Source	SS	df	MS	F	p
Gender	1.53	1	1.53	.00	.95
Control/Treatment	.34	1	.34	.08	.78
Gender x Control/Treatment	7.07	1	7.07	1.65	.21
Error	197.41	46	4.30		
Total	3169.00	50			
Corrected Total	204.50	49			

Table H13

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Summed Score, 2000 Cohorts

Source	SS	df	MS	F	p
Gender	16.70	1	16.70	.04	.83
Control/Treatment	417.69	1	417.67	1.13	.29
Gender x Control/Treatment	43.45	1	43.45	.12	.73
Error	16943.47	46	368.34		
Total	173582.00	50			
Corrected Total	17453.28	49			

Table H14

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Rhyme Awareness, 2000 Cohorts

Source	SS	df	MS	F	p
Gender	1.70	1	1.70	.32	.58
Control/Treatment	3.42	1	3.42	.64	.43
Gender x Control/Treatment	9.53	1	9.53	1.80	.19
Error	246.91	46	5.37		
Total	3541.00	50			
Corrected Total	260.50	49			

Table H15

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Beginning Sounds, 2000 Cohorts

Source	SS	df	MS	F	p
Gender	4.78	1	4.78	.54	.47
Control/Treatment	.18	1	.18	.02	.89
Gender x Control/Treatment	7.98	1	7.98	.91	.35
Error	405.56	46	8.82		
Total	2758.00	50			
Corrected Total	418.72	49			

Table H16

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Alphabet Knowledge, 2000 Cohorts

Source	SS	df	MS	F	p
Gender	126.73	1	126.73	2.63	.11
Control/Treatment	10.27	1	10.27	.21	.65
Gender x Control /Treatment	12.99	1	12.99	.27	.61
Error	2221.43	46	48.30		
Total	19378.00	50			
Corrected Total	2376.32	49			

Table H17

*Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy
Screening Letter Sound Identification, 2000 Cohorts*

Source	SS	df	MS	F	p
Gender	3.71	1	3.71	.16	.69
Control/Treatment	20.14	1	20.14	.89	.35
Gender x Control /Treatment	1.09	1	1.09	.05	.83
Error	1037.86	46	22.56		
Total	4711.00	50			
Corrected Total	1064.42	49			

Table H18

*Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy
Screening Concept of Word, 2000 Cohorts*

Source	SS	df	MS	F	p
Gender	1.74	1	1.74	.05	.82
Control/Treatment	48.70	1	48.70	1.43	.24
Gender x Control/Treatment	24.80	1	24.80	.73	.40
Error	1562.83	46	34.00		
Total	6024.00	50			
Corrected Total	1643.52	49			

Table H19

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Summed Score, 2001 Cohorts

Source	SS	df	MS	F	p
Gender	352.00	1	352.00	1.07	.30
Control/Treatment	61.22	1	61.22	.19	.67
Gender x Control/Treatment	70.02	1	70.02	.21	.65
Error	15078.38	46	327.79		
Total	144487.00	50			
Corrected Total	15556.58	49			

Table H20

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Beginning Sounds, 2001 Cohorts

Source	SS	df	MS	F	p
Gender	15.80	1	15.80	1.95	.17
Control/Treatment	10.12	1	10.12	1.25	.27
Gender x Control/Treatment	2.92	1	2.92	.36	.55
Error	373.12	46	8.11		
Total	2541.00	50			
Corrected Total	402.42	49			

Table H21

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Alphabet Knowledge, 2001 Cohorts

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Gender	2.34	1	2.34	.08	.77
Control/Treatment	44.18	1	44.18	1.59	.21
Gender x Control /Treatment	13.38	1	13.38	.48	.49
Error	1279.97	46	27.83		
Total	20546.00	50			
Corrected Total	1338.00	49			

Table H22

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Letter Sound Identification, 2001 Cohorts

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Gender	74.25	1	74.25	1.27	.26
Control/Treatment	25.85	1	25.85	.44	.51
Gender x Control/Treatment	13.05	1	13.05	.22	.64
Error	2688.15	46	58.44		
Total	7743.00	50			
Corrected Total	2802.82	49			

Table H23

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Summed Score, 2002 Cohorts

Source	SS	df	MS	F	p
Gender	247.33	1	247.33	.73	.40
Control/Treatment	5.41	1	5.41	.02	.90
Gender x Control /Treatment	600.33	1	600.33	1.77	.19
Error	15588.35	46	338.88		
Total	162254.00	50			
Corrected Total	16454.00	49			

Table H 24

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Beginning Sounds, 2002 Cohorts

Source	SS	df	MS	F	p
Gender	.71	1	.71	.14	.71
Control/Treatment	1.35	1	1.35	.26	.61
Gender x Control/Treatment	8.91	1	8.91	1.71	.20
Error	239.97	46	5.22		
Total	3185.00	50			
Corrected Total	251.22	49			

Table H25

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Alphabet Knowledge, 2002 Cohorts

Source	SS	df	MS	F	p
Gender	10.33	1	10.33	.28	.60
Control/Treatment	4.90	1	4.90	.13	.72
Gender x Control /Treatment	31.18	1	3.18	.85	.36
Error	1689.61	46	36.73		
Total	21139.00	50			
Corrected Total	1734.50	49			

Table H26

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Letter Sound Identification, 2002 Cohorts

Source	SS	df	MS	F	p
Gender	.13	1	.13	.00	.95
Control/Treatment	79.01	1	79.01	2.05	.16
Gender x Control/Treatment	106.60	1	106.60	2.76	.10
Error	1776.07	46	38.61		
Total	7559.00	50			
Corrected Total	1962.18	49			

Table H27

Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy Screening Concept of Word, 2002 Cohorts

Source	SS	df	MS	F	p
Gender	.38	1	.38	.11	.74
Control/Treatment	2.62	1	2.62	.76	.39
Gender x Control /Treatment	.57	1	.57	.16	.69
Error	158.42	46	3.44		
Total	505.00	50			
Corrected Total	161.78	49			

Table H28

Two-Way ANOVA for Gender by Treatment/Control–Phonological Awareness Literacy Screening Summed Score, 1998-2002 Cohorts

Source	SS	df	MS	F	p
Gender	.27	1	.27	.00	1.00
Control/Treatment	38.71	1	38.71	.10	.74
Gender x Control /Treatment	4.33	1	433.43	1.23	.27
Error	87048.27	246	353.86		
Total	710928.00	250			
Corrected Total	87526.98	249			

Table H29

Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy Screening Kindergarten Rhyme Awareness, 1998-2002 Cohorts

Source	SS	df	MS	F	p
Gender	22.18	1	22.18	3.24	.07
Control/Treatment	9.66	1	9.66	1.41	.24
Gender x Control/Treatment	5.79	1	5.79	.01	.99
Error	1343.01	196	6.85		
Total	14272.00	200			
Corrected Total	1375.82	199			

Table H30

Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy Screening Beginning Sounds, 1998-2002 Cohorts

Source	SS	df	MS	F	p
Gender	7.78	1	7.78	.98	.32
Control/Treatment	15.13	1	15.13	1.90	.17
Gender x Control/Treatment	.98	1	.99	.12	.73
Error	1561.03	196	7.96		
Total	11134.00	200			
Corrected Total	1584.38	199			

Table H31

Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy Screening Alphabet Knowledge, 1998-2002 Cohorts

Source	SS	df	MS	F	p
Gender	12.97	1	12.97	.36	.55
Control/Treatment	2.578	1	2.58	.07	.79
Gender x Control/Treatment	70.82	1	70.82	1.96	.16
Error	7101.37	196	36.23		
Total	80302.00	200			
Corrected Total	7187.12	199			

Table H32

Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy Screening Letter Sound Identification, 1998-2002 Cohorts

Source	SS	df	MS	F	p
Gender	13.30	1	13.30	.33	.57
Control / Treatment	2.87	1	2.87	.07	.79
Gender x Control/Treatment	8.34	1	8.34	.21	.65
Error	7848.97	196	40.05		
Total	22086.00	200			
Corrected Total	7873.02	199			

Table H33

Two-Way ANOVA for Gender by Treatment/Control--Phonological Awareness Literacy Screening Concept of Word, 1998-2002 Cohorts

Source	SS	df	MS	F	p
Gender	12.69	1	12.69	.44	.51
Control / Treatment	105.76	1	105.76	3.66	.06
Gender x Control/Treatment	8.40	1	8.40	.00	1.00
Error	5667.55	196	28.92		
Total	13032.00	200			
Corrected Total	5783.92	199			

Table H34

Two-Way ANOVA Gender by Treatment/Control--Grade Three Standards of Learning Word Analysis, 1998-1999 Cohorts

Source	SS	df	MS	F	p
Gender	.88	1	.88	.24	.63
Control / Treatment	.73	1	.72	.19	.66
Gender x Control/Treatment	6.68	1	6.68	1.81	.18
Error	354.13	96	3.68		
Total	5973.00	100			
Corrected Total	362.99	99			