

1-1-2007

Exploration into the Head Start Fade Phenomenon

Deborah Tenjeras Clarke

Follow this and additional works at: <http://commons.emich.edu/theses>

Recommended Citation

Clarke, Deborah Tenjeras, "Exploration into the Head Start Fade Phenomenon" (2007). *Master's Theses and Doctoral Dissertations*. Paper 137.

This Open Access Dissertation is brought to you for free and open access by the Master's Theses, and Doctoral Dissertations, and Graduate Capstone Projects at DigitalCommons@EMU. It has been accepted for inclusion in Master's Theses and Doctoral Dissertations by an authorized administrator of DigitalCommons@EMU. For more information, please contact lib-ir@emich.edu.

Exploration into the Head Start Fade Phenomenon

by

Deborah Tenjeras Clarke

Dissertation

Submitted to the Department of Leadership and Counseling

Eastern Michigan University

in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

Dissertation Committee:

Charles M. Achilles, EdD

Helen Ditzhazy, PhD

Ellen Hoffman, EdD

Jaclynn Tracy, PhD

July 26, 2007

Ypsilanti, Michigan

DEDICATION

This work is dedicated with love and gratitude to my mother, Patricia Tenjeras, who taught me the love of learning. Your dream for your children to receive a Bachelor's degree has been exceeded, and we know you are with us in spirit. I also want to dedicate this to my parents Doc and Judy Tenjeras, and to my family, including Margy and Nick, for always supporting me with love that has given me the courage to reach my goals. I thank my friends, Dr. Marci Erby and Dr. Dawn Pickard, and mentors, including my chair, Dr. Achilles, and committee members, Dr. Ditzhazy, Dr. Hoffman, and Dr. Tracy, for their unfailing support. I thank those mentors, Dr. T. C. Wallace, Dr. James Hawkins, and Judith White, who taught me about leadership, relationships, quality, and equity. To everyone: Each one of you has coached me and given me the gifts of knowledge, insight, and trust. I could not have taken this journey without you and your belief that together we could achieve the goal that would help every child walk a path to achieving his or her dreams.

ACKNOWLEDGEMENTS

A special thanks is offered to Delores Love who graciously assisted with the data warehouse established for the study and patiently worked with a long-distance partnership. I wish to acknowledge editor, Nancy Hoose, and statistician, June A. Cline, who assisted me early in the process. Both Jeanne Ballew and her daughter, Jenneva, were champions who offered the editorial strengths to help me complete the study in the final hours and proved that there are guardian angels.

I acknowledge school administrators and teachers who offered their thoughts, inspiration, and assistance in an effort to move this work forward. Finally, Michael Yokum and Ernie Bauer from Oakland Schools and Harvey Czerwinski from Macomb County Intermediate School District, I thank you with profound gratitude and admiration for being part of this journey.

ABSTRACT

The Head Start fade effect, documented since the 1970s, finds that students who make gains in I.Q. and social skills in the Head Start program later see those positive effects diminish in the early years of schooling and disappear altogether by the end of third grade. The hypothesis proposed in this study was as follows: Group I Head Start students who experience full-day kindergarten every day in small classes, and continue in small classes through grades one, two, and possibly three, will not demonstrate the Head Start fade effect, or will have less fade, than Group II Head Start students who have half-day kindergarten on alternating days in small classes and who have large classes in grades one, two, and possibly three. This objective of the study was to test a set of previously established theories that when applied would mitigate or eliminate the fade-out effect experienced by Head Start participants in most programs.

To accomplish this objective, this study used a used a non-experimental, longitudinal, retrospective explanatory design. The method involved tracking the progress of two groups of children, a treatment group and a control group, from three of school districts in Michigan for a period of seven years, starting with the Head Start program. It also used two formats for yearly assessment, norm-referenced tests (NRT) and criterion-referenced tests (CRT). The results of this study showed no statistically significant fade effect for the participants. It is the conclusion of this researcher that applying the theories used in this study can mitigate the Head Start fade effect for young children and strengthen their opportunities for improved achievement and long-term success.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
LIST OF TABLES	xi
CHAPTER 1: INTRODUCTION AND BACKGROUND	1
INTRODUCTION	1
BACKGROUND	2
STATEMENT OF THE PROBLEM	4
Empirical Statement	4
Normative Statement	5
PURPOSE OF THE STUDY	7
HYPOTHESIS	9
RESEARCH QUESTIONS	9
SIGNIFICANCE OF THE STUDY	10
METHODOLOGY	10
LIMITATIONS	11
DELIMITATIONS	11
STRENGTHS AND WEAKNESSES OF THE STUDY	12
DEFINITION OF TERMS	12
ORGANIZATION OF THE STUDY	14
CHAPTER 2: REVIEW OF RESEARCH, THEORY, AND LITERATURE	15
INTRODUCTION	15

BACKGROUND OF THE HEAD START FADE EFFECT.....	17
THE LINK BETWEEN POVERTY, LEARNING, AND ACHIEVEMENT.....	18
EARLY INTERVENTION AND PRESCHOOL PROGRAMS.....	24
Theoretical Basis of Head Start Program	24
<i>Findings from Formal Evaluations of the Head Start Program</i>	26
<i>Reasons for Head Start's Failure</i>	27
<i>Head Start Fade Persist Despite the Success of Other Programs</i>	29
Successful Early Childhood Education Studies.....	30
<i>The Abecedarian Program</i>	31
<i>The Chicago Child Parent Center</i>	31
<i>The Perry Child Development Center</i>	32
<i>The STAR Experiment</i>	32
<i>Three Critical Conditions for Success</i>	32
<i>Early intervention</i>	33
<i>Duration</i>	33
<i>Intensity</i>	34
<i>Other Variables That Contribute to Successful Early Education</i> <i>Programs</i>	34
<i>How Is a Program Determined to Be Successful?</i>	35
CLASS-SIZE RESEARCH	37
Early and Refining Studies	38
Large Scale Studies: Demonstration and Empirical	40
<i>1984: Texas (House Bill 72)</i>	40

<i>1984: Tennessee (DuPont)</i>	41
<i>1985: Tennessee (STAR)</i>	41
<i>Introduction to the Benefits of STAR Experiment</i>	42
<i>Recommendations for Class-Size Change</i>	44
<i>Six Elements of STAR’s Success</i>	44
<i>Associated Research on the Benefits of STAR’s Small Classes</i> ...	45
<i>Outcomes of STAR Experiment</i>	46
<i>STAR’s Impact on Impoverished Students</i>	47
<i>Benefits of Random Assignment/Heterogeneity</i>	48
<i>Gains in Psychological Sense of Community (PSOC) and</i> <i>Resulting Self-Concept</i>	50
<i>Longitudinal Gains of STAR Students</i>	51
<i>1986: Indiana (Project Prime Time)</i>	52
<i>1990: Tennessee (Project Challenge)</i>	53
<i>1991: North Carolina (Burke County)</i>	54
<i>1996: Wisconsin (SAGE)</i>	54
<i>1996: California Class Size Reduction (CSR)</i>	55
ECONOMIC RATIONALE FOR SUCCESS AND NEXT STEPS	57
Economic Rationale for Implementing Successful Pre-K and Head Start Programs	57
Steps for the Future.....	58

RATIONALE FOR EFFECTIVE THEORY TESTING.....	59
CHAPTER 3: DESIGN AND METHODS.....	63
INTRODUCTION	63
RESEARCH DESIGN	63
STUDY TYPE	66
RESEARCH METHODS	68
Actual Participants and Conditions.....	68
DATA COLLECTION PROCEDURES	69
DATA ANALYSIS.....	71
DELIMITATIONS	72
LIMITATIONS.....	72
STRENGTHS AND WEAKNESSES	73
CHAPTER 4: RESULTS OF DATA ANALYSIS.....	75
INTRODUCTION	75
HYPOTHESIS	75
DESCRIPTION OF THE SCHOOL DISTRICTS	76
DATA COLLECTION	79
RESULTS OF DATA ANALYSIS	82
QUALITATIVE DATA: DEVELOPMENT OF A THEORY TO SUPPORT REDUCING THE HEAD START FADE.....	88
Class Size	91
SUMMARY	95

CHAPTER 5: SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH	96
SUMMARY	97
Summary of Successful Early Education Programs and Class-Size Research	97
<i>Additional Elements of Successful Pre-K Programs</i>	97
<i>Alignment of Class-Size Research to Elements of Successful Pre-K Programs</i>	99
<i>STAR Outcomes</i>	100
<i>Longitudinal Gains</i>	100
Summary of the Purpose, Design, Method, Hypothesis, and Outcome of This Study	101
Outcome of the Study	102
Unexpected Limitations Encountered During the Research Process.....	103
CONCLUSIONS.....	104
RECOMMENDATIONS	105
Policy	105
Practice.....	105
Advancing the Theory in Future Program Research.....	106
REFERENCES	108
APPENDICES	120
APPENDIX A: RECOMMENDATIONS FOR CLASS-SIZE CHANGE	121
APPENDIX B: IRB APPROVAL	122
APPENDIX C: SYNOPSIS OF STAR EXPERIMENT	123

APPENDIX D: SUMMARY OF SMALL-CLASS STUDY	124
APPENDIX E: WHY SMALL CLASSES “WORK”	125
APPENDIX F: MAJOR IMPEDIMENTS TO SMALL CLASSES	126
APPENDIX G: STAR FINDINGS ON INSTRUCTION—EXECUTIVE REPORT	127
APPENDIX H: DESCRIPTIVE STATISTICS MAT 8 SUPPORT FOR TABLE 4	128
APPENDIX I: WILCOXIN MAT 8 DATA SUPPORT FOR TABLE 5	129
APPENDIX J: WILCOXIN MAT 8 DATA SUPPORT FOR TABLE 6	130
APPENDIX K: MEAP T-TEST DATA SUPPORT FOR TABLE 7	131
APPENDIX L: MEAP DATA SUPPORT GROUP 1 FOR TABLE 8	132

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1	<i>Beginning Kindergartners' School Readiness Skills by Socioeconomic Status (SES).....</i> 20
2	<i>School Demographics by District.....</i> 76
3	<i>Fourth Grade MEAP Reading Outcomes (2003 through 2006) for the Three School Districts (Percentage of Students Who either Met or Exceeded State Grade-Level Content Standards).....</i> 78
4	<i>Descriptive Statistics: Metropolitan 8 Reading Test Normal Curve Equivalent (NCE) Scores for Grades 1, 2, and 3.....</i> 83
5	<i>Wilcoxon Signed Ranks Test: Metropolitan 8 Reading Achievement Tests – First to Second Grade Comparisons.....</i> 84
6	<i>Wilcoxon Signed Ranks Test: Metropolitan 8 Reading Achievement Tests – Second to Third Grade Comparisons.....</i> 85
7	<i>T-Test for One-Sample: Michigan Education Assessment Program (MEAP) Test Scores – Fourth-Grade Reading.....</i> 86
8	<i>Descriptive Statistics of District, County, and State Comparisons of Fourth-Grade MEAP Reading Outcomes (2003 through 2006) for School District A (Percentage of Students Who either Met or Exceeded State Grade-Level Content Standards and District A Group 1.....</i> 87
9	<i>Theory Development: Examples of Enduring Effects of Small Class Sizes.....</i> 92

CHAPTER 1: INTRODUCTION AND BACKGROUND

INTRODUCTION

The Head Start Fade phenomenon has long been a concern in American education, particularly in relation to our continued failure to meet the educational needs of impoverished students. Overall, Head Start participants do not show a measurable difference in their grades two, three, and four outcomes when compared with children who have not attended Head Start. The data tell us the following: (a) Head Start participants are not mastering the readiness skills required for success in elementary grades and later schooling and (b) Subsequent schooling experiences in kindergarten (K) and the early grades (1–3) do not reinforce the Head Start efforts, so those students achieve on par with their peers who are not in Head Start.

There are, however, compelling data from empirical, scientific-based research (SBR) and theory to support a solution to this problem of Head Start fade. It can be found by examining the common factors of successful early education programs as well as the longitudinal gains associated with small-class sizes. This chapter provides background on the fade effect, the foundational theories and practices at work in successful preschool and K–3 programs, and corresponding theories and outcomes of small-class sizes. It also outlines the following: (a) statement of the problem, (b) purpose, (c) hypothesis, (d) research questions, (e) significance, (f) methodology, (g) limitations, (h) delimitations, and (i) strengths and weaknesses of this study. Finally, it presents an overview of this document’s organization and a definition of terms used throughout this document.

BACKGROUND

In 2001, the Family and Child Experiences Survey (FACES), a comprehensive, national review of Head Start programs, acknowledged that 43% achieved proficiency in reading skills (less than 30% of all participants), but that these children were not on welfare or receiving free school lunches (Department of Health and Human Services, 2003). In follow-up data, the report documented, however, that children in most programs lost the benefit of those early gains in elementary school. “Though some participants show increased gains in I.Q., those opposed to federal funding claim that the positive effects fade or completely disappear by third grade in most programs” (Barnett, 1993, p. 40). Thus, research has generally shown that students participating in the Head Start programs are not mastering the prereading and premathematics readiness skills important for success in the primary grades.

The ones paying the price for these failing Head Start programs are our nation’s poorest children. These children then grow up to be poor, undereducated, underemployed adults, thus continuing the vicious cycle of intergenerational poverty. Impoverished children who enter school behind their peers in language and math skills significantly increase their educational risk in primary and later grades (Lee & Burkham, 2002). In a meta-analysis of socioeconomic status (SES) and achievement, Sirin (2005) found that children who live in poverty are at greater risk and have less access to opportunities to participate in high quality preschool programs taught by certified teachers, thereby entering kindergarten one to three years behind peers in language and development.

The establishment of the Horace Mann Common Schools was supposed to be the “great leveler,” giving all children the benefits heretofore accorded to the wealthy (Haller &

Kleine, 2001). Later, President Johnson's 1965 "War on Poverty" legislation funded Head Start as part of the Economic Development Program to attack the effects of poverty systematically by providing academic intervention, health services for families and parenting education through county implemented preschool programs. Despite such sweeping measures, America continues to battle issues of quality and equity while the greatest risk of academic failure remains for those families on welfare.

On January 8 of 2002, *The No Child Left Behind Act* (NCLB) was signed. Although the stated intent, as indicated by the title, was to provide equal opportunities for education to children of all socioeconomic status (SES), the means to achieve that goal were founded on the notion that all schools, regardless of (SES), be held to the same academic standards and that those standards be held in place through nationally mandated, research-based testing. "Scientifically based research requires the application of rigorous, systematic, and objective procedures, systematic empirical methods, and rigorous data analyses that are adequate to test the stated hypotheses and justify the conclusions drawn" (NCLB, 2001, pp. 126–127). Schools whose students fail these tests lose funding. For many impoverished schools whose students don't meet the national testing standards, this means severe budget cuts to their already struggling programs.

In this era of high stakes standards and testing, poor children become victims of a system that fails to regard the impact of poverty on achievement. This is a problem referenced by Rothstein who suggested national standards must be measured to be equitable and "Testing alone will never address the pre-existing developmental gaps that prevent children from making academic gains" (2005, p. 36).

Educators commit to a primary assumption that all children can and will achieve given the opportunity and time. However, until national policy is created and funded at the legislative level that attends to the SBR findings on successful early childhood programs, children will not be given their right to an equal educational opportunity.

The NCLB Act is correct in its fundamental concept: early childhood education should be founded upon rigorous, SBR and sound theory. The question is which research and which theories? Since the NCLB has not produced its stated objective of leaving no child behind, perhaps it's time to consider other alternatives.

Research on small class sizes seems to be the most promising research to date. Sirin's (2005) meta-analysis of the effects of poverty on children found that small classes were among the (SBR) initiatives that provided a successful intervention for students at risk. Research on small classes, and in particular, the Tennessee Student-Teacher Achievement Ratio (STAR) experiment, provided empirical evidence that small class size does level the playing field for students, given the same conditions and program implementation (Word, Johnston, Bain, Fulton, Zaharias, Achilles, Lintz, Folger, & Breda, 1990).

STATEMENT OF THE PROBLEM

Empirical Statement

Empirical data have established that there are some short-term gains from Head Start programs but that a Head Start fade does exist. For example, several studies with strong research designs have demonstrated that Head Start has short-term cognitive, affective, and social benefits for children in poverty (Lee, Brooks-Gunn, & Schnur, 1988; Lee & Loeb, 1995; McKey, Condelli, Granson, Barrett, McKonkey, & Planz, 1985). By the same token, a

national review of Head Start programs by the Department of Health and Human Services stated the following: “When the school readiness level of the nation’s poor children was assessed, it became clear that Head Start has not eliminated the gap in educational skills and knowledge needed for school success” (DHHS, 2003, p. 1). In addition, the decline increases for those who are not involved in follow-up intervention (Copple, Cline, & Smith, 1987; Lazar & Darlington, 1982; Lee, Brooks-Gunn, Schnur, & Liaw, 1990; McKey et al., 1985; Miller & Bizzel, 1983, 1984).

Normative Statement

In addition to the empirical data on the prevailing Head Start fade experienced in most schools, there are also empirical data on the longitudinal gains of some select programs that have succeeded in preventing the fade effect. Given the longitudinal, empirical evidence that exists for these effective early childhood programs, there should not be a fade-out effect in any early childhood education program that utilizes the same successful theories and practices. In the Head Start national research and evaluation studies (DHHS, 2003), and in Barnett’s reports (1993, 1995) of 10–11 large-scale preschool and Head Start studies, two programs, the Abecedarian program in North Carolina for children age 0–5, and the Chicago Child Parent Centers (CPC) for children ages 3–9, showed persistent increases. In addition, Michigan’s Perry Child Development Center (PCDC) study for three- to four-year-old children showed a small fade effect initially but then showed robust longitudinal gains with reductions in grade retention and special education referrals. All three programs demonstrated these outcomes. Some state-level preschool studies provided similar results. In addition, Barnett (1993) conducted a cost benefit analysis of the PCDC in Ypsilanti, Michigan and found that of five studies only Perry showed persistent increases in

achievement data with significant follow-up, low attrition, and real-life outcome measures. Longitudinal studies showed significant gains in achievement, high school graduation, and college enrollment over peers not enrolled in these programs. Similar results were found in the comprehensive DHHS (2003) Head Start study report which had examples of both successful and challenged program results in the comprehensive, national review.

Another program that helped close the gap was the Michigan School Readiness Profile five-year follow-up study (MSRP, 2004), which found that four-year-old children living in poverty had success in closing the gap in student achievement and was reported to have saved the state \$11 million a year in the cost of grade-level retention and special education services.

The problem for this study, based upon the gap between the empirical and the normative when framed as a series of questions is as follows:

1. Why does this gap between the empirical and the normative persist?
2. How have the following helped researchers understand and mitigate the discrepancy between the normative and the empirical as stated here?
 - prior evaluations
 - new research
 - evolving theories
3. How might recent research and theory (1990-2007) contribute to a greater understanding of the importance of pre-K and small class sizes and their long-term, positive effects on student outcomes? (Pre-K programs in and of themselves do not offer long-term, positive effects on student outcomes, whereas small class sizes do.

PURPOSE OF THE STUDY

The purpose of this small-scale study is to explore how Head Start Fade can be eliminated or moderated by applying the key theories and practices found in research on successful early childhood programs and small class sizes. The intent is to test an evolved theory, based upon prior work, against the professed theory.

Evidence of a relationship exists between the successful elements of the CPC, PCDC, and Abecedarian programs and the successful elements of the small-class size Tennessee STAR experiment. They have all been able to offer longitudinal gains. STAR has shown that certain practices have to be in place if the benefits of the early years of schooling are to last (Word et al., 1990). Evidence of these practices exists in Head Start programs where the treatment effects last beyond high-school graduation (Barnett, 2003). The empirical data that the STAR experiment and various small class-size studies provide adds to the knowledge base and assists in explaining the fade effect by analyzing Head Start enrollees who have, or have not, participated in small classes.

The successful programs, Abecedarian, CPC, and PCDC had some characteristics in common. All had early intervention, duration of more than one year, daily instruction provided by certified teachers, and a small-class requirement of 14-16 pupils. Each program implemented high quality parent education programs, and each of the programs also provided developmentally appropriate curriculum and assessments. Some evidence of random assignment of children to teachers also exists. Equally important were the existence of transition programs to K-3 and the ability of students to establish a cohort as a result of

remaining in a program for more than a year. For example, in the CPC program, children were able to remain with the same group of children until age nine if a parent wished.

The STAR experiment provided the SBR to support the theoretic positions advanced by Ramey and Ramey (1989) in the Abecedarian study. Works reported by Finn and Achilles (1999), Finn, Gerber, Achilles, and Boyd-Zaharias (2001), Krueger and Whitmore (2000), Nye, Hedges, and Konstantopoulos (1999) and others have shown that for the early achievement gains to have enduring effects through subsequent years, three major conditions must be met:

1. *Early intervention*: when the student starts school
2. *Duration*: provide small classes for three, preferably four years so the child can learn about school (an apprenticeship for years of successful “work”)
3. *Intensity*: maintain the small class all day, every day

As previously mentioned, the Abecedarian program, CPC, and PCDC incorporated these factors into their designs. In addition to these three, the STAR experiment identified three other elements that have shown to contribute significantly to the gains that are part of this theory testing design (See Appendix A). The STAR experiment also addressed the importance of parental involvement and transition programs.

When considering Head Start fade, the gap between the empirical and the normative is clear. Forty years of empirical data show that the fade exists. In addition, the data on developmentally appropriate early childhood programs that contribute to improved student outcomes are robust.

It is in the interest of all children to have early childhood experiences that parallel the longitudinal gains made by children in Head Start and preschool programs such as PCDC,

CPC, and the Abecedarian program. Hopefully, in the 2008 reauthorization of Head Start, the credible, reliable data provided by these programs will be utilized to upgrade Head Start and moderate or put an end to the fade effect.

HYPOTHESIS

Group I Head Start students who experience full-day kindergarten every day in small classes, and continue in small classes through grades one, two, and possibly three, will not demonstrate the Head Start fade effect, or will have less fade, than will Group II Head Start students who have half-day kindergarten on alternating days in small classes and who have large classes in grades one, two, and possibly three.

RESEARCH QUESTIONS

1. Which factors in the research theory, (a) early intervention, (b) duration, and/or (c) intensity, will determine whether or not small-class Head Start students will demonstrate less fade-out effects than students not receiving those services as measured by the MEAP and other data at fourth grade?
2. Using available data, what, if any, relationship exists between academic success and (a) attendance, (b) behavior and/or (c) socioeconomic status (SES) for Head Start youth in small classes as compared to students not receiving those benefits?
3. What indicators of achievement do at-risk youth who attended a Head Start program demonstrate on the fourth-grade Michigan Educational Assessment Program (MEAP) test that other fourth-grade students who did not participate in Head Start programs show?

SIGNIFICANCE OF THE STUDY

Currently, a great deal of national attention is focused on the importance of pre-K experiences. There is agreement that services must not only be increased but proven effective and that effectiveness will come from the successful application of SBR. The STAR experiment is recognized as one of the most important pieces of research in decades (Cawelti, 2002) and important because of its length, size, scope, and empirical research base (Mosteller, 1996). As demonstrated in enduring benefits studies using STAR data, increased high-school graduation rates and college admissions and other indicators of success may prove this study's significance for one compelling reason in particular. In a state struggling with a declining economy such as Michigan, STAR research could provide the solution to increasing high-school drop-out rates if small-class size and other success factors are adopted into all pre-K and K-3 classrooms. Appropriate use of small classes could possibly provide economic relief for schools (e.g., Sharp, 2000; Achilles, 2004; Achilles & Sharp, 1998).

METHODOLOGY

This study used a non-experimental, longitudinal, retrospective explanatory design to test a theory. Johnson, the innovator of this design proposed a "new two-dimensional classification of non-experimental research" (Johnson, 2001, p. 3). Johnson (2001) described the primary purpose of the research to be descriptive, predictive, or explanatory. The second purpose addresses time dimensions such as cross-sectional, longitudinal, or retrospective. He offered this new design because "social research does not often lend itself to controlled inquiry of the experimental design; adding this model lets us consider cause and effect" (p. 3). Johnson suggested that the new design format eliminates prior design issues and lends

itself to “looking at several studies over time that more consistently align with current research where extensive development of findings cannot be falsified.”

In terms of methods, this researcher planned to follow the progress of two groups of children (Group I and Group II). Group I experienced Head Start in small classes and subsequently attended small full-day kindergarten classes and small classes in grades one, two, and three. Group II also experienced Head Start in small classes but did not subsequently attend either small, full-day kindergarten classes and/or small classes in grades one, two, and three. Grade two- and/or grade-three outcomes were examined for groups of students who had experienced small Head Start classes of 15 to 17. The researcher examined the relationship between the factors of duration and intensity and Head Start fade as revealed by achievement outcomes and scores on the fourth-grade MEAP test.

LIMITATIONS

One primary limitation of the study was the fact that the data sample was relatively small. This was due, among other reasons, to student transience and the lack (or loss) of available records. This was especially problematic when there was no central storage for classroom assignments. Without those data, it was impossible to determine if the students remained a cohort, and if they did, for how many years.

DELIMITATIONS

Delimitations were the yearly results of test data used to determine if any fade effect had occurred. The data were disaggregated by race, gender, and socioeconomic status (those qualifying for free and/or reduced lunch). Outcome data were collected from 1998–2005. A

second delimitation was that data collected on curricula, staff development and training, parent involvement, and transition programs were compiled as context data for each school as reported in the P.A.25 school improvement and annual reports.

STRENGTHS AND WEAKNESSES OF THE STUDY

The external validity of the study is strong because it used two formats for yearly assessment, norm-referenced tests (NRT) and criterion-referenced tests (CRT). Data were collected after the fact and were not manipulated or controlled. The procedures for administration and collection of the fourth-grade Michigan Educational Assessment Program (MEAP) tests were in place for all districts and were in accordance with Michigan Department of Education (MDE) guidelines. No accommodations were made to testing procedures for the study, and no special circumstances existed around testing.

Internal validity was strengthened by the use of multiple assessments that can be examined retrospectively, thereby establishing a longitudinal database over the four years time frame. The use of NCE and CRT tests align with the measurement format used in STAR research.

DEFINITION OF TERMS

For the benefit of the reader, the following are definitions of terms used in this document:

1. At-risk: At-risk is a term that applies to students who are at-risk of failure by virtue of having two or more of the conditions that may detract from academic achievement. The strongest factor is poverty, in particular, for those children whose families are supported partially through welfare. Other factors include being two or more years

- behind in grade level, having a high incidence of absenteeism and/or disciplinary incidences, being victims of abuse, or having been low-birth weight babies.
2. Class Size: “The number of students for whom a teacher is primarily responsible during a school year” (Lewit & Baker, 1997, p. 113). “This is an addition problem” (Achilles & Finn, 2002 p. 11).
 3. Early Intervention: The age at which a child begins schooling. Bloom’s research identifies the importance of early learning before traditional schooling, ideally starting from age 0-5.
 4. Head Start Program: Developed as part of the *Elementary and Secondary Education Act*, Head Start (PL 89-10), it was adopted in 1965 to deliver community-based services focused on school readiness skill development, health services, and parent-training program components.
 5. Improved Education = QE2: This means that small classes offer quality, equality, and equity. Quality means higher achievement, behavior, and citizenship. Equality means that all participants get the same treatment; i.e., no group gets more or less. Equity means that although minority and hard-to-teach youngsters benefit more from small classes than do better students, all students benefit (Achilles, Finn, & Bain, 1997; Finn & Achilles, 1990; Robinson & Wenglinsky, 1997).
 6. Program Duration: This means that students have small classes in preschool and then continue to have small classes through third grade.
 7. Pupil-Teacher Ratio (PTR): “The number of students in a school or district compared to the number of teaching professionals” (McRobbie et al., 1998, p. 4). “In some venues, all educators are part of the computation, including counselors,

administrators, etc. In this division problem, the divisor is important” (Achilles & Finn, 2002, p. 11).

ORGANIZATION OF THE STUDY

Chapter 1 provides an introduction, background, and research rationale for a Head Start Fade study. It includes research methodology, significance of the study, limitations, delimitations, and research questions. Chapter 2 provides a review of the literature, research, and theory related to the Head Start Fade effect and preschool programs that have not shown a fade. It adds small-class research and the SBR results of the STAR experiment as possible solutions to the fade effect. Chapter 3 describes the research design and methods in detail, as well as the Human Subjects Approval letter (See Appendix B). Chapter 4 presents retrospective data on small classes in elementary K–3 programs and MEAP achievement results at grades three and/or four. Chapter 5 presents a summary of findings, discussion, and conclusions, and offers recommendations for policy, practice, and further studies.

CHAPTER 2: REVIEW OF RESEARCH, THEORY, AND LITERATURE

INTRODUCTION

Why is it that children in Head Start programs who demonstrate gains in achievement and socialization begin to lose the positive effects once they enter elementary school such that by third grade gains have progressively disappeared on achievement tests (U. S. Department of Health and Human Services, 2003)? This phenomenon, known as the Head Start Fade or “fade effect” is a serious problem.

Children who have not achieved mastery of school readiness skills in pre-K programs are less likely to be proficient in reading and computation by the end of third grade than their peers who do attend pre-K programs. This early developmental gap may increase the likelihood that these students will later drop out of school. Successful schooling experiences are critical to our nation’s success in halting the debilitating, cyclical effects of failure, particularly for children trapped in “intergenerational” poverty cycles. For educators to resolve the complex problems of the fade phenomenon requires that they commit to implementing strategies theories and programs to ensure that all children master the school readiness skills required to transition successfully to K-3 and to ensure that early education programs have long-term support for learning.

This review of literature has been divided into six sections: (a) background of the Head Start Fade Effect, (b) the link between poverty, learning, and achievement, (c) an evaluative history of early intervention and preschool programs, (d) class-size research, and (e) economic rationale for success and next steps, and (f) a rationale for theory testing. The history of Head Start Fade Effect is briefly traced in the first section. The second section

outlines the overall conditions of poverty for U.S. children and the relationship between poverty, learning, and ultimately, achievement. The third section is divided into two parts: (a) an evaluative background of the failing Head Start program that addresses the flaws in its underlying theory and (b) successful pre-K and K-3 programs and their foundational theories. The fourth section explores class-size research as related to successful early education programs, with special emphasis on the six critical elements of the highly successful STAR program. Finally, the chapter concludes with an exploration of theory testing as a first step toward building successful early education programs founded on SBR and time-tested theory.

Basic questions for this literature review are as follows: (a) “How can the results from SBR in small classes for primary grades in elementary schools and other SBR help explain the fade effect?” and (b) “How might these studies help moderate the phenomenon, especially as shown in Head Start programs?” Not all early education programs discussed in this literature review employ the Head Start program. Some do, some don’t, and some simply draw on certain principles of the Head Start program without following the model exactly. Head Start has been considered a prototype for many programs and has certainly inspired them, but many educators have explored alternative approaches, as well.

Haller explained the purpose of the literature review as the identification of a problem that exists between normative and empirical conditions. Framed in a discrepancy model, the problem is explained as residing in a gap between what is and what ought to be (Achilles, 2001).

BACKGROUND OF THE HEAD START FADE EFFECT

Head Start Fade was first discussed around 1970. Nearly 25 years later, Barnett (1993, 1995) did a review of pre-K and Head Start programs and identified two comprehensive studies: one was a review of 36 national studies, and a second was a review of 22 long- and short-term programs. In these studies, young children demonstrated gains in Head Start only to have those gains diminish and/or disappear in most programs by third or fourth grade. This “fade effect,” also found in national studies by the Department of Health and Human Services (DHHS, 2001, 2003), has become a growing concern to educators. Similar studies, such as those performed by Finn, Gerber, and Achilles (2001), found that it is “. . . not uncommon that immediate cognitive benefits, reflected in tests of academic achievement tend to decrease over time so that experimental and comparison-group students were indistinguishable on tests three years after students left the program” (p. 160).

Evidence from various sources resulted in researchers developing theories that help provide answers to this problem. For example, STAR, Tennessee’s 11,600-student longitudinal experiment on class-size, and other companion studies, have provided empirical evidence to advance a theoretical direction to help explain the Head Start Fade. As Lindbloom suggested, an independent, original thesis could lead to a meaningful new approach to an existing problem (Lindbloom, 1959).

Continuation of the fade effect is one reason that Head Start programs have faced national scrutiny since the 1990s, especially when questions of program reauthorization arise. In 1998, the Family and Children Experiences Survey (FACES) documented national trends in Head Start programs (DHHS, 2003). Reported themes include the following: (a) Children in Head Start showed gains but were still behind in the skills required for kindergarten

readiness; (b) Head Start programs with lower achievement were correlated with a lack of certified teachers whereas Head Start programs with certified teachers demonstrated greater gains; and (c) A lack of coordination between county social service agencies, Head Start, and preschool programs resulted in a fragmented delivery of services to families.

In 2002, the Head Start Impact Study compared the results of those who attended the program with those who did not. Early results showed small gains in achievement for Head Start children in vocabulary and prereading skills but not in math or oral comprehension skills (Jacobson, 2007).

This distressing trend of persistent Head Start fade not only raises the question of dubious theory and practice in early education programs, but it clearly reveals the woeful lack of success in alleviating the vicious cycle of poverty and low achievement. Given that Head Start was founded to help improve the education of poor children, a brief review of poverty and education is relevant for the present study.

THE LINK BETWEEN POVERTY, LEARNING, AND ACHIEVEMENT

Without targeted and continuing intervention, each succeeding generation of poor people continues to demonstrate poverty behavior. We have had knowledge of this trend of “intergenerational” poverty from as far back as 1965, but it was recently discussed by Quinn in a dissertation on the cultures of intergenerational poverty (2005). First, a brief review of the current state of affairs for poor children in this country is presented. Hodgkinson’s material shows the continuous cycle: low birth weight, low achievement, failure, and generations of Head Start kids not making it (1992). Quinn recognized that poverty is intergenerational (2005).

In the early 1990s, Hodgkinson reported alarming statistics on children living in poverty:

- 13% of all children are regularly hungry.
- 25% are born to unmarried mothers.
- Over 20% are born to drug-addicted moms or about 350,000 children.
- Approximately 81% have health issues.
- About 2% of every 100,000 juveniles are incarcerated.
- 21% or 3.3 million are almost assured of being educationally retarded or difficult to teach.
- Every day in America, 40 teenage girls give birth to their third child. Most teenage moms have a greater probability of delivering babies with low birth weight or babies born prematurely, which can be predictors of learning difficulties. (Achilles, 1999, p. 9)

Despite the obvious and persistent condition of poverty for many U.S. children, little has been done in the way of federal or statewide measures to remediate these conditions, particularly when it comes to education. For example, in 1996, the United States spent 50% of its research budget on space exploration but less than 11% on education (Mosteller et al., 1996). Several studies point to results gained from high-quality research that would support substantial economic benefits for improving educational outcomes for students from families with low socioeconomic status (SES). Some researchers have even presented evidence that to do so would be relatively cost effective. Bracey (1999) pointed out that the cost of educating all preschool children nationally using the Perry Child Development Model would represent less than 15% of the already meager federal budget for education.

Significant research has been done on the difference between low SES children and high SES children in terms of school readiness. The research clearly outlines that low SES children are not operating on a level playing field from the very start of their education. In reviewing the impact of at-risk factors in school success, Newman (2003) found that more than one-half of all American children had one or more at-risk factors and 15% had three or more. Challenging the NCLB notion that assumes that all children start on a level playing field, Newman asked educators and policy makers to consider the following statistics (see Table 1) on school readiness levels for children in poverty (2003, p. 287).

Table 1

Beginning Kindergartners' School Readiness Skills by Socioeconomic Status (SES)

School Readiness Skill	Lowest SES	Highest SES
Recognizes letters of alphabet	39%	85%
Identifies beginning sounds of words	10%	51%
Identifies primary colors	69%	90%
Counts to 20	48%	68%
Writes own name	54%	76%
Amount of time having been read to prior to kindergarten	25%	1,000 hours
Accumulated experiences with words	13 million	45 million

Cooley (1993) found that over 60% of the reasons for performance variance attributed to school failure were related to poverty, and communities with a higher poverty rates generally experienced lower student achievement levels on state and standardized tests. The children in Cooley's study who demonstrated the lowest achievement on the state tests shared the following characteristics: (a) They were raised in single-parent families with

school-age children; (b) They resided in families in which adults had not graduated from high school; and (c) They came from families with low socioeconomic status.

The Cooley study outcomes were similar to a 2002 Michigan study by Maylone who used test scores and average family income as predictors of achievement on the MEAP test. In a summary of the findings, Maylone reported a predictive strength of 0.749 in multiple regression studies for students, using the values of three district SES factors:

1. Percentages of students receiving Free and Reduced Lunch
2. Lone-parent families
3. Mean family income

Maylone gave several examples of the phenomena. One was:

Students in the Adrian district who had a Mean Household Income of \$51,686, and where 33.2% of students received Free and or Reduced lunch, and 12% lived in Lone Parent Households, had average scores of 53 on the MEAP test. [Maylone compared] Adrian scores with the affluent West Bloomfield community where 5.5% of students had Free and or Reduced lunch, 6% lived in Lone-Parent Households, and the Mean Household Income was \$132,080. By comparison, the West Bloomfield students received an average MEAP score of 76. (pp. 96-97)

Maylone's predictions based on the three factors were all within 2% of the actual test outcomes in the districts.

Demographer Hodgkinson's (2003) report, *Leaving Too Many Children Behind*, explored similar statistics that were important in the first five years of a child's life. Indicating that our national standards for K-12 programs promote the need for quality childcare and early school programs, Hodgkinson (2003) and Newman (2003) reported that

we must recognize that children do not start the race at the same place and may not finish at the same time. Hodgkinson identified other alarming statistics such as the number of infant deaths by murder, student and family transience and mobility, and increasing poverty rates for children. The lack of adequate funds to support low-income families and assist with childcare and Head Start/preschool programs places thousands of children in an uncertain future.

In addition to U.S. researchers presenting data regarding the dire state of child poverty in the U.S., outside sources such as UNICEF have come to similar conclusions when placing the U. S. in a global context. Unicef (2007) recently compiled results of 21 of 25 industrialized nations on specific areas of overall child well-being and indicated the following:

- Children are worse off in the United Kingdom (21st) and the United States (20th).
- The United States has the most children living with step-parents and single-parent families.
- The U.S. ranks 12th on educational well-being, which combines test scores, educational attainment, and the transition to employment (next to highest ranking, 17th).
- The U.S. finishes dead last in relative poverty, 25th (On some individual components or dimensions, there are more than 21 nations for which there were data).
- The U.S. finishes 24th in infant mortality, 22nd for low birth weight, and 22nd in the percent of children who eat their main meal of the day with their parents.

The lack of medical care and poor nutrition often associated with poverty can result in children who face additional learning challenges. In a meta-analysis of SES and achievement,

Sirin (2005) found that children who live in poverty, and are therefore subject to inadequate medical care and nutrition, are at greater risk for school problems than are children from families with higher SES. Concluding that family income substantially impacts a student's achievement and knowing the relationship that SES has to school success, Sirin (2005) suggested the following:

To significantly reduce the gap in achievement between low-SES and high-SES students, policy decisions at the local, state, and federal levels must aim at leveling the playing field for students deemed to be at risk academically as a result of their family SES. (p. 446)

Other studies comparing minority children to white children demonstrated this lack of a level playing field when it comes to IQ. For The Early Childhood Longitudinal Study, Kindergarten Cohort (ECLS-K) researchers sampled 23,000 kindergarten children. It showed that at the beginning of kindergarten, African-American and Hispanic children scored more than half a standard deviation lower on reading and math, or the equivalent of eight points, on an IQ test with a standard deviation of 15 than did Caucasian children (Haskins & Rouse, 2005).

Rothstein (2005) performed an in-depth review of the achievement gap as related to low SES and offered suggestions for improving the academic and affective responses needed for children and families. Some, but not all, suggestions related to increased resources. However, the resounding message in Rothstein's work was that it will take a larger response than schools alone can provide to address these problems. According to Quinn (2006) and Edmondson (2005), correct theory requires that early intervention and parent participation address many of the issues and Head Start has done both correctly.

Finding that at-risk students are enrolled in schools that are financially challenged, (i.e., school poverty that abets student poverty), Sirin (2005) highlighted some successful interventions that have assisted students in academic outcomes. Among these are Title I and Head Start, small-school and class-size initiatives, early childhood education, and summer school programs. Achilles and Mitchel (1999) had similar findings. Sirin advocated the need to reduce the gap for children of the “haves and the have-nots” (p. 446).

Research on small classes provides empirical evidence that small class size does level the playing field for students. In particular, the Tennessee STAR experiment demonstrated increased gains for minority and Title I eligible students; moreover, the gains continued to increase after the children left the experiment (Word et al., 1990). The children’s gains were tracked in the Lasting Benefits Study (LBS) and Enduring Effects analyses (Nye et al., 1992, 1993, & 1997; Finn, Gerber, Achilles, & Boyd-Zaharias, 2001). In Michigan alone, the 2005 Michigan School Readiness Progress Profile (MSRP) for the program years 1999-2004 showed that the state would save \$11 million annually as a result of achievement gains in the program the state sponsored for four-year olds in public schools. These gains minimized the costs generated by special education referrals and grade retention for continued failure that could have been avoided.

EARLY INTERVENTION AND PRESCHOOL PROGRAMS

Theoretical Basis of Head Start Program

Created in 1965 as part of the Elementary and Secondary Education Act (ESEA), PL 89-10, and as a major component of President Johnson’s “War on Poverty,” Head Start programs were developed to increase the likelihood that all children, especially those living

in poverty, had access to preschool programs. Early Head Start services traditionally operated in community settings, employed noncertified staff to instruct children for half-day sessions, and included parent education and health services to children. There were no direct relationships or connections to K-3 school programs.

The Head Start program was conceived and designed based largely in response to the findings of the 1967 Coleman Report, as reported in (Haller & Kleine, 2001), that attempted to explain variance among academic outcomes of children from different groups. The report suggested that an increase in children's motivation could improve confidence, self-concept, and subsequently, learning. The Coleman Report determined that of the variables studied, family background had the greatest influence on student achievement, significantly more than other education variables that had been previously studied: "Family background factors were estimated to be roughly 1.5 times more powerful than were teacher quality factors (13.3% vs. 9% respectively). The remainder of the variance in student achievement, 83.6%, remained unexplained, that is, due to unstudied factors" (Haller & Kleine, 2001, p. 35).

In what has been described as perhaps the most damning paragraph ever written, the report's authors said, "Taking all the results together, one implication stands out above all: That schools bring little influence to bear on a child's achievement that is independent of his background and general social context." (Haller & Kleine, 2001, p. 37)

Research has shown that early theories centered on improving a child's self-concept and accompanying family relationships as a means to increase a student's school-readiness levels and improving developmental gains. Shapiro (1982) also proposed that improving a child's relationship with the family could improve school success. In alignment with these

theories, early Head Start programs were designed as a series of community-based programs for children and families to deliver school readiness reading, social skills, and health services.

Unfortunately, many educators began to consider family antecedents as reasons for failure. The Coleman Report led many educators to believe that family background had the most influence on achievement, even though the findings showed that 83% of factors that might have had a positive impact on achievement remained unknown. The Coleman Report results only accounted for about 17% of the reasons that contributed to achievement. It's important to remember much of the Coleman Report focused on the physical plant and other characteristics of schools not always related to learning.

Findings from Formal Evaluations of the Head Start Program

The Westinghouse Learning Corporation completed the first national Head Start Evaluation study in 1969. The evaluation assessed results from students enrolled in the Head Start summer programs for the years 1965, 1966, 1967, and 1968. No pretest was given because of the student age and development. Researchers tested 148 children along with a comparison group of children not enrolled in Head Start using two tests of cognitive ability. Though problems with validity in the evaluation were noted, results showed children not enrolled in Head Start outscored those children enrolled in the program. Magidson (1982) commented on programs that had weak quasi-experimental designs such as the original Head Start would not be likely to yield strong results from children in the summer program. The original design showed in an examination of disaggregated subtests by SES that children who scored better on the tests also came from higher SES backgrounds.

Later reviews of the data yielded other results. Among them were those of Barnow (1973) who reported that the program was beneficial for Mexican and African-American children but not for White children. Magidson (1982) found some positive results of Head Start for White children from his study, however, the author cautioned readers to be aware that the findings did not originate from a true, randomized experimental design.

Head Start analysis and results from the Westinghouse Study suggested that the early Head Start theories needed to be questioned. The initial evaluation results, as well as most national follow-up research studies of Head Start, have failed to demonstrate that theories on improved achievement were correct as originally conceived.

Reasons for Head Start's Failure

In addition to the faulty foundational theory, the fact that early Head Start programs had no direct relationships or connections to K-3 school programs has also been noted as a noted as a problem. This was identified in the FACES 2001 national study and was later identified as such in the national evaluation reports of Head Start programs (DHHS, 2001, 2003). This lack of transition and coherence, notably absent in most programs nationally, may have occurred as a result of policy makers and program implementation staff historically being engaged in a series of decisions that mirrored an early change strategy: the “Ready, Fire, Aim” method of program development (Fullan, 2001).

Another primary reason for the failure is that resources are not, and have not been, available at the federal level to enroll all eligible children in *universal* pre-K programs (open to all children in the country whose parents want them to participate). According to a recent Unicef (2007) policy brief, Gayl noted only 10% of the nation's three- and four-year-old children and 17% of all four-year-old children are enrolled in a state-funded preschool.

By comparison, Belgium, France, and Italy enroll more than 90% of their children from three to six years of age in free preschool programs. Japan joins that group with four-year-old attendees. The brief highlights a strong correlation between preschool attendance and achievement. Most nations with universal pre-K programs out-performed United States students on the Third International Math and Science Study (TIMSS). Until the nation levels the playing field, educators can be expected to fail in providing children with equal, quality educational opportunities. Regarding the pervasive power of poverty, Bracey (1999) stated the following:

Poverty, like gravity, is a condition. Gravity acts upon people in profound ways. So does poverty. To overcome the effects of gravity and fly requires great effort. Men have tried for centuries without success. To overcome the effects of poverty will require more effort than we are now making. (p. 330)

In *Class and Schools* (2004), Rothstein advises that schools must be supported in the battle to eradicate poverty at every social and economic level.

In *Savage Inequalities*, Kozol (1992) described the run-down, under-funded conditions of schools in lower SES areas. At-risk children attend school in these impoverished economic environments and enter school behind their more affluent peers in language, social, and economic capital. The phenomenon of “school poverty” reported by the U.S. Department of Education relates to lower achievement scores in schools that are poverty-stricken places (Achilles & Mitchel, 1999, p. 6). Beyond the ravages to instruction, children in these under-funded schools experience poor air quality, inadequate libraries, poor technology, and old, outdated materials. With higher rates of teacher turnover caused by good teachers transferring to less challenging schools, at-risk children often have to adjust to

new teachers more frequently and deal with the effects of instability in their classrooms. In addition, these children enter school behind their peers in language, math, and social skill development, placing them at increased educational risk in the primary grades. The risk to success for young children is greatest for families receiving welfare benefits.

Another factor to consider is that of aides versus certified teachers. Rowan (1994) suggested that a review of the Comprehensive School Reform (CSR) research should closely examine differences in how many CSR programs implement instruction. Finding that teacher aides were used extensively in CSR work, Rowan suggested that improved gains for children depended greatly on delivery of the “instructional core” (1994, p. 301). Aides, however, are not the same as teachers, a finding STAR and Prime Time in Indiana validated (Achilles, 1999). Researcher Darling-Hammond (1998) suggested that positive gains in achievement happen in significant ways when teachers who are proficient in their content knowledge deliver content instruction for all core subjects.

Head Start Fade Persists Despite the Success of Other Programs

Fifty years after the founding of Head Start, the Head Start fade remains despite the application of different theories supporting other successful early childhood programs such as the Abecedarian Program in North Carolina, the Perry Child Development Center (PCDC) in Michigan, the Chicago Child and Parent Centers (CPC) and other SBR class-size studies highlighted in this review. Sarason (1990) chastised America in *The Predictable Failure of School Reform*, noting that in the years since the Elementary and Secondary Education Act was passed in 1965, despite spending billions of dollars, education has not reached the goal of having all children achieve the necessary school readiness skills essential for later school

success. The results found short-term gains in some program areas in most states (DHHS, 2003) but a notable lack of long-term gains.

Overall, the underlying theories from the Coleman Report that provided the foundation for initial Head Start programs have not stood the test of the 2007 standard of SBR. The earliest programs were not based on empirical research or research that focused on practices to increase student learning. Both as a policy decision and as a legislative program initiative, Head Start has failed to meet its mission so far.

Based upon ideas presented up to here, it seems appropriate that other early theories and current research be reviewed. The theoretic construct for the present study, based on a collection of relevant research and theory, points in a direction that should assist in creating programs that provide children with the foundation to yield success in school. The following review includes small class and early childhood program research findings that suggest a new theoretical position that can help resolve the fade effect.

Successful Early Childhood Education Studies

In a national analysis of Head Start and preschool programs, Barnett (2003) identified programs in which students demonstrated minimal fade effect and maintained strong longitudinal results. The Abecedarian program in North Carolina and the Chicago Child Parent Center (CPC) provided the strongest evidence, and the Perry Child Development Center (PCDC) students in Michigan demonstrated a mild fade effect. The Tennessee STAR experiment in small class sizes should also be mentioned as it has had the most impact in the field in terms of positive, reliable data on the longitudinal benefits of small classes, as well as reduced fade effect. Longitudinal studies of PCDC students at age 40 showed improved graduation rates with a significant number of students continuing their education through to

college and professional careers. The same report highlighted 39 other state-funded preschool and early childhood education programs that yielded positive student achievement results.

The Abecedarian Program

The Abecedarian Program started in North Carolina in 1972. It enrolled 123 children ages four months to age five for a duration of five years. It had 57 youngsters in the treatment group that involved high-quality childcare programs taught by certified teachers and 54 youngsters in the control group that was not treated. Students were tracked for two decades, and data were collected at 12, 15, and 21 years of age. Compared to the control group, children who participated in the Abecedarian Program demonstrated the following (Kluge, 2006, p. 7):

- Higher cognitive tests scores from toddlerhood to age 21.
- Higher academic achievement in both reading and math from the primary grades through young adulthood.
- More years of education and a greater likelihood of attending a four-year college.
- Higher average age of parents when the first child was born.

Participants in the treatment group showed significantly higher scores on cognitive assessments in reading and math from primary grades through adulthood than did participants in the control group. By age 21, students who had participated in preschools with substantial funding from 1978 to 1998 experienced positive impacts in improving developmental competence in a variety of domains.

The Chicago Child Parent Centers

The Chicago CPC enrolled students who attended daily and also offered health and parent education components. At age five, children could elect to remain in the CPC program

until age nine instead of a public school. The CPC showed strong longitudinal gains in achievement through ages 15 and 21 (DHHS, 2003).

The Perry Child Development Center

The PCDC Head Start program began in 1960 and included 123 African-American three- and four-year olds from low SES families. Children were randomly assigned to attend a high quality preschool program with certified teachers. In follow-up studies 40 years later, Barnett found that the children who attended this program were less likely to repeat a grade or be referred for special education services were less likely to be referred for special education services, and were more likely to graduate from high school and attend college (DHHS, 2003). Approximately 65% of children in the program group had I. Q. scores of 90+ as compared to 28% of children from the group that had not participated in the program (Kluge, 2006).

The STAR Experiment

Tennessee's 1985 STAR experiment has come to be the landmark of empirical studies on class-size. It is known for its breadth and depth of reliable data, as well as its outstanding longitudinal outcomes. Researchers analyzed data on 11,601 students over a period of four years in its longitudinal experiment on class-size. Further details of the experiment are summarized in Appendix C.

Three Critical Conditions for Success

Three common conditions for success have emerged from research on the most successful early education studies to date. Access to data on these three conditions has been readily available, yet it has been neglected by many of today's program designers. It is hoped that by gathering the data and presenting them in a cohesive, coherent fashion that

educators will consistently implement these factors into their designs of future early education programs. The three major conditions for success as shown from the Abecedarian Program, the CPC, the PCDC, and the STAR experiment are:

1. *Early intervention*: when the student starts school
2. *Duration*: provide small classes for three, preferably four years so the child can learn about school (an apprentice for years of successful “work”)
3. *Intensity*: maintain the small class all day, every day (Ramey & Ramey, 1998)

Early intervention.

Research on cognitive development has indicated that the first five years of a child’s life are the most important with regard to language and social development. Bloom (1964) explained the importance of developmental learning, and his theories indicated the importance of starting early when the rate of growth is strongest for children in education programs. Piaget’s research also indicated that the first five years are pivotal in a child’s development. Ramey and Ramey (1998) created a framework for early interventions and articulated six principles of efficacy that built on the concept that “fragmented, weak efforts in early interventions are not likely to succeed, whereas intensive, high quality, ecologically persuasive interventions can and do” (p. 109).

Duration.

Earlier research by Reynolds (2000) showed that the duration of a program is important in assisting students in their transition from preschool to primary grades. According to Reynolds, the results in reviews show “The Abecedarian Study and [the] Chicago Child Parent Center [CPC] Programs directly support the conclusion that extended

early childhood programs can promote more successful transitions to school than preschool intervention alone” (p. 17).

Intensity.

Bloom’s (1964) early research advanced the importance of developmental timing, noting that learning occurs when children are young and the rate of growth is strongest.

In consideration of the findings of Bloom, Piaget, and others, it stands to reason that for continued student growth, educators would want to start early and continue program treatment for multiple years.

Other Variables That Contribute to Successful Early Education Programs

Although these three are the most critical, there are additional facilitating factors that contribute to success for students that have come out of the previously named studies: (a) parent involvement (Edmondson, 2005), (b) community efforts to fight intergenerational poverty (Quinn, 2006), (c) effective evaluations correlated to the program’s theory and design STAR (Word et al., 1990) and (d) program coherence (Newmann, Smith, Allensworth, & Bryk, 2001). Research on instructional and program coherence in particular have found theoretical and empirical evidence as to why coherent programs that built upon other instructional strategies and programs tended to improve achievement (Newmann et al., 2001). Sustained program coherence created opportunities for sensible connections in curriculum and instructional practices. Newmann et al. also found that programs where staff worked together on common goals, used common instructional strategies, and were organized to support innovations had a greater chance of improvements.

In addition, all instructors in the three previously named programs were certified teachers with knowledge of developmentally appropriate programs. And it should be noted

that like early Head Start programs that emphasized health, the Abecedarian Program also assigned a registered nurse who worked directly with parents and provided health services to youth.

Another factor that contributed to their success is that researchers randomly assigned students to the treatment and control groups; they formed heterogeneous groups as a result, (therefore, the terms random assignment and heterogeneity are used interchangeably throughout the rest of this chapter). In the Perry program, some children were eligible for services even though their families were not considered to have a low SES. In the Abecedarian Program, children were placed randomly at birth, making it unlikely that children would have visible indicators of ability other than birth weight. This is in keeping with Head Start programs that also practiced heterogeneous grouping. They required that all classes contained several high SES children even though the program was specifically designed for low SES children and families.

How Is a Program Determined to Be Successful?

A critical requirement for high quality interventions is that the effect lasts beyond the treatment. The Abecedarian Program, the CPC, and the PCDC all show evidence of strong longitudinal gains. Compelling results for these successful pre-K programs also included a noted decrease in grade retention and special education referrals, as well as continued gains in achievement. Results also indicated decreased juvenile and adult crime, continued gains in school attendance, stronger high school graduation rates, as well as increased college enrollment.

State-funded preschool programs recognized for having higher achievement and childcare results through agency collaboration were compiled in a U.S. Department of Health

and Human Services report (2003). In a meta-analysis of all evaluations of state programs, results showed that the Perry program enhanced pupils' school readiness up to age seven. At 14 years of age, Perry program participants outperformed nonparticipants on school achievement tests, and by age 19, Perry program participants had higher graduation rates than their peers.

In Virginia, the final report of the Governor's Task Force on School Readiness (2005) recognized that the levels of a child's skill and social development are directly related to the quality of early childcare received. Further highlighting the benefits of longitudinal studies such as the Chicago PCP or the Abecedarian Program, Michigan's Governor Jennifer Granholm referenced the importance of the first three years of a child's life in later development. Thus, there is state-level interest in early education and small classes though prior implementation, such as Prime Time in Indiana and Class Size Reduction (CSR) in California (to be discussed in the next section), have shown that interventions need to be planned and implemented in developmentally appropriate ways if they are to be successful. In other words, success means that interventions yield results such as those shown by Perry, Abecedarian, CPC and STAR.

These preschool programs are clearly "out-performers" in the research as they all provide strong evidence of underlying program theories that have continued to yield robust achievement results for young children. The intent of this researcher is to promote a pre-K through grade three theoretical framework based upon such empirical and longitudinal research, as well as other companion studies that have heightened achievement for all children. The next section emphasizes the K-3 class-size research that contributes to a theory for early success and a reduction of the Head Start "fade" phenomenon.

CLASS-SIZE RESEARCH

What follows is a summary compilation of class-size studies to date. The studies may be considered significant for their outstanding contributions or for their exceptional failings. They are presented chronologically to illustrate the impact they have had on subsequent research, as well as on current understanding of early programs. Some of the evidence that has surfaced from these studies confirms and/or supports the underlying theories of these most successful programs.

Class-size research that has built upon the results of the previously mentioned examples of successful early intervention helps provide greater understanding of the Head Start Fade effect. The empirical STAR experiment provided the evidence to support the theoretic positions advanced by Ramey and Ramey (1998) in the Abecedarian study. Researchers such as Finn and Achilles (1999), Finn et al. (2001), and Krueger and Whitmore (2000), as well as secondary analyses by Nye, Hedges, and Konstantopoulos (1999) have indicated that for early achievement benefits to continue into subsequent grade levels, the specific conditions found in the Abecedarian Programs must be met. The longitudinal and varied research on small class size has provided direction in the theory testing that is part of the present study on ways to mitigate or moderate the Head Start Fade.

Although class size has been studied in the U. S. since about 1904, fairly widespread interest in class-size studies began to appear around the 1950s when Blake published an inquiry in 1954 (Word et al., 1990). This “force of sustained inquiry” (Mosteller et al., 1996; Achilles & Finn, 2002) has created a field of knowledge about class-size studies. The field of knowledge was built from class-size research that might be considered groundbreaking

studies from about 1900-1965, evaluation work from 1966-1979, and refining studies from 1978 to the present (Achilles, 1999, pp. 5-6). Achilles noted the following:

1. Early studies emphasized “common sense” and showed some advantages of small groupings of children for schooling.
2. The “refining” studies improved the earlier studies in several ways: scope, duration, method, description, explanation, continuity, etc. The early phase (1965-79) involved evaluations of “special” efforts such as Title I and other projects. Some may have been pupil-teacher ratio (PTR) rather than class size.
3. Large scale observation studies such as those of Lindbloom and Olsen (1971) have provided insights into operations and processes. The early phase concluded with the meta-analyses by Glass and Smith (1978) and Smith and Glass (1979).
4. Large-scale and state-wide demonstrations began in 1981 with the Project Prime Time in Indiana.

Early and Refining Studies

In a review of 267 studies in the early 1900s, Blake chose 85 studies to analyze, finding that 41% (35 studies) showed improved student achievement results for smaller sizes, 21% (18 studies) showed that large classes were better, and 38% (32 studies) were inconclusive (Achilles, 1999). Later studies by Fumo and Collins in 1967 occurred over the course of five years in Baltimore. These studies included “16,449 students in reading and arithmetic in class sizes ranging from 1–25, 26–31, 32–37, and 38 or more; overall, they found greater gains in the smaller class sizes. The effect was 7.3:1 in classes of 25:1 over the larger (26 or more) in 192 comparisons; in 96 comparisons involving minority students, the

ratio increased to 21:3 to 1. Small classes increased effects for minority students” (Word et al., 1990, p. 201).

Olsen (1971) completed reviews on class size in over 112 schools, seeking to determine the effects of small classes on student learning. Olsen compiled the results into “nine general conclusions about the positive effects of small classes” (Achilles, 1999, p. 22). Lindbloom’s reviews of small classes during the 1970s found that teachers created more conditions in small classes that enhance student learning (Achilles, 1999, p. 22). In a meta-analysis, Glass and Smith (1978) combined 77 studies and made 725 comparisons on the effects of differing class sizes on student achievement. They found many trends that were not evident when reviewing individual study results. For example, as class size increased, achievement decreased. A pupil who scored at about the 63rd percentile on a national test when taught individually scored around the 37th percentile when taught in a class of 40 pupils. The Smith and Glass (1979) meta-analysis showed positive effects of small classes on teacher and student affect. Achievement gains reported by Weglinsky’s research (1997) found that fourth graders who had small classes were well ahead of peers who were not in the small class design. In fact, gains on the average were as high as a half a year ahead for those students in small classes. Overall, a survey of research from 1980s–2007 suggests that early studies were poorly designed and analyzed, too brief (lack of direction), lacking early intervention, and not longitudinal (Achilles, 1999).

Small classes are good for all students, and have implications beyond the educational setting. Achilles (1999) noted in a matrix of instructional practices (See Appendix D) that the STAR and companion studies incorporated the most examples of best practice research on effective instruction. Some of the earliest studies and meta-analyses helped guide later

research on small classes. The following section explains a series of large, state-sponsored studies (Egelson et al., 2003) to look for clues for testing theories on class size.

Large Scale Studies: Demonstration and Empirical

To start this section, a timeline of selected class-size initiatives has been compiled for easy reference:

1984: Texas (House Bill 72)

1984: Tennessee (DuPont)

1985: Tennessee (STAR)

1986: Indiana (Project Prime Time)

1985: Tennessee (Project Challenge)

1991: North Carolina (Burke County)

1996: Wisconsin (SAGE)

1996: California (CSR)

1997: Alabama: (State Board of Education decision)

1998: Florida (House Bill 367)

1998: South Carolina

1999: Michigan, Michigan School Readiness Profile (MSRP)

2000: Georgia (House Bill 1187)

1984: Texas (House Bill 72)

In 1984, the Texas legislature passed House Bill (HB) 72, the first policy decision to distinguish between pupil-teacher ratio (PTR) and class size. The Texas legislature was specific in their intent to have class sizes of no more than 22 students in K–2. Distinction

between PTR and class size did not occur in policy again until the Tennessee Star experiment.

In 1984, Texas passed HB 72, mandating small classes (no more than 22 students) in K–2. The bill was amended in 1986 to include grades K–4. By 1994, when students who started in the small-size Kindergarten took the Texas Assessment of Academic Skills (TAAS) in grade 10 as sophomores. The TAAS scores at grade 10 began their mythical rise (through 1990) showing the influence of *early intervention* (K or pre–K) and *duration* (the longer a student is in small classes, the larger the long-term impact). (Achilles, 2002, p. 2.)

1984: Tennessee (DuPont)

Early studies helped researchers and theorists understand the importance of class size. For example, the DuPont study in the fall of 1984 involved two elementary schools in Nashville and showed that students in smaller classes performed better than students in larger classes (Bain, Achilles, Dennis, Parks, & Hooper, 1988; Whittington, Bain, & Achilles, 1985). Though small, the DuPont study served an important function as it was the pilot test for the highly significant STAR experiment.

1985: Tennessee (STAR)

STAR was the largest experimental, longitudinal, and randomized education study to date at the time. STAR tracked the progress of 11,600 students over a period of four years and provided detailed records, thereby leaving a legacy for future researchers. STAR was the state's response to the legislative sponsored class-reduction program and was conducted in partnership with four university research teams. STAR researchers confirmed the founding theories of the Abecedarian research gained from earlier class-size studies. STAR's research

was designed as a method of organizing schools to ensure that all children at every ability level succeed. Ouchi (2004) noted that the structure and organization used in delivering services must change for the culture to change. The STAR study showed that the system of schooling was broken and changed the organization of the classroom delivery structures in 79 schools. The Tennessee STAR study changed the organizational unit of delivering instruction (Achilles & Finn, 2002).

Because STAR was so large and had so much impact, it deserves an extensive exploration. Thus, this section of the chapter is divided into nine parts: (a) introduction to benefits of STAR experiment, (b) recommendations for class-size change, (c) six critical elements of STAR's success, (d) associated research on the benefits of STAR's small classes, (e) outcomes of STAR experiment, (f) STAR's impact on impoverished students, (g) benefits of random assignment and/or heterogeneity, (h) gains in psychological sense of community (PSOC) and resulting self-concept, and (i) longitudinal gains.

Introduction to the Benefits of STAR Experiment

Studies of small classes, in particular the Tennessee STAR experiment, have provided empirical evidence that given the same conditions and program implementation, small classes do help level the playing field for student outcomes. STAR also showed that small classes work for all students but offer increased benefits for minority, poor, or hard-to-teach children. Other benefits of small classes include increased effects of engagement, student initiative, and on-task behavior in small classes (Word et al., 1990). Further studies have strengthened research on positive effects of these characteristics that are strongly correlated with learning. For over 50 years, educators have learned through early childhood education

studies on class size that children benefit when they read, write, and compute on grade level by the end of third grade.

Does an answer for the Head Start fade effect exist that can assist in strengthening those early years, especially for at-risk youth? If the role of schooling is to ensure that all children learn and all learning is cumulative, it follows that to change the cycle of failure, educators should go to the start of schooling and change the organization of early childhood programs. Ouchi (2004) said that changing the structure of schools will assist in helping children who live in poverty to succeed. Achilles (2003) considered Demings' research that suggests that when over 85% of participants are failing, the system is probably broken. Working with a primary research team from four universities, Achilles suggested that the answer to maximizing the potential of the early years of schooling might be to "Change the Damn Box" (2003, p. 1). In other words, it's easy enough to say that we should think out of the box, but perhaps we should consider changing the box itself, i.e., the structure of the schools. To a degree, STAR's success can be attributed to the fact that it changed the organization of the school.

There is strong evidence to support changing the system. For instance, the research team designed and implemented the STAR experiment in partnership with the Tennessee State Legislature and Department of Education. By changing the organization of schooling into three class designs, using the small class (13–17) as one of the models for the K-3 years, all children achieved at levels greater than the levels of randomly assigned peers in larger ($n = 22-25$) classes. The small class design provided by the STAR study may help explain the fade effect. STAR follow-up studies on high school completion, parallel the Michigan PCDC longitudinal research showing a high rate of graduation for students in small classes.

The Michigan School Readiness Profile was a five-year follow-up study on four-year olds who had attended the pre-K Head Start program and four-year olds who were not enrolled in a pre-K Head Start program. Researchers found that children living in poverty had success in closing the gap in student achievement when they attended small classes. Such gains were reported to have saved the state \$11 million a year in grade level retention and special education services (Xiang & Schweinhart, 2002).

The STAR study provided the evidence to support the theoretic positions advanced by Ramey and Ramey (1998) in the Abecedarian study. Analyses and articles by Finn and Achilles (1999), Finn et al. (2001, 2003), Krueger and Whitmore, (2001), Nye et al. (1999), and others have shown that for the early achievement benefits to show enduring effects in subsequent years of school, schools must employ the three previously mentioned factors of early intervention, duration, and intensity. STAR formally included those in its

Recommendations for Class-Size Change

Achilles (2007) recently summarized new findings in Hong Kong about class size. For a summary of these class-size findings see Appendix E.

Six Elements of STAR's Success

The six elements of STAR's success are as follows:

1. *Early intervention:* Start when the pupil enters "schooling" in K or even pre-K.
2. *Sufficient duration:* Maintain the small-class environment for at least three, preferably four, years for enduring effects. Encourage parent involvement in schooling.
3. *Intense treatment:* The pupil spends all day, every day in the small class. Avoid Pupil-Teacher Ratio (PTR) events, such as "pull-out" projects or team teaching. Develop a psychological sense of community (PSOC), close student-teacher relations,

- and coherence. Although teacher aides may assist in the building, there is scant evidence that they influence student outcomes positively; except as required by law, reduce aides by attrition.
4. *Use Random Assignment* or heterogeneous classes in early grades to facilitate peer tutoring, problem-solving groups, student-to-student cooperation, and active participation and engagement.
 5. *Employ a Cohort Model* for several years so students develop a sense of family or community (PSOC). STAR results show the power of both random assignment and a cohort model. “Looping” adds teacher continuity to the cohort, and may be a useful strategy for added benefits.
 6. *Evaluate Process and Outcomes* carefully, and share results. Appropriate-sized classes in elementary grades will take policy and perhaps even legislation change.

Achilles (1999) noted that a solution to Bloom’s “two-sigma problem” (1984) for educators may be that a strategy as effective as one-on-one tutoring but less expensive be found for improving the learning of young students. Small class results come strikingly close to the outcomes of one-on-one tutoring and do not require extensive resources. Research has demonstrated that class size can be reduced at little or no cost if anything that takes children away from their primary teachers such as teacher aides and/or academic services such as Title I rooms is eliminated, offering a way to keep costs reasonable (Achilles, 2005).

Associated Research on the Benefits of STAR’s Small Classes

Reviewing the findings of the Tennessee STAR experiment, Finn and Achilles (1999) reported that antecedents to the STAR study represented over 100 research efforts in reducing class size. Four significant findings emerged from that research:

1. According to Glass and Smith, reduced class size can be expected to increase academic achievement with positive effects realized even from small reductions in class size.
2. The major benefits from reduced class size are obtained as the size is reduced below 20 pupils.
3. Small sizes are most beneficial in primary grades.
4. The research consistently finds that students who are economically disadvantaged or come from ethnic minorities perform better academically in small classes (Finn & Achilles, 1999)

In addition, “STAR researchers analyzed multiple teacher behaviors and found that statistically significant differences” (Achilles, 1999, p. 68) between effective and less effective teachers on 12 variables directly related to student achievement.

Outcomes of STAR Experiment

The beneficial outcomes recorded from the STAR experiment range from raised test scores to improved behavior. “The small classes made the highest scores on the Stanford Achievement Test (SAT) and Basic Skills First (BSF) Test in all four years (K-3) and in all locations (Rural, Suburban, Urban, Inner City)” (Word et al., 1990 p. 1). Data also showed that STAR outscored those in comparison schools at every level (Zaharias, Achilles, & Cain, 1995).

Students in small classes were, on average, 7.1 months above their peers who were not in small classes. This is for students who entered at kindergarten and remained in small classes through third grade. This evidence demonstrates the importance of duration and intensity, two key correlates to increased gains for children. In a study of pre-K and Head

Start programs, Barnett found that “evidence of effects was strongest among the experimental studies that had used random assignment to form comparison groups” (1995, p. 34). Results were verified multiple times over the four years of the study by reviewing results of data collected on achievement through test scores, attendance, behavior, participation, and self-concept as measured by the SCAMIN. Students in the small classes consistently outscored their peers on all measures (Word et al, 1990).

STAR's Impact on Impoverished Students

The STAR experiment resolved many questions about how children could benefit from schooling. It showed that “Small-class advantages appeared for all students participating in the study. The gains were similar for boys and girls, but were greater for impoverished students and African-American students from inner-city school groups that are traditionally disadvantaged in education” (Biddle & Berliner, 2002 p. 17). For example, higher achievement gains for minority and free- and reduced-lunch students occurred in the STAR experiment. Mosteller (1995) noted “at the end of the second year the effect size for minorities was about double that for majorities” (Mosteller, 1995, p.119). Achilles, Nye, and Bain (1995) noted that “A pupil who entered STAR in K on average scores 18.6 points higher on grade one SAT reading test than did a pupil who entered STAR at grade one” (p. 27).

The third year of the study ended after kindergarten. In the beginning, STAR students attended kindergarten on a voluntary basis. However, researchers discovered the benefit of K-2 and based upon these results, K became required in Tennessee (Mosteller, F. 1995 p.119). Similar results were found in STAR companion studies such the Wisconsin Student Achievement Guarantee in Education (SAGE) (Molnar, Smith, Zahorik, Palmer, Halbach, &

Ehrle, 1999) and in 1991 the Success Starts Small Study as reported by Egelson, Harmon, Hood, and Achilles (2003).

Benefits of Random Assignment/Heterogeneity

Random assignment (heterogeneous grouping) of STAR students was part of student placement in the STAR experiment.

The experiment carried out in 79 schools the first year randomly assigned both children and teachers to classes; each school had at least one of each of the three class sizes, (S), (RA with Aide), and (R), so that comparisons could be carried out within the same school. Otherwise, the effects on the groups of classes might have depended on the properties of the schools presenting the teaching or the neighborhoods where the children lived. (Mosteller, Light, & Sachs, 1996, p. 815)

There is evidence that random assignment has a significant impact on student achievement in the early elementary grades (e.g., Zaharias et al., 1995) and that this positive effect lasts through at least the twelfth grade (Finn et al., 2005). Indicators of increased achievement surfaced when examining small class results with those in the comparison schools. In fact, results from the STAR comparison schools on similar tests showed that randomly assigned (heterogeneous grouping) students had greater gains than those not randomly assigned.

Zaharias et al. (1995) reported the effect of random versus nonrandom assignment on reading and math achievement on a total of 1,157 students. Random group size was 499 and nonrandom was 658. Outcome measures included the SAT Total Reading and Total Mathematics scaled scores and the BSF Total Reading and Total Mathematics raw scores. The SAT is a norm-referenced test that compares students' scores at their grade to the scores

of other students and the BSF was a criterion-referenced test used to measure mastery of Tennessee's state Comprehensive Curriculum in grades one, two, and three. In reading, the randomly assigned students outscored the non-randomly assigned students on both tests at each grade level. Differences in achievement reached statistical significance by grade three. Analyses of math scores by grade three were also statistically significant, though the trend did not emerge until third grade. Using STAR data, Boyd-Zaharias (1993) identified positive gains for randomly assigned pupils over nonrandomly assigned cohort mates in grades one, two, and three. All pupils, whether entering with kindergarten in 1985–86 or without kindergarten in 1986–87, were randomly assigned to one of three class types: S (small), RA (regular size class with an aide), or R (regular no aide). Teachers were also randomly assigned. Though teachers changed classes each year, their placement was also random. It is clear that random assignment/heterogeneous grouping seems to provide numerous benefits all around, but it also contributed to an improved Psychological Sense of Community (PSOC) and self-concept in children, as measured by the SCAMIN scores. These scores showed significant gains for all children but especially for minority and of low-SES children (Word et al., 1990).

In an analysis of available national studies on the effects of ability grouping Mosteller et al. (1996) reported no significant academic positive effect for homogenous groups though parents preferred the higher ability grouping for their children. The negative effect of homogeneous classes, however, on less academically able students in studies of self-concept showed diminishing gains for those students.

Other comprehensive reviews of heterogeneous grouping practices and random assignment by Slavin, and later by Levin and Burriss (2006), show these two factors

may have a greater impact on achievement than does class size. Slavin's (1989) early work highlighted the Joplin Plan in which students formed a heterogeneous triad. In these triads, weaker individual students, typically from low SES, made substantial gains in achievement when paired with two students with strong developmental skills. Overall, low SES student success rates increased significantly as a result of this practice.

Long-term gains from heterogeneous placements were verified through an analysis of student scores in each of the three in-school designs in the Lasting Benefits Study (LBS) (Achilles, Nye, Zaharias, & Fulton, 1993; Nye, Zaharias, Fulton, Achilles, & Pete-Bain, 1994).

Gains in Psychological Sense of Community (PSOC) and Resulting Self-Concept

Slavin's (1987) synthesis contributed the best evidence to date on the negative effects of homogeneous grouping on elementary children's instructional and affective growth. "The results of these 43 studies were reviewed within four major divisions: (1) ability-grouped class assignments (n = 14), (2) regrouping for reading and mathematics (n = 7), (3) Joplin plans (n = 14) and (4) within-class ability grouping (n = 8)" (Boyd-Zaharias, 1993, p. 35).

As further evidence, a review of grouping practices by Mosteller et al. (1996) showed the negative effects of homogeneous grouping on student achievement and self-concept, particularly for low SES students who displayed an increase in negative behaviors in homogeneous groups. Fewer than 30 works were reviewed in this study, yet the consistent findings are significant as to the harmful effects on students' esteem.

Using the SCAMIN inventory, Project STAR reported on the positive effects of small class size on students' self-concepts. The positive effect of small classes on minority students' self-concept was significant, although there was no difference for other students.

Students in inner cities had higher self-concept scores from being in small, randomly assigned classes (Word et al., 1990).

Longitudinal Gains of STAR Students

The Lasting Benefits Study (LBS) (Nye, Achilles, Boyd-Zaharias, Fulton, & Pate-Bain, 1993) demonstrated that students from small classes continued to have gains through middle school. Test results from grade seven students showed increased student achievement with no fade effect. The STAR follow-up studies through grade seven also demonstrated that small classes have a significant and lasting effect on increasing student achievement (Nye, Achilles, Boyd-Zaharias, Fulton, & Wallenhorst, 1994). The Enduring Effects Study of Small Classes noted the longitudinal gains through high school of students from small classes (Finn, Gerber, Achilles, & Boyd-Zaharias, 2001).

Three studies of longitudinal gains were conducted based on STAR data. The first one showed that students who had participated in small classes were more likely to take college entrance exams. In particular, a greater percentage of African-American students who took the entrance exams had come from small classes (Krueger & Whitmore, 2001)

A second study using STAR data found that the graduation rates of students who had been in small classes for three or more years were higher, and were, in fact, similar to their peers with higher SES. The odds of graduating for low-SES students who had been in small classes for at least four years increased by 80%. The odds of graduating increased by 67% for students who had spent at least three years in small classes (Finn, Gerber, & Boyd-Zaharias, 2005). Along with increased graduation rates, Barnett (2003) noted that students continued to show persistent gains in student achievement, I.Q., decreased crime, and delinquency rates.

The third study showed that 3,922 STAR students who participated in small classes took more advanced courses in high school (Finn, Fox, McClellan, Achilles, & Boyd-Zaharias, 2006). The results also indicated that a greater number of students from small classes enrolled in math and foreign language programs. Researchers noted that the effect continued for up to nine years after students had left small class programs with no difference noted in SES.

Finn (2006) recently used follow-up high-school data to examine heterogeneity in a post high-school population to determine both the status risk and academic risk factors for young adults on attainment and educational outcomes in post-secondary and job-related success. Working in partnership with the National Center for Education Statistics, Finn found that school academic and behavioral engagement factors contributed significantly to later success in post-secondary school and careers. At-risk students who successfully completed high school and post-secondary education had stronger employability patterns as a result of school engagement.

In summarizing the STAR experiment, Finn and Achilles (1990) commented that “the research leaves no doubt that small classes have an advantage in reading and math in the primary grades” (p. 196). The 1984-1989 STAR experiment found the following to be important: (a) reduced grade retention, (b) fewer special education referrals in later grades, (c) higher test outcomes, and (d) reduced “gaps” between groups (Word et al., 1990).

1986: Indiana (Project Prime Time)

At a cost of \$300,000, Indiana implemented a two-year class size project for the early grades from 24 randomly assigned schools (Egelson, et al., 2002). Project Prime Time dropped class size down to about 18 in grades one and two and allowed the teacher to have

an aide when class size reached 24 (Chase, Mueller, & Walden, 1986). In the second year, staff could add either full-day kindergarten or have an aide. Most staff chose the aide in the second year. Prime Time was developed with the knowledge available at the time, but it was not fully understood that the addition of the aide changed the pupil-teacher ratio and was therefore not a true reduction of class size that many thought it to be. Later, STAR also showed that the PTR strategy of adding an aide did not yield higher achievement.

Positive results, however, were not consistent in the study. The results in Indiana's class-size initiative and later results from another statewide initiative in California highlighted the importance of duration. It is worth noting that this was not empirical research but rather a demonstration project (Achilles, 1999) of both early intervention and duration, a critical factor shown to enhance long-term gains as part of the theoretical construct of the Abecedarian Program that was confirmed in Tennessee's STAR (Word et al., 1990).

1990: Tennessee (Project Challenge)

“In 1990-1994, following STAR, Tennessee initiated Project Challenge, a field trial that involved implementing small classes in grades K-3 that was conducted in 16 of Tennessee's poorest counties” (Achilles & Finn, 2002, p. 5). “In Project Challenge, K-3 students did not experience significant gains until after the third year of being in small classes.” Project Challenge later showed why Prime Time participants should have opted for the full-day kindergarten. In Project Challenge, early intervention and duration had a stronger impact on achievement than did the addition of an aide (Achilles, Zaharias, Nye, & Fulton, 1995). Project Challenge later showed why Prime Time participants should have opted for the full-day kindergarten. In Project Challenge, early intervention and duration had a stronger impact on achievement than did the addition of an aide (Achilles, Zaharias, Nye, & Fulton,

1995). From the results in STAR and Challenge, one can infer that small classes are preventative not remedial. This research strengthens the results mirrored in companion studies that emphasize the importance of early intervention in improving achievement.

1991: North Carolina (Burke County)

In 1991, North Carolina's Burke County started another small-class initiative (Egelson, Harman, Hood, & Achilles, 2003). At that time, Burke County had a free-and-reduced-lunch student count of approximately 30% that reached 38% in 2000. Over the course of the study, costs reached \$3 million. Beginning in first grade, class size dropped, and each year thereafter another grade level was added until all students in grades one to three were in small classes at a total cost of \$3 million. By 2001, Burke County elementary students had exceeded state accountability gains in every subject except third- and fifth-grade reading where gains were .03% and 2.2% lower (Egelson et al., 2003).

1996: Wisconsin (SAGE)

In 1996–1997, Wisconsin opened the Student Guarantee in Education (SAGE) giving schools with over 50% free- and reduced-count students an extra \$2,000 per pupil (Egelson et al., 2003). With this money, class size dropped to 15 students in each class. The program began in first grade and one grade level was added each year thereafter. Project Challenge, Burke County, and SAGE all built upon the theories that had proven successful with STAR, and their achievement results tracked the STAR results. SAGE also proved to offer higher achievement gains for minority and free- and reduced-lunch students (Molnar, Smith, Zahorik, Palmer, Halbach, & Ehrle, 1999).

1996: California Class Size Reduction (CSR)

Biddle and Berliner (2002) noted that the California CSR initiative, based upon the same theories as those in Indiana Prime Time was a textbook example of how a state should *not* reduce class size. Like Indiana's Prime Time program, the California CSR provided evidence that small classes should begin early. Whereas Indiana phased in grades yearly, California's CSR spanned grades one, two, and three simultaneously and added K later. The lack of physical space and staff and other unanticipated program implementation issues presented major problems. There were simply not enough classrooms and teachers. Compared to STAR, the California CSR did not yield the anticipated results once implemented. Results from the Abecedarian Program helped researchers understand duration as one possible reason why the Indiana and California programs did not yield the strong results as anticipated.

To keep costs reasonable, the Indiana and California initiatives left an option for the third-year class to have an aide or a full-day kindergarten. The Abecedarian Program, STAR, and Project Challenge supported the importance of early intervention and duration, meaning that a full-day kindergarten that added grades one and two to the long-term process enhanced the importance of duration. In Tennessee's Project Challenge, implemented in grade K–3 simultaneously, the first year of testing at grade three produced no gains; in the second year, there were some grade-three gains, but gains at grade three became most evident following the full K-1, 2, and 3 effort when students finished the fourth year of Project Challenge. This evidence of duration was also shown later in the STAR experiment.

Duration was essential; it involved a small class treatment in place for three to four years. PTR only adds an adult, like an aide, and is not the same as small class size. As stated

earlier, STAR later demonstrated that an aide in the class, effectively changing the PTR, does not have a positive impact on achievement. This finding was echoed in the Key Stage 1 and Key Stage 2 (K-1, K-2) in Great Britain (Blatchford, Russell, Bassett, Brown, & Martin, 2003) as well as demonstrated in Prime-Time and California CSR. STAR also showed no positive effect from an aide in place of a full-time classroom teacher.

In addition to previously cited U.S. studies on the benefits of small class sizes, a considerable number of international class-size studies have been implemented in England (United Kingdom), Australia, Sweden, and the Netherlands. England's KS-1 (ages pre-K through grade three) and KS 2 (ages 7-11) with 20,000 students created robust, longitudinal outcomes (Blatchford, et al., 2003), and after this extensive exploration of small-class-size research, it can be seen that small classes offer a wide range of benefits. According to Achilles (1999), small classes offer "Three socially desirable benefits that are expressed American values. Those benefits of quality, equality, and equity provide measurable positive outcomes. This idea was expressed as Education Improvement (or EI) = QE2: Small classes offer Quality, Equity, [and] Equality. They can also be the basis for value-driven education leadership" (pp. 158–159).

There have been many other studies on class size, but they are not explained here in detail. In general, well designed class-size studies, including small-scale studies such as in dissertations and those that do not confuse class size and PTR provide contrasting evidence of points made here.

ECONOMIC RATIONALE FOR SUCCESS AND NEXT STEPS

Economic Rationale for Implementing Successful Pre-K and Head Start Programs

An argument can be made purely on the basis of economic advantages to offer pre-K and Head Start in small classes. Currently, the Head Start program represents the largest pre-K initiative in the country with expenditures in the \$6 billion range—and that is for programs that are relatively ineffectual. Consider the cost of crime as a result of teens dropping out of high school. In a recent discussion about the adverse impact of inadequate noncognitive growth at early ages, Heckman (2006) cautioned the nation in a *Wall Street Journal* article to be mindful of the social ills that increase as children continue to fail in this country. It is a well-known that drop-outs who are routinely unemployed or underemployed suffer from low self-esteem as a result of failure, frequently turn to crime. As recently reported by Nobel laureate economist Heckman (2006), “the cost of crime [in the U.S.] is \$1.3 trillion per year or \$4,818 per capita” (p. 2).

Consider what cost to society could be saved if early childhood programs became highly effective. Citing the increase in high-school drop-outs, Heckman noted that data collected from the Perry Preschool program in a follow-up to age 40 showed that participants had higher achievement scores and were more successful in a number of noncognitive measures, (e.g., factors like behavior, participation, attendance, and self-concept) than those in the control group. Heckman also noted the success of the Abecedarian Program in North Carolina in actually raising I.Q. as well as achievement, most likely because of the starting age of four months. This leads one to speculate that perhaps the Perry Program might have also raised IQ had it started children earlier than it did at ages three and four.

Steps for the Future

In a recent article, Lewis acknowledged that evidence is accumulating, both theoretically and empirically, that a fade can be avoided or at least reduced given certain conditions (2006). Lewis echoed what has been established in this study that some children who have pre-K programs demonstrate continued gains; however, once children enter grades K–3, a loss of those positive effects occurs. Lewis suggested that this may result from educators failing to administer elementary programs in ways that support and extend preschool. For example, educators may continue to allow K–3 class sizes of 25 or more in contrast to the average preschool class size of 15–17, or they may only support half-day kindergartens instead of full-day kindergartens where young children reap the benefit of the time that they had in preschool. Lewis (2006) went on to suggest that legislators could responsibly address this concern with policies that mandate and fund universal pre-K programs based on theories of successful national pre-K efforts such as the Abecedarian Program in North Carolina, the CPC in Illinois, and the Perry Preschool Program in Michigan.

Early childhood education and pre-K programs similar to those highlighted in this study that have strong evidence of gains have not been adopted as a nation policy despite evidence of the positive gains available to all students. Established policy based on available empirical research could provide greater certainty that the Head Start Fade dilemma will be addressed. As this issue is critical for children, policy must be grounded in research that involves “Two or more good quality, replicable, independent, empirical, rigorous, objective, systematic studies on positive effects of small classes on short term- and especially on long-term student outcomes as usually measured” (Achilles, 2003).

A comprehensive review of large-scale demonstration and empirical studies has been completed, and the benefits of applying successful foundation theories to our current early childhood education have been established. It is clear that the future of our early childhood educational programs rests firmly on employing sound, SBR to the design of our pre-K and K–3 programs if we are to move beyond the Head Start fade effect. Therefore, a discussion of the merits of quality theory testing is appropriate.

RATIONALE FOR EFFECTIVE THEORY TESTING

Small-class research must test theories related to methods and programming if we are to improve early childhood education, especially for children living in poverty. Hawking (2006) described theory testing in this way:

A theory is a good theory if it satisfies two requirements. It must accurately describe a large class of observations on the basis of a model that contains only a few arbitrary elements, and it must make definite predictions about the results of future observations. (p. 6)

Shapiro (1982) adopted the position that, “The argument that links program activities to outcomes is called the program theory” (p. 341). Further, he asserted the following:

One effect of explicating and testing the theory underlying the program structure is that the evaluator can demonstrate the relationship between the program activities and outcomes. In addition, as suggested by Suchman (1967), the theory testing approach to evaluation allows the evaluator to distinguish program failure from theory failure. (p. 341)

In 2007, Kelly and Yin suggested the following:

. . . exploratory work can be expected to be more expansive and speculative than confirmatory trials Drawing from Stokes (1997), we recommend that authors indicate from what stage of research their claims emanate: (a) basic or exploratory research, with the goal of advancing theory or discovering fundamental processes [or] (b) movement from informed theory to informed practice. (p. 136)

Stokes (1997) advanced the idea that research can be high or low in terms of contributions to theory and/or to practice. Stokes presented Pasteur's Quadrant (Brookings Institution, 1997) to identify potential methods used to develop new research findings. The researcher in the current study suggests that the hypothesis presented advances a theory both high in practice and high on theory when supported by the longitudinal research findings of the STAR study and others.

The earliest years of a child's life are a short, important period to help children develop their own building blocks of knowledge and experience that can help establish a path of success. Educators must understand theories that facilitate student achievement, use sound professional judgment, provide research, and continuing research and evaluation as a guide to establishing administrative policy and practice. Researchers and stakeholders must ask for evidence to further define the gap between empirical and normative outcomes that may help explain the fade.

In a commentary on SBR, Stipek (2003) suggested that three key strategies should be implemented to establish an evidence-based focus on issues from educators: (a) randomized field trials, (b) commitment to higher standards of inquiry and research, and (c) research that involves collaboration between practitioners and researchers.

Also, Stipek suggested that practitioners pay close attention to context; what may work in one context may not have similar success in another. Stipek cited Barnett's review of the Perry Preschool study and California's class-size reduction (CSR) program as an example. Both used class-size reduction but had far different results.

A deeper examination of the California program by Biddle and Berliner, 2002 revealed that the program was based on an unsubstantiated theory. Both Perry and California implemented the practice of starting early, but thereafter many differences existed in the programs. For example, the California study did not use the principle of duration. Biddle and Berliner (2002) considered "the California CSR effort a near textbook example of doing CSR wrong" (p. 20). Small class implementation theories offered indicators for how to run a successful program, but California's program development staff ignored these concepts. In summary, there was a plenitude of successful theories from other studies that California could have drawn on to avoid negative outcomes.

Other examples of Abecedarian Program theory, the Project Challenge and the STAR experiment showed that starting in third grade was too late. Burke County in North Carolina started with K-1 and added another grade level each year (Egleson et al., 2003). Students made significant gains as educators implemented a reform practice using Stipek's thinking and drew from the research on reform strategy improvements.

Another aspect of good theory testing involves aligning with external research teams. The STAR experiment and other related studies were an example of Stipek's call for educators to align with external research teams. The STAR experiment used a strong theoretic research design based on the results of the Abecedarian Program, which was field tested in the DuPont small class study in Tennessee. However, as Stipek illustrated, program

theories that work in one setting may or may not be effective in another. The ability to generalize and/or replicate the research may depend on how valid and reliable the research design is and if it yields consistent results. There were concerns regarding the validity and reliability of early Head Start program evaluations.

Adopting quality theory testing into early education design takes vision, courage, and an ability to inspire others—in short, it takes effective leadership. Heifetz (1994) suggested that leadership is not simply mobilizing others to address tough issues and solve them but rather to confront those problems we have not yet been successful in resolving. Head Start fade is a problem that would greatly benefit from good leadership and sound theory.

The theories and studies provided here in sequence offer a coherent, consistent, and defensible “Big Theory” to explain the Head Start fade phenomenon. Better yet, the “Big Theory” points to empirical testing that may mitigate or reduce the fade if data are either available or collected in a timely way.

Besides a summary of the research and theory on early childhood education, successful early intervention, poverty’s influence on schooling, and class size, the review in Chapter 2 has suggested a coherent theory on Head Start fade. Chapter 3 presents the design and methods for a pilot test (case study) of the theory within the limitations of available data.

The research presented in this study provides compelling evidence that a reform strategy exists to increase the likelihood that all children can master reading by third grade. Theories from programs like Abecedarian, CPC, and PCDC, and from a host of class-size studies, as well as overwhelming evidence as to the benefits of applying theories from such works all point to the fact that there should not be a fade if educators adhere to these theories and implement their programs in accordance with them.

CHAPTER 3: DESIGN AND METHODS

INTRODUCTION

The chapter identifies the questions that guided the research, the research design, the preferred methods for the data preparation and analyses, and the limitations and delimitations for the study. There is also a brief section on unanticipated roadblocks and adjustments made, as well as issues of validity and application of the theoretic framework.

The research design was originally conceived to a) assist in building an understanding of the Head Start “fade effect,” through retrospective review of K-3 student scores on achievement tests, and b) determine if the hypothesis presented supports the conditions, as verified by the research and data, that may lead to a reduction in the Head Start Fade effect in children by the third grade.

RESEARCH DESIGN

The research design was a non-experimental, retrospective, explanatory design (Johnson, 2001) used to test a theory. For the non-experimental, retrospective design Johnson (2001) proposed a “new 2 dimensional classification of non-experimental research.” (p. 3). Because the researcher used data that had been previously collected (2002-2007), the design also had strengths of a longitudinal study. Johnson described the first dimension as the primary purpose of the research (i.e., descriptive, prediction, or explanatory). The second dimension represents a time element (i.e., cross-sectional, longitudinal, or retrospective). This design classification is useful because “social research does not often lend itself to controlled inquiry of the experimental design, adding this model lets us consider cause and effect” (p. 3). Johnson (2001) also highlighted problems resulting from data analyses, even

robust ones like ANOVA and regression research, and argued that little can be achieved from single studies. Johnson suggested that the new design format eliminates prior design ambiguities and lends itself to “looking at several studies over time that more consistently align with current research where extensive development of findings cannot be falsified” (p. 8).

Considering the large number of non-experimental class-size studies and the few examples of experimental scientifically research based (SBR) studies, positive research findings that can be replicated are not abundant. However, findings in a few (SBR) studies such as the Tennessee class-size experiment (STAR), the Perry Pre-school Program and the Abecedarian (Ramey & Ramey, 1998) research stand apart from other studies in the significant and replicated positive results from small-class sizes on short and long-term student outcomes.

The enduring effects of the conditions that lead to long-term outcomes in STAR are the correlates of the theories proposed from the Abecedarian study (Ramey & Ramey, 1998). These theoretic elements that occurred in STAR (Word et al., 1990) were as follows: (a) *early intervention* (pre-K, or when the child starts schooling), (b) *duration* (3 or more years in reduced class size), and (c) *intensity* (same teacher for academic subjects all day at each grade level). These conditions contributed to an increase in student achievement in the STAR experiment. These same three elements occurred in STAR (Word et al., 1990) companion studies such as the Burke County, North Carolina class-size program, (1991-2007), and reported in several SERVE documents (e.g., Egelson, Harman, Hood, & Achilles, 2002). Using a matched-pairs design, researchers determined that small classes contributed in significant ways to increased student achievement. As the students reached ninth grade, a

review of the results for Burke small-class students found they had scale scores on the average of 8 points higher than those students paired in large classes.

Other positive long-term effects are shown in a study by Finn, Gerber, and Boyd-Zaharias (2005) that followed STAR students into high school and showed that small-class students took more advanced courses in high school than did their randomly assigned peers in larger classes.

The present study examined students who participated in Head Start as a proxy for early intervention and subsequently experienced these “enduring effects” in small classes. The combined results of these studies lend insight into the Head Start fade effect and provide a theoretic frame for the present small-scale research to test.

The researcher proposed identifying two subgroups: (a) Group I students who began in Head Start as early intervention and then participated from kindergarten through third grade in the small class sizes (16 to 20) having duration (3-4 years) and intensity (all day, everyday); and (b) Group II students who participated in Head Start as part of the early intervention, experienced some of the enduring effects—however, not consistently administered in each program each year and were in classes with sizes larger than Group I students. The researcher examined students’ norm-referenced test (NRT) and criterion-referenced test (CRT) scores on district administered achievement tests each year over a 4-5 year period (Head Start through second or third grade) in both subgroups. By comparing the achievement results of the two subgroups, ideally with the Michigan Educational Assessment Program (MEAP) scores at 3rd or 4th grade, the researcher sought to determine if the theoretically sound enduring effects in place in Group I had a positive influence on the students’ achievement over time, to reduce or ameliorate the Head Start fade.

STUDY TYPE

This study was originally to include several coordinated case studies following a common design in districts where requisite data were available. After completing several case studies, data would be aggregated and combined into a single database for further analysis. The school districts contacted for the present study reviewed the research proposal, agreed to participate, and provided appropriate written consent for the investigator to conduct the research. The review of two or more studies where data from more than one database were to be analyzed, in retrospect, complied with the research characteristics of the Johnson (2001) explanatory, retrospective, research design (Johnson, Type I. 2001, p. 10).

The guiding hypothesis for the study was as follows: Head Start students who subsequently experienced full-day, every-day kindergarten in small classes and grades 1, 2, and possibly 3 in small classes (Group I) will not demonstrate the Head Start fade effect or will have less fade than do Head Start students who have less than full-day, every-day kindergarten in small classes and who had large classes, (more than 20 students) in grades 1, 2, and possibly 3 (Group II).

The researcher planned to use available NRT and other outcome data for Group I and Group II. Demographic data, including gender, race, and socioeconomic status (SES) using school-lunch as a proxy for (SES) were obtained from school records as available.

RESEARCH METHODS

Actual Participants and Conditions

The researcher examined grade 2 and/or grade 3 outcomes for groups of students who had experienced Head Start classes of 15 to 17 students. Group I then entered K–2 or K–3

schooling in full-day, everyday kindergarten for grades 1, 2, and 3 in small classes. Group II students experienced Head Start in small classes, but did not subsequently experience full-day Kindergarten and/or small classes in grades 1, 2, & 3. The researcher estimated how duration (years in a small class) and intensity (same teacher all day, every day) influenced Head Start fade between students who entered K–2 or K–3 conditions (large or small class size) as assessed by grades 2, 3, and/or 4, including MEAP reading scores as achievement outcomes. Because of incomplete data, the “Guiding Hypothesis” (described above) was replaced by a “Working Hypothesis” as follows: Head Start students who subsequently experience full-day, every-day small classes in K and in grades 1, 2, and 3 will not demonstrate the Head Start fade effect for achievement tests completed in succeeding grades 1, 2, and 3, as well as 4th grade Michigan Educational Assessment Program (MEAP) reading data.

Three school districts originally participated in this study. One, a medium-size urban district with a high percentage of minority (Hispanic and African American) students located in mid-Michigan, had all-day kindergarten and class sizes of up to 20 (Group I) and of 24 or higher (Group II) for K and grades 1, 2, and 3. This district is located in an urban, blue-collar area, with access to a major state university and a large technology-oriented community college in the immediate area. A large group of professional parents send their children to magnet schools in the district, and other private schools, through a school-choice option and to township schools located in the most affluent area of the county. The primary employment in the area is automotive manufacturing. The district superintendent approved a request for the study and the District Research and Evaluation Department compiled the historical, retrospective database for years leading to 2004 outcome data. The districts’ Head Start

program was administered through the Intermediate School District. However during the years covered in of the study (1998-2005), the program changed management companies three times and was provided oversight by the city, a local social service agency, and a management company in Denver, Colorado, who recently came to the area when grant funds were revoked from the county provider.

Due to the high rate of mobility of students in the district (40% and higher), the number of students for each Group (I, II) in the study was smaller than anticipated. The second district was located in a small city in a fairly affluent county in southeastern Michigan. The district was considered as low SES based on the percentage of students qualifying for free or reduced lunch programs. The district borders on a large urban area and participates in cross- district school of choice programs. Head Start is part of the school program. The two subgroups in this district to be studied included a reduced class size norm of 16 to 18 (Group I) students in all-day kindergarten and continuing through grades 1, 2, and 3. A second group with all-day kindergarten and a larger student class size norm of 20 to 24 in grades 1, 2, and 3 comprised Group II. The district has approved the study through the assistant superintendent's office, which provided some data for grades 1 through 3 achievement outcome data.

The third district, a small, blue-collar, urban school community had small classes of 16-18 students for over 10 years and has full-day kindergarten programs. The Head Start program, administered by the county, has no connection to the school program. The district could provide no comparison Group II data, but was engaged as a subgroup to compare with state data on the academic achievement as measured by the Michigan Education Assessment Program (MEAP) tests completed in fourth grades.

Delivery of instruction highlighted developmentally appropriate programs and program Intensity (all day, every day) of instruction were obtained by interviews with the Superintendent or designee and possible principals. In all districts, historical data were reported on district-wide tests completed by each student at the regional office of placed at the district administration building. Approval was given for the researcher to have discussions with the school principals if required, but any data-searching requirements beyond that required funding of a researcher from the district office staff to mine data at school sites to obtain specific information on student outcomes. All districts were asked to eliminate any identifying information (name, social security numbers, etc.) to maintain student confidentiality.

For the study, the researcher required information on the following: (a) participation in Head Start, (b) educational delivery structures in later grades (K, 1, 2, and perhaps 3); and (c) records for Head Start students in Groups I and II through yearly testing in the spring on the norm-referenced tests (NRT) through 2004 or 2005 and winter (January-February) MEAP scores for fourth grade.

DATA COLLECTION PROCEDURES

Once the researcher determined that districts met the research design for small classes, had access to Head Start programs in the area and would provide for the researcher to have access to historical available data for review, the necessary documentation was acquired from the Eastern Michigan University Institutional Review Board (IRB) for protection of human subjects. Following receipt of the IRB Board approval, timeline commitments were confirmed. To maintain the anonymity, the names of the school districts

were not used. A final step was for each district to determine what process would be used to code District data to have students maintain anonymity once data were received by the researcher. The researcher followed the IRB human subject requirement and had district personnel remove identifying information from all data provided. In two of the three districts the researcher paid a district staff person to record, in coded format, all data received for analysis in the study.

The Michigan Department of Education (MDE) MiTracker database provided all required and available achievement, behavior, and attendance data for all students as needed. This database was accessible to review MEAP state assessment data and coded results were obtained through the Superintendents' district curriculum and assessment offices. Potential documentation, as available, of the Reading First (MDE) databases were reviewed, if available, for programs of study delivered K-3. Examination of any available schedules to document supplemental courses offered such as Title I and others that may be done outside the classroom for supplemental support, if available were to be reviewed.

Collecting data became difficult or nearly impossible over a period of 2-3 years that followed each districts' prior review and approval for supplying required data and information for a number of reasons: a) student mobility within and between school districts reduced the availability of longitudinal data on specific students b) NCLB, and Michigan's charter schools and school of choice have provided educational options for parents who can move their children to different schools if they are dissatisfied with their children's present schools.

Further difficulties arose because during the study two districts changed superintendents and all three experienced personnel changes in central office Curriculum or

Assessment Department leadership. New superintendents had to be updated and informed of the data collection process. Among the three school districts that enrolled an aggregate of nearly 15-17,000 students yearly, only one district was able to provide longitudinal data on some of their children. They submitted information on 68 children in five elementary schools who had participated in Head Start and remained in the district for more than one year. The remaining school districts were unable to provide consistent data (using the same achievement tests across grade levels) for the study. Data from the one district were included in the study and the other districts were eliminated because of the data issues.

DATA ANALYSIS

The researcher examined students' norm-referenced test (NRT) test scores from the IOWA or the Metropolitan Achievement Test (MAT 8) each year over a 4 to 5 year period (Head Start through second or third grade) in both subgroups. By comparing achievement results of the two subgroups, the researcher sought to determine if the enduring program effects in place in Group I influenced the students' achievement in this group. Data collected from the school districts were combined to form a single data set on which analyses were completed using SPSS –Windows, ver. 15.0. The data analysis used t-tests for one sample to compare MEAP reading scores to the state passing score of 400. Because of the mobility of the students in these districts, students with data for first, second, and third grades were included in the analysis. In addition, MAT 8 scores between first to second grades, and second to third grades were compared using the Wilcoxin test for the grade-to-grade score analysis within the district and the MEAP cut score of 400 to determine the extent to which

Head Start fade seemed to be occurring. All decisions on statistical significance of the findings were based on a criterion alpha level of .05.

DELIMITATIONS

Delimitations imposed by the researcher were the outcomes on test data yearly were used to determine if any fade effect had occurred for the purpose of exploring the hypothesis or outcomes of the study, disaggregated by race, gender, and free and reduced lunch for the students in each group as possible. The years the data were collected for the outcomes will be a second delimitation for those districts without a data warehouse. The data collected on curricula, staff development, parent involvement, and transition programs were compiled as context data for each school as available, but were limited by P.A. 25 School Improvement Plans and Annual Reports reported on the Michigan Department of Education (MDE) website.

LIMITATIONS

Limitations in the study were sample size for data to be collected caused by student transience. Although the original design called for 3 districts to be involved, it became impossible to obtain longitudinal data due an absence of data tracking systems in the districts. Data collected for sub-groups were small, a factor caused for reasons previously named.

A key limitation was the available data required a switch from the planned parametric analyses to use of the non-parametric statistics for small samples. Samples of students who

met the required conditions declined each year of analyses (e.g., from 45-29-18), such that analyses could only establish a trend, not clear statistical evidence.

In relation to understanding programmatic issues in school interviews, the mid-Michigan achievement database was available only for identified students in alphabetical order by grade, disaggregated by race and gender for the district as a whole. It was difficult with no central storage of classroom assignments to determine if any cohort remained together, and for the number of years desired in the research. No available documentation of school attendance for the student database was available; therefore, any interviews were generic in nature. It was possible to secure some teacher records with cooperation from principals' records and from other program information. However, program structures and documentation were generic in nature, not lending themselves to deeper understanding of context issues when reviewing scores.

STRENGTHS AND WEAKNESSES

The external validity of the study in using two formats for yearly assessment, both norm-referenced test (NRT) normative data and criterion-referenced test (CRT) data, is strong. Data were collected after the fact by one or more years and no variable was manipulated or controlled by the researcher. The procedures for MEAP test administration, data collection and reporting were in place in the Michigan Department of Education (MDE) for all districts. There were no accommodations to testing procedures for the study and no special circumstances around testing.

Internal validity was strengthened by use of multiple points of assessment, provided verifications were in place that allowed the assessment of the subgroups' progress statewide

for comparison NCE tests and individually for student content strengths and weaknesses on criterion-referenced tests (CRT) like the MEAP.

The internal validity should be strong because test items for the MEAP are correlated to the state contents standards to be taught. These grade-level content standards (GLCS), part of the state curriculum, are used for MEAP tests by MDE.

Chapter 3 provided the research hypotheses and a description of the methods and steps used to acquire the necessary data to complete the study. Chapter 4 presents the analyses of the data for the study.

CHAPTER 4: RESULTS OF DATA ANALYSIS

INTRODUCTION

This chapter provides the results of the statistical analyses used to address the research question developed for this study. The chapter has four sections. The first section provides a description of the three school districts selected to participate in the study. The procedures used to collect the data in the study are presented in the second section. The third section provides the results of the quantitative statistical analyses that were used to test the research hypothesis. The qualitative information that presents the theory development is presented in the fourth section of the chapter.

The researcher's purpose for this study was to explore how the Head Start fade can be eliminated or moderated by applying the research findings of successful early childhood programs and key theories and practices that provide the foundation for successful pre-school and early childhood programs identified in the research literature.

HYPOTHESIS

The working hypothesis developed for this study is as follows: Head Start students who experience full-day, every-day kindergarten in small classes and in grades 1,2,and 3 in small classes will not demonstrate the Head Start fade effect for achievement tests completed in grades 1, 2, and 3, as well as 4th grade Michigan Educational Assessment Program (MEAP) reading data.

A non-experimental, retrospective research design was used in this study. The data were obtained from three school districts on standardized test outcomes for

children who met the criteria for inclusion in the study. Group 1 included children who met three conditions: (a) *early intervention* when the students start schools (i.e., attended Head Start programs and full-day kindergartens), (b) *program intensity* (i.e., same teacher all day for all academic classes), and (c) *duration* (i.e., small class sizes from kindergarten through third grade). In addition, three other conditions ideal for this study was random assignment/heterogeneous groups, cohort effect, and program evaluation. Ideal class size would be 15 to 20 students in a class. Group 2 students would not have the same conditions and they would be assigned to classes with more than 20 students.

DESCRIPTION OF THE SCHOOL DISTRICTS

Three urban school districts from three counties (Macomb, Oakland, and Saginaw) were selected to participate in the study. Selection was based on districts that indicated they had student data that matched the research design that they would be willing to share with the researcher. The data presented in Table 2 were obtained from the SchoolMatters website and reflect school demographics in 2006. These school districts are described in detail.

Table 2

School Demographics by District

Demographic Characteristics	School District A	School District B	School District C
Students (N)	2,820	3,793	11,593
Spending Per Student			
District	\$9,839	\$9,619	\$9,111
State	\$8,625	\$8,625	\$8,625
County	\$8,437	\$9,754	\$8,211
Economically Disadvantaged (Qualify for free or reduced lunch)	70.0%	52.0%	74.0%
Students with Disabilities	19.6%	9.8%	21.5%
Racial/Ethnic Distribution			
African American	48.7%	91.0%	63.4%
White	46.5%	7.0%	22.4%
Hispanic	2.7%	.5%	13.0%
Other	2.1%	1.5%	1.2%

District C had the largest school population (n = 11,593 students) and had the lowest spending (\$9,111) among the three school districts. School District B, with a school population of 3,793 students, had the lowest percentage of students qualifying

for free or reduced lunch programs (52.0%). School District C had the highest percentage of students with disabilities (21.5%), with School District B having the fewest students with disabilities (9.8%). School District B had the highest percentage of African American students (91.0%) and School District C was the most racially diverse (African American [63.4%], White [22.4%], Hispanic [13.0%] and other [1.2%]). The racial distribution in School District A was almost equal with 48.7% African American students and 46.5% White students.

The academic outcomes presented on the SchoolMatters website for 2006 for the three school districts provide an overall view of how students have performed over time. Table 3 presents the percentage of students who either met or exceeded Michigan grade level content standards on the fourth-grade reading tests (MEAP).

Table 3

Fourth-Grade MEAP Reading Outcomes (2003 through 2006) for the Three School Districts (Percentage of Students Who either Met or Exceeded State Grade-Level Content Standards)

Year	State	School District A	School District B	School District C
2006	83.2	74.4	75.1	72.6
2005	82.4	70.6	74.7	74.3
2004	79.7	73.3	78.7	74.5
2003	74.5	67.4	65.2	68.8

None of the three districts that were included in the study exceeded the percentage of students who either met or exceeded state grade-level content standards statewide. Each of

the school districts showed general improvement in the percentage of students who passed the MEAP Reading tests from 2003 to 2006. However, in 2005, all three school districts showed decline, possibly due to a change in the timing of the test from the end of January to the first week of October. While School Districts A and B both showed improvement from 2005 to 2006, School District C experienced a slight decrease in the percentage of students who either met or exceeded state grade-level content standards.

DATA COLLECTION

The superintendents for the three school districts to be included in the study were contacted to arrange for data collection. The researcher met with the superintendents and/or their designees (e.g., assistant superintendents, elementary building principals, and clerical staff) from each school district to review the research proposal and establish procedures for data collection. The superintendents agreed to allow their data to be used in the study, with the stipulation that names and identifying information for the included students had to be eliminated.

Head Start programs were available in the three school districts, but they were administered separately by county agencies. It was assumed that participation in Head Start was included in the student records; however, in District B, the liaison for the school district in some cases had to obtain test information from student files to determine who had attended Head Start programs. Once this determination was made, the liaison had to search the students' records to obtain information on their assignment to full-day kindergarten and classes with less than 20 students for all years of study in the program. Student data for participation in Head Start was then collected for the years 1997-2001. Thereafter yearly

achievement scores in reading were sought from each district's IOWA, TERRA NOVA and MAT 8 reading tests in grades 1-3 and the 4th grade with MEAP reading scores sequentially for each student in the study. The final year when MEAP scores were collected was 2005-2006. Achievement test scores for each of these students were obtained for first, second, and third grade, as well as MEAP reading scores for fourth grade. To be included in the study, the students had to have attended the same district for all five years, and remain in the small class program, with data available for the five-year period. In District A the lists of Head Start attendees were provided for the district, but principals had to search individual student files to determine if the students were in the district in grades 1,2, and 3, and if so, they then could get achievement data to be coded for the researcher's use.

Over the next 18 months, as the researcher visited each district separately to work with the superintendent's designee(s), it became increasingly apparent that the data were not available. When data were collected, information from School District B and School District C was too sparse and too disconnected to be of value. Either student data could not be tracked across the five years in these school districts, or the students did not remain in the district for all years in the study. As a result, they were eliminated from the study.

Some specific problems associated with collecting data from the school districts included the following:

1. Students in these districts are generally transient, attending different schools across the five years. Because Michigan has both school of choice and charter schools legislation, students may attend one school in the district for kindergarten, another for first grade, and then return to the first school for second grade. The economic turmoil and high unemployment rates may cause parents to move frequently. As a result, the

- children are moved from schools either within the school district or across individual school districts in the School of Choice or Charter School programs. However, records were not maintained across the district(s) and consistent data were not available for the students for the five years.
2. It was planned that the results from School District B, with half-day kindergarten programs, would be compared with results from the other school districts with full- and half-day kindergarten programs. However, district data were sparse and not usable. Over the two years when data were being collected, changes in personnel occurred in all three school districts. These changes affected data collection as new personnel responsible for assisting the researcher had to be apprised of data collection procedures.
 3. During the data collection period, the researcher also found that much of the data were missing and/or unavailable either because of student mobility or data collection and storage issues.
 4. During the data collection period, funding for a pilot study of small-class sizes was available from the state of Michigan, but the class-size pilot was not administered consistently. Further problems with funding for these programs included decisions by the state not to allow Title I funds or funded resources to be used for small-class programs. At the present time (2007), Title II funds can no longer be used to reduce class size.
 5. During the course of the study (2002-2003), changes occurred at the state level in disseminating MEAP data. The MEAP data originally were scored by the Michigan Department of Education, then moved to the Michigan Treasury Department, and

- subsequently returned to the Michigan Department of Education. With these changes came differences in scoring, test development, timing of test administration, and dissemination of scores.
6. Appendix F provides additional data problems encountered in understanding small class programs.

RESULTS OF DATA ANALYSIS

Reading scores from the Metropolitan 8 Achievement Test for students who had attended full-day kindergarten and continued in small classes for grades 1, 2, and 3 were used to determine the extent to which a fade effect might be occurring between first, second, and third grades using descriptive statistics and then using nonparametric procedures. Results of these analyses are shown in Table 4. (See Appendix H for descriptive statistics for MAT 8 support for Table 4.)

Normal curve equivalent (NCE) scores for the three subscales and total reading improved for the study group across the three grade levels. The number of students who remained in the same school from kindergarten through fourth grade declined each year. This result was not unexpected given the low SES of persons in the school district and the transient nature of the families in the district. Data on some students who were in the same schools were missing because the children were not tested. Possible reasons for not testing a student were absences or the students' individual education plans (IEPs) indicated that they should not be tested.

Table 4

Descriptive Statistics: Metropolitan 8 Reading Test Normal Curve Equivalent (NCE) Scores for Grades 1, 2, and 3

Reading Test	First Grade			Second Grade			Third Grade		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Total	45	523.84	41.8	29	561.3	37.57	18	607.17	23.21
			8		8				
Sounds & Print	45	545.80	46.1	29	573.2	39.38	18	601.44	30.95
			0		1				
Vocabulary	45	493.02	85.7	29	558.0	41.61	18	616.39	28.00
			7		0				
Comprehension	45	526.67	41.3	29	558.9	44.14	18	606.11	29.35
			1		0				

To determine if the students were experiencing a test score fade, scores from first to second grade were compared using Wilcoxon signed ranks test. The results of these tests are presented in Table 5 (See Appendix I for Wilcoxin MAT 8 data support for Table 5).

Table 5

Wilcoxon Signed Ranks Test: Metropolitan 8 Reading Achievement Tests – First to Second Grade Comparisons

Reading Tests	N	1 st > 2 nd	1 st = 2 nd	1 st < 2 nd	Z Value	Sig
Total	21	1	1	19	-3.85	<.001
Sounds & Print	21	2	1	18	-3.25	.001
Vocabulary	21	0	1	20	-3.92	<.001
Comprehension	21	3	1	17	-3.06	.002

The results of the Wilcoxon signed ranks analysis support the hypothesis in that the students improved in all aspects of reading as measured by the Metropolitan 8 (MAT) reading achievement tests. Most of the students (n = 19) improved on the total score, $Z = -3.85$, $p < .001$. Eighteen students showed positive improvement on the sounds and print subtest, $Z = -3.25$, $p = .001$. The improvement of 20 (of 21) students in reading vocabulary was statistically significant, $Z = -3.92$, $p < .001$, while 17 students had higher scores for comprehension, $Z = -3.06$, $p = <.002$. These findings suggest that the Head Start fade was not occurring from first to second grade.

To determine the extent to which a fade effect was occurring from second to third grade, the results of a Wilcoxon signed ranks test for the students from Grade 2 to Grade 3 were examined. Table 6 presents results of this analysis (See Appendix J for Wilcoxin MAT 8 data support for Table 6).

Table 6

Wilcoxon Signed Ranks Test: Metropolitan 8 Reading Achievement Tests – Second to Third Grade Comparisons

Reading Tests	N	1 st > 2 nd	1 st = 2 nd	1 st < 2 nd	Z Value	Sig
Total	12	0	12	0	-3.06	.002
Sounds & Print	12	1	11	0	-2.59	.010
Vocabulary	12	1	11	0	-2.90	.004
Comprehension	12	1	11	0	-2.98	.003

The results of the Wilcoxon signed ranks test indicated that 12 students who had reading scores from second to third grade showed significant improvement in total reading scores, $Z = -3.06$, $p = .002$. Eleven students improved significantly in sounds and print from second to third grade, $Z = -2.59$, $p < .010$, with a similar number of students showing positive outcomes in vocabulary, $Z = -2.90$, $p = .004$ and comprehension, $Z = -2.98$, $p = .003$. Based on these findings, it appears that Head Start students who had been in full-day kindergarten classes and small-class sizes in first, second, and third grades were not exhibiting Head Start Fade at the end of third grade.

To examine if full-day kindergarten and small class sizes in early elementary (Grades 1, 2, and 3), the students' fourth grade Michigan Education Assessment Program (MEAP) scores were compared to 400, the cut score at which students were considered to have either exceeded or met the Michigan standards. The results of the t-tests for the single sample

available for this comparison are presented in Table 7. (See Appendix K for MEAP t-test data support for Table 7).

Table 7

T-Test for One-Sample: Michigan Education Assessment Program (MEAP) Test Scores – Fourth-Grade Reading

MEAP	N	Mean	SD	DF	t-Value	Sig of t
Reading Scores	16	389.44	46.44	15	-.91	.377 (ns)

Comparison of the mean score of 389.44 (sd = 46.44) for the group sample in this study with the cut score of 400 considered passing the MEAP reading test was not statistically significant, $t(15) = -.91$, $p = .377$. The fourth-grade students who met all conditions for inclusion in Group 1 were not differing significantly from the point on the MEAP at which Grade 4 students were considered to be reading at their grade levels. This finding supports that students in the fourth grade were not experiencing a measurable Head Start fade, as suggested by the theoretic consideration and issues generated and explicated in Chapter 2 of this study.

The percent of students who either met or exceeded the state grade-level standards for the MEAP reading test in District A for 2003 through 2006 are compared to the state and county results. Table 8 presents results of this analysis. (See Appendix L for MEAP data support group 1 for Table 8).

Table 8

Descriptive Statistics of District, County, and State Comparisons of Fourth-Grade MEAP Reading Outcomes (2003 through 2006) for School District A (Percentage of Students Who either Met or Exceeded State Grade-Level Content Standards and District A Group 1

Year	State	County	District A	District A Group 1
2003	74.5	79.8	67.4	—
2004	79.7	82.0	73.3	—
2005	82.4	86.0	70.6	—
2006	83.2	87.0	74.4	62.5
Change from 2003 to 2006	+8.7	+7.2	+7.0	—

The descriptive statistics described in Table 8 show the percentage of students, 74.4%, who either met or exceeded the state grade-level standards for reading. In the targeted Group A, the percentage was lower than the district or state percentages. Although the scores for District A were lower than the county results, their scores showed consistent improvement, with the exception of 2005. In this urban district, all K–5 classes have only 17–19 pupils per classroom, which helps explain the larger percentage of students scoring higher on the MEAP test. The percentages of students in the county who either met or exceeded state grade-level standards for reading were generally three or four points higher than the state average. Although the scores for School District A, Group I students were

lower than the county results, their scores showed consistent improvement (except for 2005), with the county four year gain of 7.2 and Group I gain of 7.0.

The students in subgroup A are the most economically challenged in this district. The minor fade in performance is not significant and it supports the hypothesis that the fade effect from the small-class treatment will minimize or decrease the fade effect.

There are a number of reasons that explain the results. First, students in the final year were tested on the MEAP during the fall when the children returned from summer break. This represented a change from earlier years when students had later testing dates. The earlier testing date in the fall does not take into account summer decrement, or the loss of learning young children experience after three months away from school over the summer. There was also a much smaller representative sample size. Finally, the districts with high levels of special education referrals (19.6%) and low socioeconomic factors represented very challenged children.

QUALITATIVE DATA: DEVELOPMENT OF A THEORY TO SUPPORT REDUCING THE HEAD START FADE

A major component of this study was to test a theory, based upon the research and literature, to determine if the effect of a comprehensive, coherent, Pre-K through grade 3 program for young children would mitigate or eliminate the Head Start “fade effect.” Evidence that supports this theory, when viewed in the research literature, presents data that showed an unintentional relationship existed between program elements in the Pre-K and Head Start programs such as the CPC, and PCDC. The theory tested in the Abecedarian program of duration, intensity and early intervention were elements found in some small

class studies. Later these elements described below proved to be implemented as the Tennessee Student Teacher Achievement Ratio (STAR) (Word, et al., 1990) empirical research experiment.

As such, the purpose of the present study was to investigate if the application of the theoretic position of Ramey and Ramey (1989) in the Abecedarian study, then later advanced as the theoretic construct tested in the STAR study, demonstrated a connection in the current research that leads to a reduction in the fade phenomenon.

Research on cognitive development has indicated that the first five years of a child's life are the most important with regard to language and social development. Bloom (1964) forwarded the importance of the principle of developmental learning and his theories indicated the importance of starting early when the rate of growth is strongest for children in educational programs. Piaget's research also indicated the importance of the first five years as pivotal in a child's development. Ramey and Ramey (1998) created a framework for early interventions and articulated principles of efficacy. The three major conditions of the work of Ramey and Ramey (1998) are as follows: (a) *early intervention*—when the student starts school; (b) *duration*—provide small classes for three, preferably four, years so the child can learn about school (an apprentice for years of successful “work”); and (c) *intensity*—maintain the small class all day, every day. Other elements that research has shown to significantly contribute to the gains that are part of this theory-testing design are as follows: (a) parent involvement, (b) transition programs (Pre-K to grades K-1), (c) random assignment or heterogeneity, and (d) the cohort effect (See Appendix C). The principles are built on the concept that “fragmented, weak efforts in early interventions are not likely to succeed,

whereas intensive, high quality, ecologically persuasive interventions can and do” (Ramey & Ramey, 1998, p.109).

The STAR (Word et. al., 1990) study provided the SBR evidence to support the theoretic positions advanced by Ramey and Ramey (1989) in the Abecedarian study. Work reported by Finn and Achilles (1999), Finn, Gerber, Achilles and Boyd-Zaharias (2001), Krueger and Whitmore (2000), Nye, Hedges, and Konstantopoulos (1999), and others have shown that for the early achievement benefits to show enduring effects for subsequent years in school, some conditions must be met.

The construct for the theory testing implemented in this study was to identify if the theoretic position of Ramey and Ramey (1989), which later became one theoretic base of the STAR study (Word et al., 1990), when applied to Head Start, the early intervention, will show that the enduring effects of duration and intensity will decrease or eliminate the Head Start fade when implemented correctly.

An important requirement for high quality interventions is that the effect of the early intervention last beyond the treatment period. Pre-K and Head Start programs with longitudinal data that indicated programs that have met this requirement are the Abecedarian Program in North Carolina; the Chicago Child & Parent Center (CPC) in Illinois; and the Perry Pre-School Program in Michigan. Each program’s success was created as a result of the inclusion, to varying degrees, of the three characteristics: early intervention, duration, and intensity. Evaluations of these programs demonstrate the benefit of each indicator in the program design that strengthened the theory tested in the current study. These evaluations showed further benefits, including: students who attended the pre-K/Head Start programs experienced decreased numbers of grade retention and special education referrals. All

programs provided some evidence of transition activities or programs that involved parents and all participants experienced continued gains in achievement. Participants in the treatment group showed significantly higher scores on cognitive assessments in reading and math from primary grades through adulthood than did participants in the control group. By age 21, students who had participated in preschools with substantial funding from 1978 to 1998 experienced positive impacts in improving developmental competence in a variety of domains, including school attendance and performance as well as reduced subsequent grade retention.

Class Size

Class-size research on students in grades K-3 with similar results to those obtained in the Perry Child Development Center (PCDC), the Abecedarian program, and the Chicago Parent and Child Center (CPC) provide strong examples of early intervention and help educators gain greater understanding of the Head Start fade effect. The continuing longitudinal research on small-class size has provided further direction in testing the theory in an effort to eliminate or moderate the Head Start fade. The following presents a composite of the research to date on the importance of small classes in effecting positive gains in learning for all students, and, particularly, for at-risk and minority youth.

Results of many studies have been published on the effects of small-class sizes on student achievement where one teacher is assigned to 14-18 students per classroom. Some of the studies provide evidence that small-class programs have positive effects on student outcomes throughout elementary, middle, and high school, as well as through college. Table 9 presents a summary of studies that have supported the “enduring effects” and other indicated outcomes of small classes in early grades, Pre-K through grade three or four.

Table 9

Theory Development: Examples of Enduring Effects of Small-Class Sizes

Theory	Source	Comments
Small classes (15 to 17 students) and one teacher	HEAD START and Pre-K Programs: Head Start (1965) Perry Child Development Center/ Perry Pre-School (PCDC), (ages 3-4 years)	Small classes are not pupil-teacher ratio (PTR), but a different way of organizing and delivering instruction. PTR just puts another adult in the classroom, usually an aide, who conducts instruction in the same way as a teacher.
	Chicago Parent and Child Center (CPC) (ages 3-9 years) N. Carolina's Abecedarian Program (ages 0-5 years)	Small classes are good for all children. STAR (Word, et al.) results demonstrated increased gains accrued for minority and at-risk children and that all children advanced in learning to greater degrees than children in large classes. Barnett (2003) reported the Perry Pre-K program had higher gains shown by African-American males.
	K-3 PROGRAMS: Texas HB 72 (1984) DuPont Study (1984) STAR (1985-1990) Indiana Prime Time (1986) Lasting Benefits Study (1989-1996) N.C. Burke County (1991) Challenge (1989-1995)	

Table 9 continues

Table 9 (cont.)

(Program Intensity Same teacher every day, all day for care academics)	CPC Abecedarian DuPont Study (1984) STAR (1985) Burke County, N.C. (K-3; 1993) Kizer-Kling (1993) Challenge (1995) Wisconsin's Student Guarantee in Education (SAGE) (1996)	Program intensity was advanced by Ramey and Ramey (1998) in the Abecedarian program and later in STAR (Word et. al., 1990). Kizer-Kling's (1993) Success Starts Small and Project Challenge (1993) results clearly supported gains in learning similar to those reported in the STAR (Word et al., 1990) empirical study.
Early Intervention (When the child starts school)	Head Start (1965) PCDC, CPC Abecedarian Program Student Achievement Guarantee in Education (SAGE) Michigan School Readiness Program, (age 4), (MSRP, 1999)	Because Head Start does not enroll all eligible students, the Michigan MSRP program attempted to enroll some of the children not admitted to Head Start.
Duration after Pre-K (All day K and 3-4 years in small-class programs)	DuPont STAR Burke County, N.C. Success Starts Small (SSS)-Kizer-Kling (1993) SAGE Abecedarian Program Draper Elementary School, S.C. STAR Project Challenge	The Abecedarian program, STAR (Word et al., 1990) and Challenge were examples of 3 to 4 years of program duration with stronger effects that occurred after the third year in the program. Abecedarian began at 4.4 months on average and the other two programs were kindergarten through third grade.
Transition Programs: Pre-K, K through 3	CPC Abecedarian PCDC MSRP-Michigan STAR SAGE	All preschool programs offered services to children through at least third grade. CPC retained students in program ages 3 to 9 years. CPC and Perry Preschool provided tutors for students in public school programs K-3. MSRP is delivered in MI Schools and provides transition programs to K-3.

Table 9 continues

Table 9 (cont.)

Parent Involvement	Head Start Abecedarian CPC, STAR Burke County Kizer-Kling (SSS) date deleted SAGE	All programs found parents become more involved in small class programs in the early years. They tend to remain more involved in school through elementary school.
Random Assignment/ Heterogeneity	STAR PCDC Head Start Burke County	In Project STAR, Boyd-Zaharias, Achilles et al. followed through research on the positive effects of heterogeneity. These findings were further supported by Mosteller's meta-analysis on the importance of random assignment.
Cohort Effect Students travel with their class through grades K-3	CPC STAR SAGE	Students benefit academically from what Saranson calls a psychological sense of community (PSOC). At-risk and minority students had better students outcomes from the cohort effect on assessments for self-concept in the study.

Note. Chapter 2 contains further definition and examples of the conditions and programs.

The results of the present study support the theory-testing component of the study for positive program effects or increases in achievement. These results confirm that there should be no fade effect, or a minimal one, when the eight elements of the theory are present in a school on a consistent basis. When students have positive educational experiences that result from early interventions (participation in Head Start and small classes from kindergarten through grade 3), the positive effects achieved from Head Start can be enduring and not expected to fade over time. While these programs should be positive for all children, program gains have been found to be stronger for African-American children and other children considered to be at risk for school problems or failure.

SUMMARY

The results of the data analysis provided support for the diminution of, or lack of, a Head Start fade for students who had attended Head Start, full-day kindergarten, and been in small-class sizes for first, second, and third grade and regularly experienced the other services. Conclusions and recommendations based on these findings are presented in Chapter 5.

CHAPTER 5: SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH

Empirical research has supported the existence of the Head Start Fade despite years of effective pre-K, Head Start, and class-size studies that have offered alternative theories and evidence to negate the fade effect. Because this study did not lend itself to direct empirical work, a dual strategy of pre-K to grade three retrospective data analysis and theory testing was implemented. This small study was conducted to evolve a theory based upon prior work in pre-K and small class K–3 programs and then pilot test the theory with a small sample of data from Head Start pupils who attended small K-3 classes in public schools. One purpose of the study was to see if the evolved theory would mitigate or moderate the Head Start fade effect. The study did support the efficacy of the theory, and although a small fade effect was noted, gains continued to increase K-3 on the Metropolitan Achievement Test (MAT) with no statistically significant fade on the Michigan Educational Assessment Program (MEAP) evidenced over the four years. The results also tracked the importance of preschool programs for children similar to the Abecedarian, Perry Child Development Center (PCDC), and Chicago Parent Center (CPC) programs.

Chapter 5 begins with a statement of the issue addressed in the study, a summary of the literature on successful pre-K programs that parallel with small-class size research, the methodology and hypotheses of this study, and the findings of this study and how they compare with the literature and prior research. The chapter ends with a conclusion and recommendations. Recommendations are presented in three areas: (a) policy, (b) practice, and (c) advancing the theory in future program research.

SUMMARY

After an exhaustive review of the literature on successful early education programs and research on class size, the most significant concepts became apparent. They are presented below.

Summary of Successful Early Education Programs and Class-Size Research

Six elements for a successful early education program (See Appendix C) were identified or verified in the 1985-2007 Tennessee Student Teacher Achievement Ratio (STAR) experiment and related studies, the largest, longest, and most well respected empirical class-size study to date. The six elements are as follows: (a) early intervention, (b) duration, (c) intensity, (d) random assignment or heterogeneity, (e) cohort effect, and (f) strong evaluations (e.g., Word et al., 1990; Achilles, 1999). Two other elements that were present include parent involvement and transition. The three most successful early education programs, the Abecedarian Program, the CPC, and the PCDC, exhibited five of the six elements: (a) early intervention, (b) duration, (c) intensity, (d) transition services, and (e) parent involvement. This lends support to the theory that these elements are foundational to the design of any early education program. In terms of early intervention, the Abecedarian program mirrors Bloom's (1964) research on starting early to maximize developmental gains by enrolling children ages four months to five years, though the other programs identify similar evidence. Each program had program duration of two or more years.

Additional Elements of Successful Pre-K Programs

The national Department of Health and Human Services (2001, 2003) and Barnett (2003) have provided the educational community with research on effective pre-K and Head Start programs that have successfully closed the developmental learning gap impoverished

children face when entering school. The three programs identified in those studies in which there was little to no fade effect included the previously mentioned North Carolina Abecedarian Program for children ages four months to five years, the Michigan Perry Child Development Center (PCDC) for three and four year olds, and Illinois' Chicago Child and Parent Center (CPC) for three to nine years olds. All three, as well as other STAR experiment programs, had robust longitudinal gains through the teen years and longer. The PCDC gains continue to be monitored through adulthood by Ypsilanti, Michigan's well-respected High Scope Foundation.

In addition to the already mentioned developmentally appropriate strategies of early intervention, duration, and intensity, these programs employed other strategies that contributed to their success. These include the following:

- *Certified teachers:* All the daily core-learning activities were taught by certified teachers.
- *Small-class sizes:* Each program had small-class sizes of 10–15 or fewer students per classroom.
- *Parent involvement:* Since parent education and participation was required by Head Start, all three included this element, and in some cases, required it, for students to be eligible for enrollment.
- *Grade-level transition programs:* The Abecedarian program and the Perry program provided tutoring and support for students when they entered K-3 programs. CPC offered elementary school services.

- *Random grouping and/or heterogeneity:* Random grouping and/or heterogeneity was found in PCDC and was required for any component of a Head Start funded program. Federal funding requires that 10% of participants not be at-risk.
- *Random assignment and heterogeneity:* Random assignment was part of the design of STAR and companion studies (Word et al., 1990; Zaharias et al., 1995).
- *At-risk children:* All three programs enrolled at-risk, low SES children, and all programs created stronger gains for at-risk youth. In the case of PCDC, Barnett found gains were strongest for black males (Barnett, 2003).
- *Strong evaluations:* Finally, each program was tied to strong program evaluations that have been done at various times in the participants' lives. This is also one of STAR's six elements of success.

Alignment of Class-Size Research to Elements of Successful Pre-K Programs

In a review of decades of research on small classes, the STAR experiment has provided the most comprehensive database for subsequent research. Researchers using STAR data have developed multiple forms of formative and summative research to create ongoing analyses of STAR's original sample of 11,601 students. From 1985 to 1989, while STAR was still in operation, new findings emerged from the ongoing research based on STAR data.

Based on the STAR data, other significant studies have replicated important elements found in successful early education programs. For example, Boyd-Zaharias (1993) and Zaharias et al. (1995) completed a study and reported on random assignment effects in STAR in 1995, and Finn et al. (2001) examined the decrease in behavior issues that could be directly attributed to the benefits of a psychological sense of community (PSOC), increased self-concept, and achievement gains.

STAR Outcomes

For the purposes of the present study, two significant outcomes of the STAR experiment were as follows: (a) early intervention prevented an achievement gap and (b) there was no fade effect. The Head Start fade effect had been identified by researchers using STAR data in grades K-3. STAR and companion studies tracked students' progress in those grades with multiple assessments, and no fade was evident. In addition, STAR demonstrated that students in small classes had stronger gains than their peers did in the other two study groups: (a) 22-25 students with a teacher and (b) 22-25 students with an aide and a teacher. Samples were diverse, and STAR showed robust gains for minority students, similar to the PCDC. Additional evidence of strong outcomes included a later decrease in special education referrals and reductions in grade-level retention. These results continued in STAR follow-up studies such as the Lasting Benefits Study (LBS), Challenge, and Enduring Effects. All STAR companion studies such as the DuPont Study, SAGE, Challenge, and Burke County mirrored the STAR results.

Longitudinal Gains

Researchers doing longitudinal studies with STAR data reported that more students who had been enrolled in the STAR experiment took advanced coursework in high school, took college placement tests such as the SAT and ACT, and enrolled in college (Krueger & Whitmore, 2001; Krueger et al., 2000). In addition, students in all of STAR's pre-K and small-class companion studies showed evidence of increased high-school enrollment and graduation and gains in student behavior, as well as a decrease in involvement in crime. A recent study by Levin (Levin, Belfield, Muennig, & Rouse, 2007) highlighted five programs

that raise the graduation rate from high school demonstrably: (a) “First Things First, (b) CPC, (c) Teacher Salary Increase, (d) PCDC, and (e) Class Size Reduction (CSR)” (p. 4).

Summary of the Purpose, Design, Method, Hypothesis, and Outcome of This Study

In this study, once a theory was proposed, the intent was to examine available achievement data, retrospectively, for evidence of a fade effect or no fade effect with two subgroups in three districts in Michigan. School districts in which children had experienced Head Start as a proxy for early intervention were selected to ensure access to Head Start data. The selected districts also indicated that they had access to achievement data for grades 1–3 from connected public schools to examine trends in achievement scores, both aggregated and disaggregated for SES, race, and gender. Most districts do not have a direct connection from Head Start to the public schools. The researcher hoped that data would subsequently assist educators in discovering correlates that contributed to the Head Start Fade, and by understanding it, employ program strategies from successful empirical, longitudinal studies that demonstrated no fade effect. If possible, long-term gains would be that young children, irrespective of race, gender, or socioeconomic status (SES) would have an equal place at the starting gate for schooling.

This researcher used a nonexperimental, longitudinal, retrospective explanatory design (Johnson, 2001) to structure the pilot study and test a theory. The duration of the study was seven years, from 1999-2006, and included scores from hundreds of students at the outset but concluded with only 12 students for whom norm-referenced test (NRT) outcomes and fourth-grade MEAP scores were available.

The method is described as follows: Two groups of children (Group I and Group II) were tracked for academic achievement. The Group I children were enrolled in small Head

Start classes and subsequently attended small, full-day kindergarten classes and small classes for grades one, two, and three. The Group II children may not have been enrolled in small Head Start classes, may have attended full-day kindergarten classes, and did not have small classes for grades one, two, and three. However, because of the unavailability of data, the actual method followed was to examine grade two- and/or grade-three outcomes for students who had experienced Head Start and small classes grades K-3 of 15 to 17. The researcher examined the relationship between the factors of duration and intensity and Head Start fade as revealed by achievement outcomes and scores on the fourth-grade MEAP test.

Outcome of the Study

Although the cumulative data from the study over four years were small, they tended to support the study's actual hypothesis. Group I Head Start students who experienced full-day kindergarten everyday in small classes and continued in small classes through grades one, two, and sometimes three, did not demonstrate a statistically significant Head Start fade effect when compared to Group II on state and county scores of the fourth-grade MEAP tests.

Similar to students in the Perry program, a minor fade effect over the four years occurred in a random pattern and did not reflect a consistent decline in subsequent performance as has been noted in most Head Start programs nationally. The students in this study did not experience a consistent fade effect from grade level to grade level tests, nor did MEAP data reflect a statistically significant fade effect. Although grade level to grade level was not part of the original hypothesis, it is worth noting this result because it tracks with data from the three previously mentioned pre-K programs, STAR, and companion studies. In this study, students continued to experience growth from grade level to grade level, no fade effect was

noted, nor was there a statistically significant fade effect on the MEAP results. Thus, the data supported the hypothesis.

An explanation for this minor fade over the four years can in part be attributed to summer decrement. Beginning in 2006, the MEAP test was given the first month of school rather than in mid-year. For impoverished children not participating in summer programs, it is probable that a loss of learning occurred, which likely affected gains on the MEAP.

This study employed the same three elements of early intervention, duration, and intensity, as did the Abecedarian program, the CPC, and the PCDC. For example, the students in this study started at age four, had small classes for six years, and had the same teacher all day, every day, for core-content classes. Further, since random grouping is required in every Head Start program, the students were randomly assigned in their Head Start programs although there is no conclusive evidence of heterogeneous grouping for grades K–4. Other outcomes were not able to be included because of the lack of sufficient data from the school districts. Additionally, school districts that received money for small classes had developmentally appropriate curriculum for their reading programs.

Unexpected Limitations Encountered During the Research Process

There were significant problems encountered throughout the study because when superintendents were asked to participate, they said they had the data required. However, as the research progressed, it became increasingly evident that the data would not be available. Either the data had not been retained or student mobility had made the data ineffective. Contributing factors to this mobility included parents opting for schools of choice and charter schools, as well as a struggling urban economy that required low-SES families to move in

order to secure work. A similar problem was found in Ed Becker's 2006 dissertation as he collected data in Michigan.

Another issue affecting longitudinal data collection was that the location of much of the data kept changing. In 2002, the Michigan Department of Education (MDE) had their archives moved to the Michigan Treasury and then back to MDE a year or two later. In the process, the Intermediate School Districts stopped storing data for county districts in their jurisdiction; this process required districts to keep all data.

Another issue that impacted the integrity of the data had to do with the constant shifting of staff in these offices. Cuts to personnel and an increased workload for employees who were retained resulted from a lack of state funding. As a result, data were not maintained.

CONCLUSIONS

The subgroup was too small to have further implications; however, the pilot results inform us that the study design was successful, the students were available ($n = 12$), and the secondary purpose of the study, creating a robust design that could be replicated nationally by other researchers, was indeed accomplished. The strength of this study, therefore, lies in the development of a theory in the carefully designed, thought-out research process that is further supported by the empirical work of STAR and other pre-K and small-class studies. The evidence presented in these findings have shown promise toward the goal of eliminating the Head Start fade effect and need to be tested with a broader audience.

RECOMMENDATIONS

This recommendation section is divided into three parts. These parts include the following: (a) policy, (b) practice, and (c) advancing the theory in future program research.

Policy

Texas was the first statewide small-class study to make the distinction in the legislation (H. B. 72, 1984) between class size and pupil-teacher ratio (PTR) until STAR (Word et al., 1990). Usually, increasing PTR means adding an aide to the classroom to assist the teacher. This does not decrease class size, but merely add another adult to the classroom. The findings of this research concur with H. B. 72 that small-class sizes with one teacher are more effective than large classes with a teacher and an aide. The addition of an aide merely places another adult in the room but does not increase the quality of education. It is the recommendation of this researcher that every state adopt a bill similar to Texas H. B. 72 bill and that they all implement policies that will ensure small-class sizes for all early education programs that, minimally, are founded upon the three principles of early intervention, duration, and intensity.

Practice

Unfortunately, data were not forthcoming as agreed upon by school districts; this experience provided some important lessons. Thus, there are specific recommendations for future researchers who would like to replicate this study:

1. Researchers should secure data at the beginning of the study to determine if the district qualifies for participation.
2. Once other retrospective studies have been analyzed, an empirical study should be conducted.

3. Evidence of results should reach legislative channels to assist in supporting universal pre-K for any family desiring it for their children.
4. Evidence of high-school and college gains should be presented to state and national professional associations like the American Association of School Administrators (AASA) and the National Association of Secondary School Principals (NASSP).
5. A series of required benchmark collection dates over a short span of time should be set in cooperation with the districts. (The researcher was repeatedly sent to delegated staff in all three districts to collect data. This was inefficient and time consuming and made it extremely difficult to maintain continuous communication with districts.)
6. Get prior knowledge as to the efficiency of state databases when undertaking longitudinal research.

Advancing the Theory in Future Program Research

If we are to be true to the educational needs of future generations, particularly to those of the poorest in our nation, we first need to determine our priorities. “The highest priority in education should be preparing the very young children from poor families for school” (Sirin, 2005, p. 4). The findings from this study are extremely promising and many others indisputably support the clear and present benefits of small-class sizes for Head Start through fourth-grade students. As stated by Achilles (1999), small classes offer the following:

Three socially desirable benefits that are expressed American values Improved education = QE2: Small classes offer quality, equity, [and] equality. They can also be the basis for value-driven education leadership *Quality*: achievement, behavior, and citizenship. *Equality*: All participants get the same treatment; no group gets more

or less, and *Equity*: Minority and hard-to-teach youngsters benefit more from small classes than do better students, but all students benefit. (pp. 158–159)

Future studies that test the evolved theory advanced in this small study should provide increased evidence for the importance of small classes for young children but especially for those who seriously need the opportunity to learn and succeed.

REFERENCES

- Achilles, C. M. (1999). *Let's put kids first, finally: Getting class size right*. Thousand Oaks, CA: Corwin Press.
- Achilles, C. M., & Finn J. D. (2002a, February.) *Making sense of continuing and renewed class-size findings and interest*. Paper presented at the annual meeting of the American Association of School Administrators, San Diego, CA. (ERIC Document Reproduction Services No. ED464393)
- Achilles, C. M., & Finn, J. D. (2002b, November). *Constructing "common sense" from the class-size research, theory, and use. The strange gloom and doom scenario of class-size implementation*. Symposium conducted at the meeting of the Annual Mid-South Education Research Association, Nashville, TN.
- Achilles, C. M., Kizer-Kling, K., Owen, J., & Aust, A. (1994). *Success starts small: Life in a small class*. Greensboro, NC: University of North Carolina. (ERIC Document Reproduction Services No. ED419 288).
- Achilles, C. M., & Nye, B. A. (1995-1996). Let's do it right at the start. *National Forum of Educational Administration and Supervision Journal*, 15(1), 6.
- Achilles, C. M., Nye, B. A., & Bain, H. P. (1994–1995). The test score "value" of kindergarten for pupils in the three class conditions of grades 1, 2, and 3. *National Forum of Educational Administration and Supervision Journal*, 12, 25–32.
- Achilles, C. M., Nye, B. A., Zaharias, J. B., & Fulton, B. D. (1993). *The lasting benefits study (LBS) in grades 4 and 5: A legacy from the Tennessee's four year (K–3) class-size study. (1985–89) Project STAR*. Paper presented at the annual meeting of the

- North Carolina Association for Research in Education, Greensboro. (ERIC Document Reproduction Services No. ED356559)
- Achilles, C. M., Zaharias, J. B., Nye, B. A., & Fulton, D. (1995). *Analysis of policy application of experimental results: Project Challenge*. Paper presented at the annual meeting of the American Educational Research Association, Biloxi, MS. (ERIC Document Reproduction Service No. ED393151)
- Aleman, A. M. (2006, October). Latino demographics, democratic individuality, and educational accountability: A pragmatist's view. *American Educational Researcher Journal*, 35(7), 25-31.
- Barnett, S. (1993). Does head start fade out? *Education Week*, 5, 40.
- Barnett, S. (1995, Winter). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, 5(3), 25–50.
- Barnett, S., & Hustedt, J. T. (2003). Pre-school: The most important grade. *Educational Leadership*, 60(7), 54–57.
- Biddle, B. J., & Berliner, D. C. (2002). Small-class size and its effects. *Educational Leadership*, 59(5), 12–23.
- Blatchford, P., Russell, A., Bassett, P., Brown, P., & Martin, C. (2003). *The role and effects of teaching assistants in English primary school years (406) 2000–2003: Results from the class-size and pupil-adult ratios*. University of London, Institute of Education.
- Bloom, B. S. (1964). *Stability and change in human characteristics*. New York: Wiley.
- Bloom, B. S. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-on-one tutoring. *Educational Researcher*, 13(6), 4–16.

- Boyd-Zaharias, J. (1993). *The effect of random class assignment on elementary students' reading and mathematics achievement*. Unpublished doctoral dissertation, Tennessee State University, Nashville, TN.
- Bracey, G. (1999). Poverty and achievement. *Phi Delta Kappan*, 77(7), 510–512.
- Burris, C. C., Herbert, J.P., & Levin, H. M.(2006). Accelerating mathematics achievement using heterogenous grouping. *American Educational Research Journal*, 43(1), 105–136.
- Cawelti, G. (1999). Lessons from research that changed education. *Educational Leadership*, 60(5), 18–21.
- Chase, C. I., Mueller, D. J., & Walden, J. D. (1986, December). *Prime Time: Its impact on instruction and achievement*. Indianapolis, IN: Indiana Department of Education.
- Cooley, W. W. (1993, May 12). *The difficulty of the educational task*. Pennsylvania Educational Policy Studies, University of Pittsburg, Pennsylvania Department of Education Policy Studies. Policy Paper No. 16, 1–15.
- Copple, C. W., Cline, M. G., & Smith, A. N. (1987). *Path to the future: Long-term effects of Head Start in the Philadelphia school district*. Washington, DC: U.S. Department of Health and Human Services.
- Darling-Hammond, L. (1998). Teachers and teaching: Testing policy hypotheses from a national commission report. *Educational Researcher*, 27(1) 5–15.
- Edmondson, B. P. (2005). *The effects of parental involvement and small- class sizes on the academic achievement gap*. Unpublished doctoral dissertation, Eastern Michigan University, Ypsilanti.

- Egelson, P., Harman, P., Hood, A., & Achilles, C. M. (2003). *State-level class-size reduction initiatives: How class size makes a difference*, (SERVE Education Laboratory ED-01-CO-0015) National Institute for Education Sciences, U.S. Dept of Education.
- Entwisle, D. R., & Astonem, N. M. (1994). Some practical guidelines for measuring youth's race/ethnicity and socioeconomic status. *Child Development*, 65(6), 1521–1540.
- Finn, J. D. (2006). *The adult lives of at-risk students, the roles of attainment and engagement in high school* (NCES No. 2006-328, pp. 1–15). Washington, DC: U.S. Department of Education.
- Finn, J. D., & Achilles, C. M. (1999). Tennessee class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, 21(2), 97–109.
- Finn, J. D., Pannoizzo, G. & Achilles, C. M. (2003, Fall). The “whys” of class size: Student behavior in small classes. *Review of Educational Research*, 73(3), 321–368.
- Finn, J. D., Fox, J. D., McClellan, M., Achilles, C. M., & Boyd-Zaharias, J. (2006). Small classes in the early grades and course taking in high school. [Electronic version.] *International Journal of Education Policy and Leadership*. 1(10), 1–13.
- Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2001, April). The enduring effects of small classes. *Teachers College Record* 103(2), 145–183.
- Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology*, 97(2), 214–223.
- Fullan, M. (2001). *Leading in a culture of change*. San Francisco: Jossey-Bass.
- Gall, J. P., Gall, M. D., & Borg, W. R. (1999). *Applying educational research: A practical guide* (4th ed.). New York: Addison Wesley Longman.

- Gerber, S. B., Finn, J. D., Achilles, C. M., & Boyd-Zaharias, J. (2001). Teacher aides and students' academic achievement. *Educational Evaluation and Policy Analysis*, 23(2), 123–143.
- Glass, G. V., Cahen, L. S., Smith, M. L., & Filby, N. (1982). *School class size: Research and policy*. Beverly Hills, CA: SAGE.
- Glass, G. V., & Smith, M. L. (1978). *Meta-analysis of research of the relationships of class size and achievement*. San Francisco: Far West Laboratory for Educational Research and Development.
- Haller, E. J., & Kleine, P. F. (2001). *Research and practice: The case of E. E. O. using educational research: A school administrator's guide*. New York: Longman.
- Haskins, R., & Rouse, C. (2005). *Closing the achievement gap*. Princeton, NJ: Princeton University and Brookings Institute, Woodrow Wilson School of Public and International Affairs.
- Hawking, S. (2006, May) Viral Marketing. *Harper's* 312(1,872), 321.
- Heckman, J. J., (2006, January 10) Catch 'em young. *The Wall Street Journal*, p. A14.
- Heifetz, R. (1994). *Leadership without easy answers*. Cambridge, MA: Harvard University Press.
- Hodgkinson, H. (1992, June). *A demographic look at tomorrow*. Washington, D.C.: Institute for Educational Leadership.
- Hodgkinson, H. (2003). *Leaving too many children behind*. Washington, DC: Institute for Educational Leadership.
- Jacobson, L. (2007). For Head Start, a marathon run. *Education Week*, 26(34), 31-33.

- Johnson, B. (2001). Toward a new classification of non-experimental quantitative research. *Educational Researcher*, 30(2), 3–13.
- Keedy, J. L., & Achilles, C. M. (1996). The need for school constructed theories in practice in U.S. school restructuring. *Journal of Educational Administration*, 35(2), 102–120.
- Kelly, A. E., & Yin, R. K., (2007). Strengthening structured abstracts for education research: The need for claim-based structural abstracts. *Educational Researcher*, 36(3), 131–138.
- Kluge, P. M. (2006, December). *The initial report of the Start Strong Council: Executive*, 1-23. Retrieved April 10, 2007 from: [www.education.virginia.gov/Initiatives/Start Strong/index.cfm](http://www.education.virginia.gov/Initiatives/StartStrong/index.cfm)
- Kohn, A. (2004). Challenging students and how to have more of them. *Phi Delta Kappan*, 86(3), 186–196.
- Kozol, J. (1992). *Savage inequalities*. New York: Harper Collins.
- Krueger, A. B., & Whitmore, D. (2001). The effect of attending a small class in the early grades on college test-taking and middle-school test results: Results from Project STAR. *Economic Journal*, 111, 1–28.
- Lazar, I., & Darlington, R. (1982). *Lasting effects of early education: A report from the consortium for longitudinal studies*. (Monograph No. 195) Society for Research in Child Development.
- Lee, V. E., & Loeb, S. (1995). Where do Head Start attendees end up? One reason why preschool effects fade out. *Educational Evaluation and Policy Analysis*, 17(1), 62–68.

- Lee, V. E., Brooks-Gunn, J., & Schnur, E. (1988). Does Head Start work?: A one-year follow-up comparison of disadvantaged children attending Head Start, no pre-school and other pre-school programs. *Developmental Psychology, 24*(2), 210–222.
- Lee, V.E., & Loeb, S. (1995). Where do Head Start attendees end up? One reason why preschool effects fade out. *Educational Evaluation and Policy Analysis, 12*(1), 62–68.
- Lee, V. E., Brooks-Gunn, J., Schnur, E., & Liaw, F. R. (1995). Are Head Start effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no pre-school and other pre-school programs. *Child Development, 61*, 495–507.
- Levin, H., Belfield, C., Muennig, P., & Rouse, C. (2007, March 5). *Straight A's: The cost and benefits of an excellent education for America's children*. Columbia Teachers College Press, Columbia University, 7(5) 1, 2, 3, pp.1-24. Retrieved May 5, 2007, from: <http://www.cbcse.org/pages/cost-benefit-studies.php>
- Lewis, A. (2006). We think we can . . . *Washington Commentary, 88*(1), 3–4
- Lewit, E., & Baker, L. S. (1997). Class size. *The Future of Children, 7*(3), 112–121.
- Lindbloom, D. H. (1970). *Class size as it affects instructional procedures and educational outcomes*. (ERIC Document Reproduction Services No. ED059 532)
- Magidson, J., & Sorbom, D. (1982). Adjusting the confounding factors in quasi-experiments: Another reanalysis of the Westinghouse Head Start evaluation. *Educational Evaluation and Policy Analysis, 4*(3), 321.
- Maylone, N. (2002). *The relationship of socio-economic factors and district scores on the Michigan educational assessment program tests: An analysis. 1996–1997*. Unpublished doctoral dissertation, Eastern Michigan University, Ypsilanti.

- McKey, R. H., Condelli, L., Granson, H., Barrett, B., McKonkey, C., & Planz, M. (1985). *The impact of Head Start on children, families, and communities*. Final report of the Head Start Evaluation Synthesis and Utilization Project. Washington, DC: Comprehensive School Reform.
- McRobbie, J., Finn, J. D., & Harman, P. (1998, August). Class size reduction: Lessons learned from experience. (Policy Brief 23). San Francisco, CA: West Ed.
- Miller, L. B., & Bizzel, R. P. (1983). Long-term effects of four pre-school programs: Sixth, seventh, and eighth grades. *Child Development, 54*, 727–741.
- Miller, L. B., & Bizzel, R. P. (1984). Long-term effects of four pre-school programs: Ninth- and tenth-grade results. *Child Development, 55*, 1570–1587.
- Molnar, A., Smith, P., Zahorik, J., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis, 21*(2), 165–177.
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children: Critical Issues for Children and Youth, 5*(2), 113–127.
- Mosteller, F., Light, R. J., & Sachs, J. A. (1996). Sustained inquiry in education: Lessons from skill grouping and class size. *Harvard Educational Review, 66*(4), 797–828.
- Newman, S. (2003). From rhetoric to reality: The case for high-quality compensatory pre-kindergarten programs. *Phi Delta Kappan, 84*(4), 286–291.
- Newmann, F. M., Smith, B., Allensworth, E., & Byrk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis, 23*(4), 297–322.

- Nores, M., Belfield, C. R., Barnett, W. S., & Schweinhart, L. (2005). Updating the economic impacts of the High/Scope Perry Preschool Program. *Educational Evaluation and Policy Analysis, 27*(3), 245–261.
- Nye, B. A., Achilles, C. M., Boyd-Zaharias, J., Fulton, B. D., & Wallenhorst, M. (1994). Small is far better. *Research in the Schools, 1*(1), 9–20.
- Nye, B. A., Boyd-Zaharias, J., Fulton, B. D., Achilles, C. M., & Pate-Bain, H. (1994). *The lasting benefits study: Fourth-year report*. Nashville: Tennessee State University, Center for Excellence for Research in Basic Skills.
- Nye, B. A., Hedges, L. V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class-size experiment. *Educational Evaluation and Policy Analysis, 21*(2), 127–142.
- Ouchi, W. G. (2004, August). Tilting the balance: A management consultant's prescription for changing the equilibrium between central office and schools. *The School Administrator, 61*(7), 19–22.
- Quinn, K. C. (2005). *Beliefs about parenting and how it affects a child's development*. Unpublished doctoral dissertation, Eastern Michigan University, Ypsilanti.
- Ramey, C. T., & Ramey, S. (1989). Early intervention and early experience. *American Psychologist, 53*, 109–120.
- Reynolds, A. J. (2000, June). The added value of continuing early intervention into primary grades. *The CEIC Review, 9*(3), 16-17.
- Robinson, G. L. (1990) Synthesis of research on the effects of class size. *Educational Leadership, 47*(7), 80–90.

- Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the achievement gap*. New York: Columbia University, Teacher's College, Economic Policy Institute.
- Rothstein, R. (2006, February). The limits of testing. *American School Board Journal*, 36.
- Rowan, B. (2007). Comparing teachers' work with work in other occupations: Notes on the professional status of teaching. *American Educational Research Journal*, 44(2) 298–338.
- Sarason, S. B. (1990). *The predictable failure of educational reform: Can we change course before it's too late*. San Francisco: Jossey-Bass.
- Sawhill, I. (2006, Fall) Opportunity in America: The role of education. *The Future of Children*, Princeton University, Princeton, NJ: Brookings Institute, 1–8.
- Schweinhart, L. J., & Weikert, D. (1986). *The effects of the Perry preschool program on youth age 15: A summary*. Hillsdale, NJ: Consortium for Longitudinal Studies, 71–101.
- Shapiro, J. (1982). Evaluation as theory testing: An example from Head Start. *Educational Evaluation and Policy Analysis*, 4(3), 341–353.
- Sirin, S. R. (2005). Socio-economic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–463.
- Slavin R. E. (1987). Ability grouping and student achievement in elementary schools: A best evidence synthesis. *Review of Educational Research*, 57(3), 293–336.
- Slavin, R. E. (1989). *School and Classroom Organization*, Hillside, NJ: Erlbaum.

- Smith, M. L., & Glass, G. V. (1979). *Relationship of class size to classroom processes, teacher satisfaction, and pupil affect: A meta-analysis*. San Francisco, CA: Far West Laboratory for Educational Research and Development.
- Stipek, D. (2005, March 23). Scientifically Based Practice: It's about more than improving the quality of research. *Education Week*, 24(28), 44.
- Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technical innovation*. Washington, D.C.: Brookings Institute.
- Suchman, E. (1967) *Evaluative research*. New York: Russell Sage Foundation.
- U.S. Department of Health and Human Services. (2001, January). *Family and children experiences survey*. Washington, DC: Administration for Children and Families.
- Head Start FACES. Retrieved on January 23, 2004 from:
http://www.acf.hhs.gov/programs/core/ongoing_research/faces/faces_pubs_reports.html
- U.S. Department of Health and Human Services. (2003). *Strengthening Head Start: What the evidence shows*. Washington, DC: Author.
- Unicef. (2007). *Innocenti Research Center Report Card 7. An overview of child well-being in rich countries*.: www.unicef-icdc.org/publications/pdf/re7-eng.pdf
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 91(1), 3–14.
- Wenglinsky, H. H. (1997). How money matters: The effects of school-district spending on academic achievement. *Sociology of Education*, 70, 221–237.
- Whittington, E. H., Pate-Bain, H., & Achilles, C. M. (1985). Effects of class size on first-grade students. *ERS Spectrum*, 3(4), 33–39.

- Word, E., Johnston, J., Bain, H., Fulton, D. B., Boyd-Zaharias, J., Achilles, C. M., Lintz, M. N., Folger, J., & Breda, C. (1990). Student Teacher Achievement Ratio (STAR): Tennessee's K-3 class-size study. Nashville, TN: Tennessee State Department of Education.
- Xiang, Z., & Schweinhart, L. J. (2002, January). *Effects five years later: The Michigan school readiness program evaluation through age 10*. Ypsilanti, MI: High Scope Educational Research Foundation.
- Zaharias, J., Achilles, C. M., & Cain, V. (1995). The effect of random-class assignment on elementary students' reading and mathematics achievement. *Research in the Schools*, 2(2), 7-14.
- Zigler, R., & Styfco, S. J. (1994). Is Perry preschool better than Head Start? Yes and no. *Early Childhood Research Quarterly*, 9, 269-287.

APPENDICES

APPENDIX A: RECOMMENDATIONS FOR CLASS-SIZE CHANGE

Educators have much information available to implement appropriate-sized classes in America's public schools. From years of studying and observing small classes, researchers and practitioners have compiled a research base, theories, and exemplary practices of outstanding teachers to guide effective small-class implementations. Informed Professional Judgement or IPJ is at the heart of class-size changes. **SMALL CLASSES ARE NOT SIMPLY HIRING TEACHERS AND DOING BUSINESS AS USUAL.** A class-size initiative should incorporate what long-term class-size research has determined are important steps for obtaining successful schooling outcomes.

- *1. EARLY INTERVENTION. Start when the pupils enters "schooling" in K or even pre-K.
- *2. SUFFICIENT DURATION. Maintain the small-class environment for at least 3, preferably 4 years for enduring effects. Encourage parent involvement in schooling.
- *3. INTENSE TREATMENT. The pupil spends all day, every day, in the small class. Avoid Pupil-Teacher Ratio (PTR) events, such as "pull-out" projects or team teaching. Develop a psychological sense of community (PSOC), close student-teacher relations, and coherence. Although teacher aids may assist in the building, there is scant evidence that they influence student outcomes positively; except as required by law, reduce aids by attrition.
- *4. USE RANDOM ASSIGNMENT in early grades to facilitate peer tutoring, problem-solving groups, student-to-student cooperation, and active participation and engagement. (STAR).
- *5. EMPLOY A COHORT MODEL for several years so students develop a sense of family or community (PSOC). STAR results show the power of both random assignment and a cohort model. "Looping" adds teacher continuity to the cohort, and may be a useful strategy for added benefits. (More research is needed here.)
- *6. EVALUATE processes and outcomes carefully, and share results. Appropriate-sized classes in elementary grades will take policy and perhaps even legislation change. (Transparency).

The difference between the PTR and actual class size in U.S. elementary schools (about n=10) provides flexibility. If the site has a PTR of 12:1, that suggests enough staff to work toward class sizes of 15 or so, K-3, and still have personnel for special assignments.

Adding endless "projects" ala Title I and continually disrupting the teacher's and students day and continuity (e.g., coherence and stability) are not what the class-size research has shown. To avoid needless costs and confusion, start in K and 1, add a grade per year through the third grade. Reduce "specials" as small-class benefits will allow and relocate personnel to teach small classes.

APPENDIX B: IRB APPROVAL



EASTERN MICHIGAN UNIVERSITY

April 21, 2006

Ms. Deborah Tenjeras Clarke
Department of Leadership and Counseling

RE: "Head Start Fade Phenomena."

The Human Subjects Institutional Review Board (IRB) of Eastern Michigan University has granted approval to your proposal: "Head Start Fade Phenomena".

After careful review of your application, the IRB determined that the rights and welfare of the individual subjects involved in this research are carefully guarded. Additionally, the methods used to obtain informed consent are appropriate, and the individuals are not at a risk.

You are reminded of your obligation to advise the IRB of any change in the protocol that might alter your research in any manner that differs from that upon which this approval is based. Approval of this project applies for one year from the date of this letter. If your data collection continues beyond the one-year period, you must apply for a renewal.

On behalf of the Human Subjects Committee, I wish you success in conducting your research.

Sincerely,

A handwritten signature in cursive script that reads "Patrick".

Dr. Patrick Melia
Administrative Co-Chair
Human Subjects Committee

CC: Dr. Steve Pernecky, Faculty Co-Chair
Charles M. Achilles
C.M. Gibelles

APPENDIX C: SYNOPSIS OF TENNESSEE STAR EXPERIMENT

Table 1. Synopsis of Tennessee's STAR Experiment (1985-1989) and follow-up analyses.

- 1) Star built on principles recognized in prior research. The intervention began in the primary grades. Small classes had fewer than 20 students. STAR's design enables researchers to assess class-size effects on minority as well as on majority students. The design required "real" class-size differences from an average of 24 pupils to an average of 15.
- 2) STAR was a controlled experiment that permitted, to the extent possible with empirical data, causal conclusions about outcomes. Pupils entering K were randomly assigned to a small class (S; 13-17), a regular class (R; 22-25) or a regular class with a full-time teacher aide (RA). Pupils entering in later years were assigned at random to classes as were replacement students. Randomization and testing were monitored carefully. Beginning in grade 1, a blind sample of comparison schools (n = 21) each matched closely to a STAR school, provided external "benchmark" test data. (No intervention.)
- 3) With minor exceptions, students were kept in their class grouping in grades K, 1, 2, and 3. A new grade-appropriate teacher was assigned to the cohort each year. STAR was a 4-year longitudinal class-size study using cohorts.
- 4) Norm-referenced tests (NRT), and criterion-referenced tests (CRT) and measures of self-concept and motivation were administered each spring. Researchers used a post-test only design. (Campbell & Stanley, 1963). Interviews, observations, logs, questionnaires, demographics, etc. provided added data.
- 5) Samples were large and diverse. The K year involved 6300 students in 329 classrooms in 79 schools in 49 districts. The first grade sample was larger still. The large samples were maintained for four years, producing an excellent longitudinal database of n = 11,601 with all test data.
- 6) Classes were maintained throughout the day, all year long. There was no intervention other than the class size and teacher aides. No special training was provided except for a small sample in second grade; no special curricula or materials were introduced. Neither training nor aides affected outcomes.
- 7) Students were evaluated after STAR ended in grade 3; most graduated in 1998. Their college-entrance test results were monitored. (Krueger & Whitmore, 1999), drop out rates, courses taken, and other long-term outcomes have been analyzed. All are positive. (e.g. Finn et al., 2001, 2006).
- 8) STAR activated other TN class-size studies [e.g., Lasting Benefits Study (LBS) Challenge, Enduring Effects, and studies elsewhere (e.g. SAGE)].

APPENDIX D: SUMMARY OF SMALL-CLASS STUDY

Table 6. Summary of Small-Class Benefits: Source, Study, (X= Yes, as included in the source named) STUDY or SOURCE *

Observed In-Class Changes	Lind-bloom 1970	Olson 1971	Glass & Smith 1978	Smith & Glass 1979	Burke Co. 92 - 02	SAGE 2000 +	SSS 1994 1995	FCPS 1997	Cooper 1989	STAR 1985+	Project Success 1994	Teacher "Stories" 2001+	Tot. of 12
A. Increases:												ALL	
• Time on Task		X	X	X	X	X	X	X	X	X	X		11
• Hands-on	X	X		X	X	X	X	X		X	X		10
• Indiv. Attn.	X	X	X	X	X	X	X	X	X	X	X		12
• Diagnosis	X			X	X	X	X	X		X	X		9
• Social Climate	X	X		X	X	X	X		X	X	X		10
• Management	X	X	X	X	X	X	X	X	X	X	X		12
• Participation	X	X	X	X	X	X	X	X		X	X		11
• Academics	X	X	X		X	X	X	X	X	X	X		11
• Parent Involv.					X	X				X	X		5
• Early ID of Spec. Ed.					X					X	X		4
• Morale	X	X		X	X	X	X		X	X	X		10
• Space					X		X			X	X		5
• Enrichment	X	X			X		X	X	X	X	X		9
• Text/Methods	X	X			X	X	X	X	X	X	X		10
• Group Work	X			X	X		X	X		X	X		8
B. Decreases:												ALL	
• Indiscipline	X	X		X	X	X	X		X	X	X		10
• Retention					X					X			3
• Spec. Ed.					X					X	X		4
• Stress		X		X	X	X	X		X	X	X		9

* SSS: Success Starts Small; Achilles et al. (1994); Kiser-Kling (1995); SAGE: Student Achievement Guarantee in Education, MoInar (1998); Project Success from Achilles et al. (1994); FCPS: Fairfax County (1997); STAR (Word et al., 1990); Teacher stories are from CA, NC, SC, TN, and WI. Other authors are listed in References. Adapted from Tables 6.12, p. 104 (Achilles, 1999).

APPENDIX E: WHY SMALL CLASSES WORK

Small-class (K-3) Benefits Obtained in the STAR Experiment are Supported by Research and Theories about Learning, Teaching and Contexts.

I. LEARNING

- A. Task Induction: Learn About School (Student's Work)
- B. Participation, Engagement, Identification.
- C. Time On Task Increases.
- D. Mastery of Basics in Less Time.
- E. Appropriate Use of Homework.
- F. Developmentally Appropriate Activities.
- G. Early Intervention, Duration.
- H. More Materials,: "Hands On".
- I. Opportunity to Learn (OTL)*.

II. TEACHING

- A. Teach to Mastery.
- B. Immediate Reinforcement.
- C. Early Diagnosis and Remediation of Learning Difficulties.
- D. Individual Accomodation.
- E. Assesment (In-Class).
- F. Effective Teaching Methods.
- G. Portfolios, Running Records, etc.
- H. Planned, Coherent Lessons. (Seamless Transitions).
- I. Opportunity to Teach (OTT)*.

III. CLASSROOM/CONTEXTS

- A. Classroom Environment (E.G. : Air Quality, Space, Crowding, Noise).
- B. Variable Room Arrangements (E.g., Learning Centers, Groups).
- C. Inclusion, Special Needs.
- D. Classroom Management.
- E. Less Indiscipline.
- F. Opportunities for Volunteers.
- G. Mixed ability/Random Assignment.

IV. "OTHER"

- A. Increased Parent Interest.
- B. Reduced Grade Retention/Dropout.
- C. Increased Student/Teacher Morale and Energy.
- D. Teacher Accountability and Responsibility.
- E. Few projects and "Pull Outs". Intensity (all day, each day).
- F. Student-Led Activities.
- G. Assesment (Outcome).
- H. Field Trips with Fewer Adults/Small Vehicles.

V. STUDENT BEHAVIOR (B)**

1. Class size and Engagement: Students are more engaged in learning and pro-social (B) and less in disruptive (B) . Principles: 1) "Visibility of the Individual" a) Time per student, b) Diffusion of Responsibility and c) Social loafing; 2. Sense of Belonging a) Group norms [e.g., Learning (B)] influence all members, b) Psychological Sense of Community (PSOC): Results are similar to some school-size work. [Equity/Gaps/Graduation/Dropout/Retention.]

*Correct numbers of students provide teachers the Opportunity to Teach (OTT) and students the Opportunity to Learn (OTL). (OTT and OTL are reciprocal.)

** Finn, J.D. Pannozzo, G.M. & Achilles, C. (2003, Fall). The "Whys" of class size: Student behavior in small classes. Review of Educaional Research, 73(3), 321-368.

APPENDIX F: MAJOR IMPEDIMENTS TO SMALL CLASSES

Pupil-Teacher-Ratio is NOT Class size. (References available)

In Educational Policy Systems, Iannaccone (1975) emphasized the issue driving this confusion: “descriptive reference is the first and most essential sense in which a concept has meaning” (p.13). He explained that: “One source of error in the scientific venture is lack of precision in referent of the concepts. Lack of precision leads to lack of reliability in the concepts.” (pp.13-14)

An economist who often criticizes small classes using PTR arguments makes the same point as Iannaccone. Hanushek¹ (1998) noted that 1) “pupil-teacher ratios are not the same as class-sizes,” and 2) “The only data that are available over time reflect the pupil-teacher ratios” (p.12). Hanushek proceeds to substitute one term for the other in his work and criticize class-size. (Emphasis Added).
Some Definitions:

Average Class Size is the sum of all students regularly in each teacher’s class divided by the actual number of regular teachers in those specific classes. If four 2nd grade rooms have 14, 16, 18, 18 students (n=65) students, the average (not actual) grade-2 class size is 16.25. (or 16).

Class Size(s): - “The number of students for whom a teacher is primarily responsible during a school year (Lewit & Baker, 1997, p.113).” This is an addition problem. Class size is an organization for instruction important to teachers, parents, students.

Class-Size Reduction (CSR) includes the process to achieve class-sizes small than the ones presently in place, such as changing the class size from 25 to 16 or so. One needs accurate pre and post data to support the change process.

Pupil-Teacher Ratio PTR – “The number of students in a school or district compared to the number of teaching professionals” (McRobbie et al., 1998, p.4). Some times all educators are part of the computation, including counselors, administrators, etc. In this division problem, the divisor is very important. The difference between PTR and class size in USA elementary schools is about n=10 (Achilles & Sharp, 1998). PTR is a formula and process for equitable allocation of resources important to administrators, policy persons, etc. Class size is an organization for provided instructional and education services to clients. Akerhielm (1995) determined that the definition of class size influenced inconsistent results in studies. Lacking actual class-size data, researchers often used PTR as the indicator of class size. When an aggregate ratio differs from a student’s actual class sizes, measurement error can bias the “class size” variable toward zero. (E.g., special low-density classes).

Valid and reliable ways to get class-size data are 1) to count the students in a class and/or 2) to establish class sizes and monitor them as in Tennessee’s STAR study. Surveys and databases usually generate PTR’s. Class size and PTR are different constructs that should not be confused. Each has research purposes and related outcomes.

¹ The Evidence on Class Size (1998). Rochester, NY: W.A. Wallis Institute, Excerpts. (Emphasis Added), Biddle and Berliner (2002) critiqued Hanushek’s work in “Small class size and its effects,” Educational Leadership. 59 (5), 12-23 (Esp. pp.14-15), as did Krueger (2000). The NCES Digest of Education Statistics contains separate PTR and class-size tables. (e.g., Tables 65 and 69 in 1999). See Akerhielm, K. (1995). Does class size matter? Economics of Education Review, 14(3), 229-241

APPENDIX G: STAR FINDINGS ON INSTRUCTION—EXECUTIVE REPORT

Based on four years of interviews, patterns emerged in kindergarten and continued through the third grade. The following advantages were apparent for instruction in small and regular/aide classes:

1. Basic instruction was completed more quickly, providing more time for covering additional basic material,
2. Use of supplemental text and enrichment activities,
3. More in-depth instruction regarding the basic content,
4. More frequent opportunities for children to engage in first-hand learning activities using concrete materials,
5. Increased use of learning centers and
6. Increased use of highly desirable primary grade practices.

Improved individualization of instruction also emerged as a dominant theme in small and regular/aide class teachers' perceptions. Teachers reported: 1) increased monitoring of student behavior and learning, 2) opportunities for more immediate and more individualized reteaching or enrichments, 3) more frequent interactions with each child, 4) a better match between each child's ability and the instructional opportunities provided, 5) a more detailed knowledge of each child's needs as a learner, and 6) the necessary time to meet individual learner's needs using a variety of instructional approaches. Significant reduction of class size or the addition of a full-time teacher aides also made positive changes in the physical, social, and emotional environments in primary grade classroom. Classrooms were more pleasant work environments for both teachers and students. Teachers and students were under less stress, and learning occurred in a more relaxed atmosphere. Students were less likely to get lost in the crowd.

Conclusions

The results of STAR are definitive enough to elicit the following statement by Dr. Jeremy Finn:

This research leaves no doubt that small classes have an advantage over larger classes in reading and mathematics in the early primary grades. This experiment yields an unambiguous answer to the question of the existence of a class-size effect, as well as estimates of the magnitude of the effect for early primary grades.

These are strong words for research in education and are possibly due to the design and power of Tennessee's Project STAR, which has paid considerable attention to maintaining the required research standards and controls.

APPENDIX H: DESCRIPTIVE STATISTICS MAT 8 SUPPORT FOR TABLE 4

Statistics

		total reading score	sounds and print	reading vocabulary	reading comprehension	total reading score
N	Valid	45	45	45	45	29
	Missing	23	23	23	23	39
Mean		523.8444	545.8000	493.0222	526.6667	561.3793
Median		521.0000	541.0000	482.0000	512.0000	558.0000
Std. Deviation		41.88076	46.10196	85.77150	41.31091	37.57128
Minimum		456.00	464.00	52.00	444.00	497.00
Maximum		605.00	655.00	615.00	597.00	647.00

Statistics

		sounds and print	reading vocabulary	reading comprehension	total reading score	sounds and print
N	Valid	29	29	29	18	18
	Missing	39	39	39	50	50
Mean		573.2069	558.0000	558.8966	607.1667	601.4444
Median		552.0000	552.0000	561.0000	607.0000	598.5000
Std. Deviation		39.38489	41.60700	44.13967	23.20814	30.95010
Minimum		524.00	488.00	482.00	570.00	552.00
Maximum		662.00	637.00	645.00	646.00	666.00

Statistics

		reading vocabulary	reading comprehension
N	Valid	18	18
	Missing	50	50
Mean		616.3889	606.1111
Median		605.0000	609.5000
Std. Deviation		28.00029	29.34559
Minimum		582.00	558.00
Maximum		670.00	654.00

APPENDIX I: WILCOXON MAT 8 DATA SUPPORT FOR TABLE 5

Ranks

		N	Mean Rank	Sum of Ranks
total reading score - total reading score	Negative Ranks	1 ^a	2.00	2.00
	Positive Ranks	19 ^b	10.95	208.00
	Ties	1 ^c		
	Total	21		
sounds and print - sounds and print	Negative Ranks	2 ^d	9.00	18.00
	Positive Ranks	18 ^e	10.67	192.00
	Ties	1 ^f		
	Total	21		
reading vocabulary - reading vocabulary	Negative Ranks	0 ^g	.00	.00
	Positive Ranks	20 ^h	10.50	210.00
	Ties	1 ⁱ		
	Total	21		
reading comprehension - reading comprehension	Negative Ranks	3 ^j	7.67	23.00
	Positive Ranks	17 ^k	11.00	187.00
	Ties	1 ^l		
	Total	21		

- a. total reading score < total reading score
- b. total reading score > total reading score
- c. total reading score = total reading score
- d. sounds and print < sounds and print
- e. sounds and print > sounds and print
- f. sounds and print = sounds and print
- g. reading vocabulary < reading vocabulary
- h. reading vocabulary > reading vocabulary
- i. reading vocabulary = reading vocabulary
- j. reading comprehension < reading comprehension
- k. reading comprehension > reading comprehension
- l. reading comprehension = reading comprehension

APPENDIX J: WILCOXON MAT 8 DATA SUPPORT FOR TABLE 6

Ranks

		N	Mean Rank	Sum of Ranks
total reading score - total reading score	Negative Ranks	0 ^a	.00	.00
	Positive Ranks	12 ^b	6.50	78.00
	Ties	0 ^c		
	Total	12		
sounds and print - sounds and print	Negative Ranks	1 ^d	6.00	6.00
	Positive Ranks	11 ^e	6.55	72.00
	Ties	0 ^f		
	Total	12		
reading vocabulary - reading vocabulary	Negative Ranks	1 ^g	2.00	2.00
	Positive Ranks	11 ^h	6.91	76.00
	Ties	0 ⁱ		
	Total	12		
reading comprehension - reading comprehension	Negative Ranks	1 ^j	1.00	1.00
	Positive Ranks	11 ^k	7.00	77.00
	Ties	0 ^l		
	Total	12		

- a. total reading score < total reading score
- b. total reading score > total reading score
- c. total reading score = total reading score
- d. sounds and print < sounds and print
- e. sounds and print > sounds and print
- f. sounds and print = sounds and print
- g. reading vocabulary < reading vocabulary
- h. reading vocabulary > reading vocabulary
- i. reading vocabulary = reading vocabulary
- j. reading comprehension < reading comprehension
- k. reading comprehension > reading comprehension
- l. reading comprehension = reading comprehension

APPENDIX K: MEAP T-TEST DATA SUPPORT FOR TABLE 7

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
MEAP standard score	16	389.4375	46.43701	11.60925

One-Sample Test

	Test Value = 400					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
MEAP standard score	-.910	15	.377	-10.56250	-35.3070	14.1820

APPENDIX L: MEAP DATA SUPPORT GROUP 1 FOR TABLE 8

Statistics

MEAP dichotomized

N	Valid	16
	Missing	52

MEAP dichotomized

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not meet state standards (<400)	6	8.8	37.5	37.5
	Met state standards (>400)	10	14.7	62.5	100.0
	Total	16	23.5	100.0	
Missing	System	52	76.5		
Total		68	100.0		