2010

The Influence of Small Class Size, Duration, Intensity, and Heterogeneity on Head Start Fade

Christopher D. Huss
Seton Hall University

THE INFLUENCE OF SMALL CLASS SIZE, DURATION, INTENSITY, AND HETEROGENEITY ON HEAD START FADE

by

Christopher D. Huss

Dissertation

Submitted to the Department of Education Leadership, Management, and Policy

Seton Hall University

in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

Dissertation Committee:

Charles M. Achilles, EdD, Mentor
Michael Osnato, EdD
Christopher H. Tienken, EdD
Deborah T. Clarke, EdD

February 5, 2010

SETON HALL UNIVERSITY
COLLEGE OF EDUCATION AND HUMAN SERVICES
OFFICE OF GRADUATE STUDIES

APPROVAL FOR SUCCESSFUL DEFENSE

Doctoral Candidate, **Christopher Huss,** has successfully defended and made the required modifications to the text of the doctoral dissertation for the **Ed.D.** during this **Spring Semester 2010.**

DISSERTATION COMMITTEE

(please sign and date beside your name)

Mentor:
Dr. Charles Achilles C M. Achilles 4-8-10

Committee Member:
Dr. Michael Osnato Michael Osnato 2-5-10

Committee Member:
Dr. Chris Tienken 2.5.10

Committee Member:
Dr. Deborah Clarke Deborah Clarke 2.5.10

External Reader:
_____

The mentor and any other committee members who wish to review revisions will sign and date this document only when revisions have been completed. Please return this form to the Office of Graduate Studies, where it will be placed in the candidate's file and submit a copy with your final dissertation to be bound as page number two.

DEDICATION

This dissertation is dedicated with love and gratitude to my family. First and foremost, I want to dedicate this work to my parents who devoted their lives to making sure my five siblings and I had everything we needed to become successful adults even when it meant sacrificing so much of themselves. I also want to dedicate this to my five siblings: John, Deborah, Michael, Joseph, and Thomas, who were always there for their baby brother through the good times and the bad. Your support and love has given me the courage and confidence to strive to do more and to be more. To my wife, Jamie, and children, Matthew and Jason, to whom I owe the most gratitude – thank you. Without you by my side supporting me each and every day, I would have never been able to complete this journey. I am eternally grateful for your patience, understanding, and love.

ACKNOWLEDGEMENTS

A special thank you is given to my mentor, Dr. Chuck Achilles. You have been an inspiration since I first entered your "Policy" class as a member of Cohort XI at Seton Hall University. You continually required me to think critically and to work diligently all the while supporting and instilling confidence in me. I am honored and grateful for your assistance and support during this process. I also wish to acknowledge Dr. Chris Tienken for the guidance and clarity you offered throughout this journey. You are not only a wealth of information; you are a great motivator and leader. I thank you profoundly for agreeing to be a member of my committee and for teaching me so many meaningful lessons. I wish to acknowledge Dr. Deborah Tenjeras Clarke for providing me with the idea, direction, and support for this study and Dr. Michael Osnato for providing encouragement. In addition, my endless gratitude goes out to all of the members of Seton Hall's Cohort XI for their support and friendship throughout this adventure. This especially applies to Steve Wisniewski who has been my "wingman" since long before the start of this journey and who continues to be there for me to this day. My appreciation goes out to all of you.

ABSTRACT

The researcher conducted a nonexperimental study to investigate and analyze the influence of reduced class sizes, intensity (all day and every day), duration (five years), and heterogeneity (random class assignment) on the Head Start Fade effect. The researcher employed retrospective data analysis using a longitudinal explanatory design on data obtained from students in an urban-fringe district in New Jersey. The academic achievement data used as the primary measures of achievement in this study were the 1st and 2nd grade Terra Nova test results and outcomes on the 3rd grade New Jersey Assessment of Skills and Knowledge (NJASK3).

Data revealed that students who were members of the treatment group, under most of the theory's conditions, did not perform significantly better on multiple independent t-tests than students who were not members. Data revealed that being a member of the treatment group minimally influenced achievement, as measured by using Cohen's d (effect size). Data also revealed that students who were members of the treatment group performed nearly equal to, or lower than, their peers, as measured by state and district factor group (DFG) averages on the NJASK3.

At first, study findings appear to be inconsistent with the theoretic construct and scientific-based research on the influence of early intervention, class size, intensity, duration, and heterogeneity, especially with regard to students from lower economic stature. Upon closer scrutiny, however, study findings provided data that confirmed what is stated in the theory, that when not implemented correctly, class size reduction (CSR) initiatives are ineffective. In this study, the researcher also tracked the influence that effective programs such as the Perry Child Development Center, the Abecedarian

Program, and the Chicago Parent Centers had on eliminating or moderating the fade effect. The researcher compared and contrasted these model programs with the one offered in the present study to highlight the importance of consistency when implementing conditions of a theory.

Table of Contents

LIST OF TABLES

CHAPTER 1: INTRODUCTION AND BACKGROUND

Students from homes and neighborhoods of low socioeconomic status (SES) have historically performed less well on traditional measures of academic achievement (e.g., standardized tests) than have peers, and have consequently been presented with fewer opportunities to earn higher wages and increase their quality of life. While educators cannot immediately improve the SES of their students, they are continually faced with the task of trying to overcome what appears to be the biggest factor affecting school performance: poverty. The effects and influences of poverty on achievement have been well-documented in numerous studies, including the meta-analyses of Sirin (2005) and Berliner (2006), which are discussed in the literature review of this study. These findings have led to numerous agendas on school reform, especially in impoverished neighborhoods, and a particularly significant interest and increase in funding for public preschools in such areas. Head Start and similar programs have been funded to provide early intervention in terms of school readiness, health care services, and nutrition to the nation's neediest students. The abundance of data, however, show that not all publicly funded programs, and in fact very few of them, have been effective at providing measurable positive results (Westinghouse Learning Corporation, 1969; Lee & Loeb, 199, Zill et al., 2003). Additionally, the data show that many of the gains made by students in the initially successful programs fade out over time (Head Start Fade).

There are, however, compelling findings from valid and reliable studies, as well as promising ideas put forth by education theorists that support the idea that the harmful effects of poverty on schooling can be moderated. Several model preschool programs

and a number of class-size studies have demonstrated that there are effective ways to reach all students, regardless of their SES, and achieve measurable and sustainable gains on standardized tests. Chapter 1 includes a review of poverty and its effect and influence on education, provides a background of practices and theories that have some success in alleviating the obstacles presented by poverty, and presents an overview of this study that will explain a test of theories and practices in a site-based public school environment.

Background

In the summer of 1965, the United States government formed a blue ribbon task force to develop the first publicly funded early intervention program for young poor children. This program, called Head Start, was an eight-week program created to help improve the physical and emotional health of disadvantaged children, as well as develop their mental processing abilities and skills prior to entering school for the first time. Since 1965 and the enactment of the Elementary and Secondary Education Act (ESEA) (P.L. 89-10), a program related to President Lyndon B. Johnson's "War on Poverty", the federal government has continually increased its financial involvement in compensatory education, especially in impoverished neighborhoods. Head Start has evolved into what is now mostly a national full-year program that promotes school readiness, with a special focus on developing early reading and math skills, for economically disadvantaged children and their families.

Head Start programs provide educational, health, nutritional, and social services to over 900,000 pre-kindergarten children and their families, at an average cost per child of $7,209 (OHS 2008). In the 2006 fiscal year, over $6.7 billion was spent on Head Start programs in the United States of America (USA), and in the years of Head Start's

existence, over $66 billion of taxpayer money have been spent (OHS, 2008). Still, the student achievement outcomes related to Head Start are inconsistent, and the evidence to definitively determine if such a large public expenditure is effective in helping the nation's neediest children remains elusive.

Researchers such as Barnett (2002) conducted extensive studies on the topic of early intervention. He argued that "Nearly four decades of research establish that Head Start delivers the intended services and improves the lives and development of the children and families it serves" (p. 1). However, not everybody agrees. In fact, some critics, such as Hood (1992), have gone as far as calling Head Start a scam and have argued that the program has little, if any, long-term impact on children. He claimed that early intervention studies showing positive results have been conducted on model programs that have little relevance in the real world.

Between these extremes lie various opinions and studies on Head Start and its effectiveness. The original researchers who conducted the Westinghouse Head Start Evaluation in 1969 found that Head Start attendees showed some immediate gains in first grade but that these gains, when measured using a variety of tests and methods, faded out over time – as early as second and third grade (Westinghouse Learning Corporation, 1969). These early findings about Head Start lead to the commonly used term "Head Start Fade". Other academics and researchers have conceded the point that early cognitive gains may be lost or level out over time, but they have argued that, when measured using long-term cost/benefit tests, both cognitive and positive societal outcomes do prevail, which alone justify the costs of publicly funded preschool. In 2002, Currie stated "A simple cost-benefit analysis suggests that Head Start would pay for itself

in terms of cost-savings to the government if it produced even a quarter of the long-term gains of model programs" (p.1).

A thorough review of literature and research related to Head Start and pre-kindergarten programs yielded inconsistent results. The findings range from strong relationships between attendance in Head Start programs and student achievement to no significant correlation at all. While debates about the promotion of universal pre-k and various studies that support opposing viewpoints persist, a substantial amount of money continues to be spent on Head Start and other publicly funded preschool programs. Without a definitive answer as to whether the money spent is worth it in terms of measurable student outcomes on standardized tests, the federal government seems willing to continue funding Head Start for the foreseeable future based on a no-harm-done rationale. For example, President Obama has set aside billions of dollars from the economic stimulus project to fund "cradle to career" learning and promote early childhood education programs (Colvin, 2009, p. 1). Thus, it has become evident that an unbiased site-based study to examine some of the issues presented in the literature related to Head Start and publicly funded preschool is justified. In fact, state and local education administrators need evidence of the effectiveness of Head Start, both long- and short-term, in order to justify their expenditures on Head Start or other publicly funded pre-kindergarten programs.

In 2007, Clarke explored the Head Start Fade phenomenon and the influence that the application of research-based practices had in reducing the deterioration of initial gains. Specifically, Clarke investigated whether the application of "previously established theories could mitigate or eliminate the fade-out effect experienced by Head

Start participants" (p. iv). Clarke proposed that students who stayed in small, heterogeneously grouped, classes in the three or four grades following pre-kindergarten, and who remained with the same teacher for most of the school day each year, would not experience Head Start Fade. These treatment conditions were labeled as early intervention, duration, intensity, and heterogeneity. Clarke found: "These results confirm that there should be no fade effect, or a minimal one, when the elements of the theory are present in a school on a consistent basis" (p.95). Thus far, the evidence suggests that meeting conditions put forth by Clarke's theoretical model could prevent or reduce Head Start Fade. Clarke's initial findings have led to the present follow-up study, in which the theories of early intervention, duration, intensity, and heterogeneity were tested in a new setting and with larger, more diverse samples of students.

In this study, a quantitative analysis was conducted using grouped data collected on students who participated in a preschool that met the requirements of the Head Start curriculum (early intervention) and who met the majority of the requirements of duration, intensity, and heterogeneity as explained by Clarke (2007) and tested by Clarke and Achilles (2008). The study's results will contribute to the existing body of quantitative research on Head Start and similar publicly-funded programs, provide a site-based analysis on the effectiveness of a specific program, and test the theory of the effects of early intervention, duration, intensity, and heterogeneity.

The ideal study would replicate Clarke's study under experimental conditions and provide definitive evidence to support or discredit the proposed theories. This study, however, was not conducted under experimental conditions. It was a nonexperimental, longitudinal, explanatory research study (Johnson, 2001), using matched pairs, formed

from preexisting, intact groups, to replicate experimental conditions. This study extended Clarke's study in a different state with more robust data including more students and multiple groups of students' data. Properly done, it could present more definitive evidence on the effectiveness of Head Start and the elimination of Head Start Fade.

The research conducted for this study is relevant because every Head Start program, although federally funded, is administered locally and somewhat independent of all others. In addition, national Head Start standards were put into place under the No Child Left Behind Act (NCLB) (P.L. 107-110) in 2002, and Head Start has become a model for many locally funded pre-kindergarten programs and for the universal pre-kindergarten movement which is gaining national momentum and support under the government's administration. Currie and Thomas (1994) explained:

> Given that there are over 1,300 Head Start programs, all administered at a local level, and that the program guidelines are not specific about how the goals of the program are to be attained, there is bound to be a great deal of heterogeneity in program content. The quality of programming is uneven across the country (p.22).

The present study provides municipality-specific, relevant data analyses, something that national studies, comparing participants from numerous locations and from differing programs, are unable to do. The researcher tested the theory presented by Clarke (2007) and examined possible ways to eliminate the diminution of initial gains made in Head Start (The Fade).

Problem Statement

A distinct and persistent academic difference in achievement, as evidenced by standardized test measures, exists between students from homes and neighborhoods of low socioeconomic status (SES) and their peers who reside in wealthier neighborhoods. Consequently, students from lower SES communities are presented with fewer opportunities to earn higher wages and increase their quality of life. One promising intervention that has been implemented since 1965 by the federal government to remedy this problem is federally funded early intervention through Head Start. Reliable data in support of Head Start's lasting positive influence, however, is not abundantly available. Parents, educators, and legislators must be presented with this data to determine if Head Start, and other publicly funded early intervention programs, is a viable solution for reducing the measurable difference among low SES students and their peers.

Various studies indicate that properly delivered early intervention and education experiences influence achievement. Additional early intervention and class-size studies suggest that continuing intervention and education programs for at least four years under specific research-based conditions positively influences the endurance of academic and social gains. Therefore, this researcher tested a municipality-specific Head Start program and addressed strategies that have historically yielded positive academic outcomes for the potential to reduce the persistent achievement gap between students from low SES environments and their peers.

Purpose

The purpose for this study was to explain the influence of research-based practices on Head Start Fade to determine whether the fade effect can be eliminated or

reduced if research-based practices and demonstrated theories are implemented in the early intervention program and continued in subsequent years. The intent was to explain the influence of implementing preschool follow-up conditions of reduced class sizes, instruction delivered by the same teacher all day and every day, and participation in heterogeneously formed classes on later academic achievement. To demonstrate this, the researcher compared and contrasted grouped academic achievement data from students who met, to some degree, all of the following conditions: (a) participated in a preschool program delivered under Head Start conditions, (b) attended school in reduced-sized classes for four or five continuous school years, and (c) stayed with the same teacher for the majority of the school day in heterogeneously grouped classes to students who did not meet the same criteria. The researcher determined what relationship existed between attendance in a Head Start program with the desired follow-up conditions proposed by Clarke (2007) and later academic achievement. The researcher explored if "Head Start Fade", a term used to describe the loss of advantages gained from Head Start, could be supported by these data. The researcher examined the data to determine if the phenomenon of "Head Start Fade" was statistically significantly reduced and practically reduced (calculated effect size) if most, but not all, of the treatment conditions proposed in Clarke's study (early intervention, duration, intensity, cohorts, and heterogeneity) were met in subsequent years following preschool by comparing groups of students who (a) do meet, and (b) do not meet established criteria.

Research Questions

Answers to the following questions were used to attain the purpose of this study:

1) What are the criteria for early intervention, duration, intensity, and heterogeneity as derived from prior research and theory?

2) What are the test results of students who meet these criteria on standardized tests, and do these outcomes show evidence of a diminution of "Head Start Fade"?

3) How do these results compare with test outcomes of comparison groups of students who did not meet the study's conditions using measures of statistical significance (Independent t-tests) and practical significance (Cohen's d)?

4) How do these results compare to the NJ state average results on the Grade 3 statewide assessments of language arts and mathematics and to the average results of students from the same DFG (baseline)?

Significance of the Study

The study adds needed empirical data to the body of work related to Head Start without using model programs, privately funded programs, or a collection of data from multiple programs. The study was conducted using student scores from a certified Head Start program who all attended a program in the same municipality, thus addressing arguments about preschool studies being conducted using model programs or collated data. According to Hodges and Cooper (1981), "If projects from all over the country are lumped together for analysis, the ineffective projects cancel the effects of effective projects and the result is no difference" (p.227). The parameters for this study led to useful information more readily than nationwide studies described by Hodges and Cooper and provided more relevant information than do studies of private or so-called model programs.

The researcher recorded data using two different measures of student academic achievement at the elementary level. These data were examined to determine if statistically significant correlations existed among attendees of a Head Start program who meet the theory-supported criteria for success: early intervention, duration, intensity, and

heterogeneity as theorized by Clarke (2007) by analyzing their academic success as measured by the Grades 1 and 2 Terra Nova and the Grade 3 New Jersey Assessment of Skills and Knowledge (NJASK3). In addition, the same data were compared and contrasted with data from Head Start, or pre-k, students who did not meet the established criteria to determine if attendees performed as well as, better than, or worse than their peers in programs meeting selected criteria. These data were examined to determine if Head Start Fade exists among the study's participants. Finally, these data were analyzed to determine if early intervention, duration, intensity, and heterogeneity affects students' achievement when compared to state DFG averages.

Issues addressed in this study provided insight into if Head Start attendees benefitted academically from the program, and if benefits endured if the conditions were properly met in subsequent years as outlined in recent theory. The study provided valuable results in this area of research because the design included groups of students from the same municipality. This selection process helped assure that the participants varied less than those used in nationwide studies. Participants in this study were products of the same district factor group (DFG), the same schools, and had similar demographics. Although residing in the same municipality does not equate to equal living conditions and home environments, it does control for many differences which are evident and unavoidable in nationwide studies. The study participants have all attended pre-kindergarten in the district's public school setting, as opposed to a multitude of private agencies as is often the case. This resulted in data being more readily available than in many instances and provided a consistent and uniform delivery of services and curriculum to the participants. Ideally, this study, which has evolved from a smaller

study and current theory, has provided robust enough information to lead to an experimental study under similar conditions.

Based on results of this study, the researcher should be able to identify strategies used in the municipality's Head Start program and in subsequent years in the elementary schools which result in improved academic performance and decreased fade-out effect. Good results would allow legislators, administrators, and educators to prevent the fade effect from occurring and save the inordinate cost of remediating the long-term problems following early failure in school. In their paper on the cost-effectiveness of preschool, Temple and Reynolds (2007) stated, "The findings suggest that investments in preschool had substantially higher net benefits and benefit-cost ratio than did several education, job training, and health service interventions" (p.142). Thus, it appears reasonable to spend money early in a child's life and reap the continual rewards than to spend exorbitantly into adulthood with fewer benefits.

Limitations

Limitations in this study are as follows: The sample size for the grouped data was limited to the number of students enrolled in the public preschool in the chosen years and who met most of the conditions of the pre-established criteria as proposed by Clarke (2007). Random assignment was not possible because this was a retrospective study and the students whose test results were used have already exited the program. In addition, "true" or complete cohorts of students did exist, due to reassignment of students to new classes each year. Class assignments have already occurred. The reassignment of students to groups and to teachers might have resulted in different instructional approaches used by teachers and different experiences of attendees based on their varied

classroom assignments. Class assignments were not made with random assignment of students, thus cause and effect cannot be determined. Another limitation is the loss of students through attrition. For example, of the 287 students who began the Head Start program in 2002, 117 were no longer registered with the district, and out of the 191 students who began the program in 2003, by 2009 54 were no longer registered in the district. These students' data were not available for the study.

Delimitations

Delimitations for this study, imposed by the researcher, were the numbers of years that participants received the treatment conditions and the amount of data collected to evaluate academic success. The years of interest for this study were school years 2002 through 2008. These years were chosen based on the year attendees began Head Start in the fall and the year in which they were administered the NJASK3 state standardized test. The data used to determine academic success have been delimited to Terra Nova and NJASK reading, language arts, and mathematics scores, as available.

Definition of Terms

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

1. Class Size: "The number of students for whom a student is primarily responsible during a school year" (Lewit & Baker, 1997, p. 113).
2. Cohort: A group of individual students who stay together as a class for multiple years in school.
3. District Factor Group (DFG): Classification of school districts based on SES used only by the state of New Jersey. See Appendix B for detailed information.

4. Duration: Students remained in small classes from the start of the study until its conclusion – preschool through third grade.
5. Early Intervention: attendance in a preschool program
6. Head Start: "Head Start is a national program that promotes school readiness by enhancing the social and cognitive development of children through the provision of educational, health, nutritional, social and other services to enrolled children and families" (OHS, 2008). In this study, "Head Start" was used frequently to refer to any publicly funded preschool program.
7. Heterogeneity: Classes are comprised of dissimilar and/or diverse students in terms of race, gender, ability-level, SES, etc.
8. Intensity: a) classes are assigned to the same teacher b) for all academic subjects c) each school day.
9. 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) (emphasis added).
10. Socioeconomic Status (SES): A family's socioeconomic status is based on family income, parental education level, parental occupation, and social status in the community (Demarest, et al., 1993, p. 1).

Organization of the Study

Chapter 1 has provided an introduction to the concept of Head Start Fade and to the research-based practices and theories which were tested to determine if they have an

influence on the popular phenomenon. It included brief statements of research methodology, significance of the study, definitions of terms, limitations, delimitations, and research questions. Chapter 2 includes a review of current early intervention literature, a review of research and theory related to the Head Start Fade effect, and an exploration of the research and theory related to small class sizes. Chapter 3 explains the research design and methods in detail, while Chapter 4 presents retrospective data on Head Start attendees and analyses of those data. Chapter 5 presents a summary of findings, discussion and conclusions, and offers recommendations for policy, practice, and further studies.

CHAPTER 2: REVIEW OF RELEVANT RESEARCH, THEORY, AND LITERATURE

Numerous studies have shown the debilitating effects of poverty on children's academic achievement (e.g., White 1982, Lee & Burkham, 2002; Sirin, 2005; Quinn, 2005). In the state of New Jersey, there is a direct correlation between the neighborhood which students come from and their performance on the state's standardized test (Tienken, 2008) (See also Appendix B, Table 1). With this knowledge, politicians, researchers, theorist, and practitioners have attempted numerous interventions to help overcome this seemingly insurmountable obstacle. Examples of such interventions are after-school tutoring programs, free and reduced lunch programs, and integrated school-based social service programs. In addition to these programs, preschool has been financially supported by the federal government in the form of Head Start since 1965. In spite of longstanding support and financial backing, Head Start in general has failed to produce the type of long-lasting results that would be considered successful.

A review of research and education theory provided insight into why Head Start has not made the significant impact that was expected of it. Model early intervention programs which have shown sustainable results and have had significant influence on the lives of impoverished children have tightly structured programs which include small class sizes, duration of the program beyond the first year, heterogeneously grouped classes, and considerable amounts of follow-up activities. Similarly, several class-size studies have found that meeting these same conditions in the primary grades produce sustained measurable results in the areas of academic achievement and affective development.

This chapter contains a review of:

(a) The background and history of Head Start

(b) The origin and analysis of Head Start "fade"

(c) The purported benefits of Head Start over the past 40 years

(d) Several analyses of Head Start in terms of cost vs. benefits

(e) The evidence presented by staunch supporters of Head Start and well as by those who oppose the program

(f) The impact of poverty on academic achievement

(g) Model programs which have produced significant results

(h) Class size studies

(i) The development of the theory of early intervention, duration, intensity, and heterogeneity.

The chapter also includes the reasons that these areas of focus are important to the study and have led to its development and design. Finally, in this chapter, the researcher will attempt to show the reader the importance of the study and why its findings may be useful for other researchers, policy makers, and education practitioners.

Questions for this literature review are as follows:

(a) What have studies shown about the effects of Head Start on impoverished children?

(b) What have studies of class size and model early intervention programs found effective for influencing academic achievement in students?

(c) What theories, practices, and policies should be used from class-size studies and model early intervention programs to improve the effectiveness of Head Start?

By discovering the answers to these questions, the researcher will identify the problems with the current design and implementation of Head Start. The researcher will also offer suggestions which can be used to ensure that Head Start and other publicly funded preschool programs are being delivered most effectively.

Background on Head Start

"Head Start is a national program that promotes school readiness by enhancing the social and cognitive development of children through the provision of educational, health, nutritional, social and other services to enrolled children and families" (OHS, 2008). Often thought of as public preschool, Head Start encompasses much more than what one would consider as such. What started as a federally funded summer program in 1965 as part of President Johnson's "War on Poverty" has evolved into an all-year service dedicated to providing comprehensive child-development services. Although administered at the local level, Head Start is funded by federal grants and must abide by federal guidelines: "The guidelines specify that, in addition to providing a nurturing learning environment, Head Start should provide a wide range of services. These include, for example, facilitating and monitoring utilization of preventive medical care by participants, as well as providing nutritious meals and snacks" (Garces et al., 2002).

The seven original objectives of Head Start put forth in 1965 were:

(a) improving the child's physical health and physical abilities; (b) helping the emotional and social development of the child; (c) improving the child's mental

> processes and skills; (d) establishing patterns and expectations of success; (e) increasing the child's capacity to relate positively to family members and others, while at the same time strengthening the family stability and capacity to relate positively to the child; (f) developing in the child and his or her family a responsible attitude toward society, and fostering constructive opportunities for society to work together with the poor in solving their problems; and (g) increasing the sense of dignity and self-worth within the child and his family" (Hodges & Cooper, 1981, p. 225).

Knowledge of these objectives, some of which are affectively and emotionally based, are of high importance considering that today most measures of Head Start's effectiveness are based on school readiness and academic achievement. In fact, the Office of Head Start itself described its program as a service "with a special focus on helping preschoolers develop the early reading and math skills they need to be successful in school" (Wilkins, 2007, p. 1). This significant shift in thought from the original purpose of the program reflects today's thirst for accountability and of the high demands placed on the nation's schools and children.

Who is Eligible for Head Start?

Head Start is available to preschool-aged children from economically disadvantaged families. "Federal guidelines require that 90% of the children served be from families with incomes below the federal poverty line; recently, more than 95% of children served have been below the poverty line" (U.S. DHHS, 1993 as cited in Currie et al., 1994). Early Head Start was established in 1995 to accommodate the nation's neediest children from birth to three years. Head Start services were granted based on

financial need and not contingent on race, ability, or health despite that in 2005-2006 12.1% of Head Start enrollment consisted of children with disabilities, 30.7% of attendees were African American, and 34% of attendees were Hispanic/Latino (Office of Head Start FY 2007). In addition to this, some "typical" students or those who are not disadvantaged attend Head Start programs for a small cost. This is to increase heterogeneity, provide mixed-ability class conditions, and provide positive role models for the neediest children. It is also because, in many instances, Head Start services are provided by private agencies which are granted funds from the government. Furthermore, many non-Head Start federal, state, and locally funded pre-kindergarten programs have been modeled after the Head Start curriculum and program parameters.

Benefits of Head Start

Since 1965, countless studies have been conducted on Head Start, and Head Start-like programs. Many studies resulted in praise of the program for its short-term benefits. Although very few, if any, have gone as far as Zigler and Muenchow (1992) to claim that Head Start is "America's Most Successful Educational Experiment", many researchers have shown that Head Start had an immediate impact on attendees' cognitive development and academic achievement, and others have credited the program with longer-term effects related to success in life as measured by societal gains. The following is a brief review of several such studies.

In 1997, Head Start began conducting a study termed "the Family And Child Experiences Survey (FACES)". This study was a random national sample of Head Start projects which detailed characteristics, experiences and outcomes of children and families who participated in Head Start programs. The 1997 study was followed by a

more thorough study in 2000 when researchers collected data on a national cohort of 2,800 Head Start attendees and their families from 43 programs. The 2000 FACES researchers used assessments of children, interviews with parents, teachers, and staff, and classroom observations as their data sources.

Researchers who conducted the 2000 FACES study reported on attendees' cognitive and social development along with numerous social findings. In general, Head Start students did show improvements in most areas studied. Importantly, "The gap between Head Start children and other preschool-age children narrowed during the Head Start year, especially with respect to vocabulary knowledge and early writing skills" (Zill, et al., 2003, p. 15). Additionally, "Gains of between a third to more than half a standard deviation were observed in vocabulary, early math, and early writing skills during kindergarten" (Zill et al., 2003, p.16). Despite these outcomes, researchers from this study readily admitted that Head Start attendees remained behind their more advantaged peers in early achievement and below national averages. This finding is not surprising and may even be expected when studying the nation's neediest children (Zill et al., 2003).

Similar to the FACES study, researchers conducted longitudinal studies on preschool children in Salinas, California and in Tulsa, Oklahoma. Although not conducted on Head Start programs per se, these preschool studies did find significant evidence to support the need for schooling at this age, especially for students living in poverty. According to these studies, "The achievement gap can be closed for children of poverty and preschool offers one of the best methods to achieving that goal" (Slaby, Loucks, & Stelwagon, 2005, p. 5), and "A program that targets four-year-olds, like the Oklahoma pre-k program, can have positive effects on the cognitive, language, and motor

skills of young children, especially disadvantaged children" (Gormley & Gayer, 2007, p.30).

Head Start Fade

Not all studies suggest that Head Start is as effective as is claimed in the studies cited above. Most researchers acknowledge that Head Start appears to make an educational impact early-on but argue that these benefits quickly fade. The most common assertion is that short-term effects of early intervention programs on cognitive development are apparent, but that these effects quickly decline after children leave the programs and are eventually lost all together.

The origin of the term "Head Start Fade" is credited to the Westinghouse Learning Corporation's study of Head Start in 1969 which stated, "Summer programs were found to have no lasting impact. Full-year programs resulted in cognitive and language arts gains at the first grade level but appeared to "fade out" by second or third grade" (McGroder, 1990, p. 2). Since then, this phraseology has become well-known and is commonly seen in preschool and early intervention studies.

To respond to the idea of fade, supporters of Head Start or similar publicly funded programs, such as Barnett (2002), have pointed out flaws in the methodologies of studies, conducted cost-benefit analyses, and have drawn attention to the long-term gains of those who have participated in preschool. Others, such as Clarke and Achilles (2008) and Lee and Loeb (1994) argued that many elementary schools' follow-up to early interventions are flawed. These flaws may result in losses in early cognitive gains, not in the preschool programs themselves. These researchers also contended that if elementary schools follow

current theory and scientific-based practice the fade effect can be and should be eliminated.

In 1982, Magidson and Sorbom explained that preexisting differences and the absence of truly equivalent comparison groups often skewed the outcomes of studies conducted in the social sciences. They attributed differences in socioeconomic status (SES) between the comparison group and the Head Start children in the Westinghouse study as one example of this. They explained, "Although the comparison children outscore the Head Start children on each of the two tests, they are also higher on each indicator of SES. Thus, it seems reasonable to conclude that, if pretest data were available, it would similarly show the comparison children outscoring the Head Start children, even before the Head Start experience" (p. 323). Because of this phenomenon, Magidson and Sorbom pointed out, *"one must rely on theory to help interpret the results"* (p. 321) (emphasis added). This argument can be, and has been, used to quell the idea of Head Start Fade and to seek and test refinements in the theories. Many researchers point to the fact that Head Start serves the nation's neediest and most impoverished students who have no equivocal comparison group, making the measurement of gains extremely difficult.

In 2002, Barnett argued that the Head Start Fade is completely a myth. He contended that "Head Start studies have systematically erred in the collection of achievement test data in ways that caused the appearance of fade-out," and that "Studies that do not have these design flaws find persistent effects on achievement test scores together with effects on grade repetition, special education, and graduation" (p.1). Barnett has consistently argued that substantial evidence is already present in studies that

are done correctly and when the results are examined properly. In 1998, he conducted a thorough review of preschool literature specifically to address this issue. In his report, he used 38 studies whose researchers estimated effects of early childhood education programs (before age 5) on the cognitive development or school success of children in poverty at least through grade 3. Barnett found that "In many studies, effects on achievement *appear* to fade out, but this is primarily due to substantial and selective attrition in follow-up that reduced the statistical power to detect effects and biases estimates" (p. 205). In studies that did not suffer from those methodological flaws, Barnett found lasting effects in both experimental and quasi-experimental studies. In addition, he saw "sizable effects" in true experiments with adequate sample sizes and minimal attrition (p.205). Along with this, Barnett's review discovered that "There is highly uniform evidence of long-term positive effects on school success as measured by rates of grade retention, special education, and high school graduation" (p. 205).

The inconsistencies related to Head Start Fade draw attention to the need for additional studies. Barnett's points about flawed studies, error in collection, and attrition need to be given credence and addressed in future studies. Researchers need to examine data from groups where participants in the study are chosen based on full participation from beginning to end (pre-k through grade 3). This will eliminate problems stemming from attrition, lack of available data, and other data-collection problems. Researchers should also employ the use of multiple well-designed tests of statistical significance and practical significance. This will reduce the chance of flawed methodology pointed out by Barnett. Likewise, studies need to take into consideration the differences in SES among participants and attempt to minimize this limitation. Conducting a site-based study where

participants are all residents of the same district and come from similar, albeit not identical, economic backgrounds may reduce this as a factor. Also, selecting matched pair groups in a study with attention to SES will minimize this as a limitation. Furthermore, future research must be designed to address the idea of inconsistent follow-up conditions, as proposed by Clarke and Achilles (2008). As a result of meeting these parameters, future studies can eliminate the factors most frequently cited as causes of Head Start Fade and more fairly explain the phenomenon.

Long-Term Benefits vs. Costs

Currie and Thomas (2004) studied the effects of Head Start using the National Longitudinal Survey of Youth and the National Longitudinal Survey's Child-Mother file. These researchers, who analyzed a national sample of children, found large and significant gains on test scores associated with Head Start attendance. The researchers found that African American students quickly lost their gains, while whites retained many benefits. However, in other papers the researchers are quick to point out that beyond test scores, Head Start was associated with long-term benefits for both races. The study provided evidence that whites who attended Head Start had increased chances of completing high school, attending college, and earning more money in adulthood. African Americans who participated in Head Start were less likely to have been charged with a crime. Additionally, the study showed an increase in college entrance test-taking among African American students which reduced the existing gap by 54% (Garces, Duncan & Currie, 2002). Results such as these, which show significant data on the cost effectiveness of quality education beyond test scores, have led to many studies conducted on the value of public preschool programs and early education. Numerous examples can

be found in the book, *The Price We Pay: Economic and Social Consequences of Inadequate Education* (Belfield & Levin, 2007), all of which illustrate the impact on society and the importance of receiving quality schooling regardless of the price.

Using information on long-term effects, researchers claim that money spent on early childhood interventions are worth the large investment in the long run. Currie (2001) stated, "A simple cost-benefit analysis suggests that Head Start would pay for itself in terms of cost savings to the government if it produced even a quarter of the long-term gains of model programs" (p.1). The model programs which are referenced are the Perry Preschool Project, the Abecedarian Project, and the Chicago Child Parent Centers. Ludwig and Phillips (2007) went beyond suggesting this point. They contended that even if Head Start Fade does exist, the benefits shown in multiple studies, including Currie and Thomas's, result in benefits that greatly outweigh the costs. They drew on statistics which demonstrated that, among Head Start attendees, Hispanics are more likely to complete high school, African Americans are less likely to be arrested, special education placements are decreased, and grade retention is lessened to support their claim. Using precise, detailed calculations, Ludwig and Phillips showed mathematically that a cost of $9,000 per child will be sufficiently exceeded in terms of short- and long-term benefits.

In a similar study, researchers pointed out that "Whites who attended Head Start are, relative to their siblings who did not, significantly more likely to complete high school, attend college, and possibly have higher earnings in their early twenties. African Americans who participated in Head Start are less likely to have been booked or charged with a crime" (Garces, et al., 2002, p. 999). Likewise, in a study on juvenile delinquency,

Mann and Reynolds found that "Preschool intervention was associated with reductions in the incidence, frequency, and severity of juvenile delinquency by age 18" (2006, p.153).

Bracey and Stellar (2003) compiled and aggregated data from the three largest preschool studies - High/Scope Perry Preschool, The Abecedarian Project, and The Chicago Child-Parent Center Program. In their study, they found "strong evidence for long-term positive outcomes for high-quality pre-school programs" (2003, p. 780). Even though the data were gathered from so-called model programs, "Findings demonstrate the importance of early intervention and schooling factors in reducing delinquency and highlight the benefits of early intervention" (Mann & Reynolds, 2006, p. 153).

In 2003, the National Institute for Early Educational Research (NIEER) compared the costs and benefits of providing children with high-quality preschool programs. Like many others, this study was conducted using data from so-called model programs, not specific Head Start schools. Nonetheless, "The NIEER researchers found that such programs truly pay off, with every dollar paid out generating a four dollar return to the children, their families, and society" (Barnett, March 2003, p. 1). In addition, the NIEER researchers found that "school districts can save more than $11,000 per child on special or remedial education" (p.1) by enrolling students in high-quality early intervention programs. Similar results have been well-documented by class size studies conducted in the early grades, such as in Tennessee STAR.

After extensively reviewing the studies above and similar cost-benefit studies, Temple and Reynolds (2007) have drawn the following conclusions:

> "The consistency of the findings of the economic effects of these programs despite major differences in social context and location, time period, and

curriculum approach are encouraging evidence in favor of expanding preschool access. Unlike a decade ago, scientific support for the benefits of preschool programs is strong. A major conclusion from these findings for early childhood policy is that for the first time a critical mass of evidence exists that preschool programs have comparatively high levels of cost effectiveness" (p. 142).

Clearly, examining Head Start, and similar programs which are publicly-funded, using the aforementioned cost-benefit studies and analyses highlights the effectiveness and worth of such a large expenditure and persuades legislators and educators to support public pre-kindergarten. However, not everyone agrees with these findings or interpretations of the data, as will be seen in the following section of this study.

Numerous cost-benefit analyses of early intervention programs appear to support the funding of public preschool. Most studies, however, do not test, or even address, benefits versus costs. The focus of the majority of studies is based on achievement as measured by standardized test scores or performance comparisons to peers. Perhaps using data from existing studies, following the participants beyond third/fourth grade, and determining a cost-benefit analysis is a worthwhile endeavor. As such, it will be included as a suggestion in the further research section of this study.

Dissenters

Despite the evidence presented by the aforementioned researchers, there are still those who discredit the effectiveness of Head Start and question the justifications for the expenses the program incurs. In 1992, Hood claimed that "Head Start's popularity is due more to slick salesmanship and superficial thinking about childhood development than to proven success" (p.1). Hood also claimed that there was no evidence that Head Start

provided any long-term benefits. In his critique of the research, Hood (1992) contended that studies to the contrary of his viewpoint examined model programs (like Perry Preschool, Abecedarian, and Chicago Parent Child Centers) which are unlike the actual Head Start programs that are provided to most students in the nation. "Policymakers have gotten the wrong impression about Head Start by listening to enthusiastic boosters who cite the success of model preschool programs as though it proved the efficacy of Head Start. The distinction between studies of Head Start and those of other preschool programs is crucial – all preschools are not created equal" (1992, p.2). Hood defended these claims by saying that using a model program which served hundreds to defend spending on public programs which serve hundreds of thousands is irresponsible. "When dealing with complex issues such as child development, researchers and policymakers must seek out a consensus – not simply hype a few best cases," Hood explained (p.4).

In this evaluative and provocative article, Hood asserted that money spent carelessly on Head Start would be better used in the forms of vouchers and/or tax relief: "Policymakers should convert Head Start funds into direct grants to families, thus allowing poor parents to choose among care providers" (p. 9). " If the federal government converted the amount of money spent on Head Start into vouchers – which would significantly defray the cost of attending most private schools – each year as many as 1.1 million poor children would have the chance to get a decent education in a local school of their parents' choice" (p. 9). Hood's concluding arguments were not against helping the nation's neediest students get a "head start". They were, rather, against the government's involvement in funding, implementing, and administering a public program which private institutions could provide for less cost and, in his opinion, with better

services. Hood did not, however, offer evidence showing that vouchers or tuition tax credits will correct the problems with Head Start which he brought to light in his critique.

In a study on the impact of intergenerational Head Start participation, Caputo (2004) found that "Graduates of Head Start appear roughly comparable to other adolescents in regard to highest grade completed, a sense of mastery, perceived health, and levels of depressive symptoms. They do not attain the levels of achievement as other adolescents in regard to reading comprehension and years of living above the poverty level" (p. 199). At first glance, data from this study appear entirely negative and unsupportive of Head Start. But, in the interpretation of the data, Caputo conceded that it is impossible to determine where these students would have ended up without Head Start. It is indeed possible that measurable gains were attained to bring these needy inner-city children up to par with their peers in some areas. Also, although still below their peers in other areas, the results are most likely higher than they would have been without early intervention. Thus, Caputo's study does not entirely discount the value of Head Start. It does, however, raise questions regarding the program's effectiveness and worth which require further study.

In more direct and poignant fashion, Haskins and Rouse (2005) proclaimed that "Studies consistently show that poor and minority children have already fallen behind before they enter the public schools. "Unless one believes that this poor performance is due entirely or primarily to genetic factors, it follows that the preschool environments of poor and minority children are deficient in supplying the types of experiences that promote school readiness" (p. 2). In their paper, Haskins and Rouse drew attention to the failures of Head Start and similar publicly funded programs by pointing out their

shortcomings. Afterward, they conceded that public funds should be diverted to pre-kindergarten programs to help the nation's impoverished children but that *existing programs should rely more on sound theory and research-based practices (such as small class sizes and duration of program)* (emphasis added), follow more stringent guidelines, and be subject to more scrutiny - similar to the model programs: Perry Preschool and the Abecedarian Program.

In a 2004 paper written for the Heritage Foundation, Kafer explained "Nearly four decades since Head Start was launched, the school readiness gap between poor children and their middle-class peers remains stubbornly large. There is no clear evidence that these programs have helped poor children gain any advantage that can be maintained over time" (p.2). While Kafer did not go as far as calling Head Start useless, the somber words begged the question of whether any program or any amount of money spent on schooling could counteract the effects of being raised in extremely impoverished households and neighborhoods. Because of this question's value and import in the discussion of Head Start, the question is explored further in this paper.

After more than 40 years of Head Start's primary implementation and hundreds of studies on it, one thing remains clear: There is still no consensus on the program's effectiveness. Viewpoints vary depending on which study one reads, what prior bias one brings to the literature, and what data one chooses to accept, regardless of the quality. Therefore, future studies must address these inconsistencies to add valuable information and data to the knowledge dynamic. Studying "real world" programs in future studies, and not model programs administered under experimental conditions, are one way to accomplish this goal. Findings from such studies will not be subject to the criticism

found in Hood's critiques about studying "best cases" (1992). Future studies should also be designed to test existing theory and evidence from scientific-based research. This will eliminate the critique that the findings of fade result because the programs themselves, and the follow-up conditions that the children are subjected to, are not adequate as was proposed by Haskins and Rouse (2005).

Influence of Poverty on Achievement

Perhaps more than any other factor, poverty has been shown to have a tremendously negative effect on education and achievement. Children in schools located in low SES areas suffer from poor air quality, inadequate libraries, poor technology, old and outdated materials, high teacher turnover, and poor home-school relationships (Kozol, 1992; White 1982; Lee & Burkham, 2002; Sirin, 2005; Quinn, 2005). There are obvious reasons why children from impoverished neighborhoods are referred to as "at risk"; unfortunately, obvious solutions are seldom used in the schools. Bracey (1999) explained, "Poverty, like gravity, is a condition. Gravity acts upon people in profound ways. So does poverty. To overcome the effects of poverty will require more effort than we are now making" (p. 511). Greater attention to theory and research can lead to the application of better practices, such as early intervention, duration, intensity, and heterogeneity, and the attainment of better results. This concept will be explored in depth later in the chapter.

In 2005, Sirin conducted an empirical meta-analysis review of the effects of SES on academic achievement. The analysis included data from 101,157 students, 6,871 schools, and 128 school districts gathered from 74 independent samples (Sirin, 2005, p. 417). Using Cohen's (1977) guidelines, the overall effect size of the study reflected a

medium level of association between SES and academic achievement at the student level and a large degree of association at the school level. As Sirin explained, "Of all the factors examined in the meta-analytic literature, family SES at the student level is one of the strongest correlates of academic performance. At the school level, the correlations were even stronger" (Sirin, 2005, p. 438). Overall, the findings reflected the significant effect that social capital has on academic achievement of students and illustrated the need for public policy to address SES as a key factor affecting schools (and society as a whole).

In 2006, Berliner also drew attention to the fact that among the lowest social classes, environmental factors, particularly family and neighborhood influences, are strongly associated with academic performances. He suggested that the most effective policy for improving student achievement is a reduction in family and youth poverty. Although his paper supported education programs to help impoverished children, one of them being publicly funded preschool, Berliner argued that "Schooling alone may be too weak an intervention for improving the lives of most children now living in poverty" (p. 955). He argued that legislators had to do much more than fund school-based programs if they wanted to raise academic achievement of the nation's poorest students.

Drawing on data from numerous studies, Berliner explained, "Although the power of schools and educators to influence individual students is never to be underestimated, the out-of-school factors associated with poverty play both a powerful and a limiting role in what can actually be achieved" (p. 950). With an understanding of these points and drawing upon similar sentiments, supporters of Head Start claim that the program's early interventions are aimed at counteracting the environmental influences of a poor

neighborhood and providing a chance for children whose biggest problem is that they are not wealthy. The contention was that educators cannot, and should not, give up because of the obvious uphill battle they face.

Berliner's claim that "School reform is heavily constrained by factors that are outside of America's classrooms and schools" (p. 950) is well-supported by data. Even a cursory glance at the New Jersey School Report Card shows that the schools in the lowest DFGs (A) suffer from the lowest standardized test scores while schools in the highest DFGs (I&J) celebrate success (NJDOE, 2008) (information available in Appendix B) (Tienken, 2008). Still, the question of what to do with impoverished children remains for educators. Legislators and government officials have yet to break up the concentration of poverty that exists in America's urban and rural areas, and provide the panacea for what is arguably the largest plague facing the education system. While the wait continues, Head Start and publicly funded preschool remain viable options available to educators who try to help these impoverished victims.

In 2005, Lewis also addressed the impact of poverty on American schools and acknowledged its profound effect on achievement. In his commentary on the need for quality pre-kindergarten interventions, Lewis pointed out that a substantial gap in readiness for learning exists in kindergarten and stems primarily from income and race. Lewis, too, conceded that it is a stark challenge for the best-intentioned teachers and school leaders in the primary grades to remedy the problem (p.1).

Acknowledgement of the impact poverty has on education can be found in abundance (Kozol, 1992; White 1982, Lee & Burkham, 2002; Sirin, 2005; Quinn, 2005), and many of the concessions follow with an acknowledgement of the fact that schools

alone cannot fix the problem. Some of the many examples include: (1) "Twenty-two years after the creation of the preschool program for low-income children, its cofounder, E. Zigler, acknowledged, "We simply cannot inoculate children in one year against the ravages of a life of deprivation" (Kafer, 2004, p. 1) (2). "Perhaps no government program can ever sufficiently make up for what a hard life takes away" (Kafer, 2004, p. 4) (3). "The problem of underachievement by poor and minority students has confounded us. High-level commissions issue warnings, governors hold summits, think tanks produce reports, scholars write books, and Congress passes laws. But the U.S. has failed to deliver on its promise to provide a high-quality education to every child" (Boyd-Zaharias & Pate-Bain, 2008, p. 40).

This review of education and economic literature clearly illustrates that poverty is a pandemic in education. Multiple societal and legislative changes must be made to address the issue and assist educators in improving the achievement of all. Head Start and school reform projects alone will not provide all of the answers for helping impoverished students. This paper, however, will draw attention to scientific-based practices and theory which have proven worthy of attention and financial investment and attempt to explain what interventions could assist with remedying the impacts of poverty.

Effective Early Intervention Programs

Within the body of research on early intervention programs, four programs have repeatedly been found to have statistically significant lasting influence on the academic achievement and social development their participants. These studies are: The High/Scope Perry Preschool Project, The North Carolina Abecedarian Study, the Chicago Child-Parent Centers, and the Michigan School Readiness Programs. All four of these

studies examined programs that were administered with strict attention to structure and intervention implementation, as suggested by Haskins and Rouse (2005). They also employed well-designed study methodology, data collection, and instrumentation under experimental conditions to validate their results as was suggested by Barnett (2002). A review of these four programs supports Barnett's claims that correctly-done studies, with properly examined results, clearly demonstrate the positive and significant influence of quality early childhood education on its participants. In addition, all four of these studies have shown that Head Start Fade does not occur when early interventions are administered correctly. These findings have been shown by numerous follow-up and longitudinal studies as outlined in the following section.

The High/Scope Perry Preschool Program

One of the most renowned, and most often referred to, early childhood education programs in education research is the High/Scope Perry Preschool. The project was developed by the Division of Special Services of the Ypsilanti School District in Michigan starting in 1962. Because of the year it began, similarities in design, and a plethora of data related to the program, the success of Perry Preschool's attendees is often used to defend the funding of Head Start and other public pre-kindergarten programs in the United States. However, it is also commonly referred to as a "model" program with stark differences to Head Start, in terms of program design, which are not delivered to the typical Head Start student today.

The High/Scope Perry Preschool Program served 58 African American children, 3-4 years of age, from low-income homes and deemed at risk of school failure because of environmental factors and low IQ scores. The children

> participated in the program for approximately two years. In addition to defined classroom activities, teachers visited the children's homes weekly and had monthly meetings with parents (Schweinhart, 2004, p. 1).

Since the program's inception, Perry Preschool has gained national attention as an exemplar of high-quality preschool. Numerous studies have been conducted on the program's participants, producing reliable, relevant, and significant data. This illustrates why the program deserves accolades.

First, the study is revered for its quality and strength. As one of the pivotal researchers involved with Perry Preschool studies explains, "The High/Scope Perry Preschool Study is one of the most convincing studies of the long-term benefits of good preschool programs for young children living in poverty. This study has three essential strengths: random assignment of its 123 study participants to a program group and a no-program group; virtually no attrition of study participants; and a plausible, consistent pattern of causes and effects from preschool to adulthood" (Schweinhart, 2000, p. 136). These sentiments are generally accepted in the field of education research and very few, if any, have argued with the design of the original Perry experiment or any of the follow-up studies which have been conducted.

In the original study, pre- and post-IQ tests were administered along with kindergarten readiness tests to determine an initial impact of the program. As was expected, the program had an immediate impact on the students. Consistent with most studies of students who attend preschool, immediate gains are noticeable and significant. What has gained the Perry Preschool much of its notoriety, however, is the body of

follow-up studies which have shown significant and lasting impacts on students who attended the program when compared to students who did not participate.

The first notable such study took place several years after the students completed preschool and the following sentence explains the findings:

> The Perry study of the effects of the preschool program on children through age 10 found that fewer children who had been enrolled in the program were held back a grade or placed in special education than children who had not been enrolled in the program and that there were consistent, nearly significant program effects on achievement test scores from first through fifth grades (Schweinhart, 2002, p. 1).

Researchers involved with the Perry study were optimistic about their findings and about the long-term impact of participation in a quality preschool program.

The researchers conducted a second follow-up study several years after the first study. In that study, the researchers found "a substantial program effect on achievement test scores for 14-year olds." "The effect was actually bigger than it was for children in their earlier years, and it definitely was statistically significant" (Schweinhart, 2002, p. 2). Not only did the researchers find promising results in their years of work, they found compelling reasons to continue conducting follow-up studies on the program's participants and the long-term effect of the program itself.

Nearly 25 years after the study began, researchers tested yet again the long-term impact of participation in the Perry Preschool Program. A study of participants at age 27 showed that "The lifetime economic benefits to the preschool program participants, their families, and the community far outweigh the economic cost of their high-quality, active

learning preschool" (Texas Youth Commission, 1993, p. 1). These findings were based on significantly higher monthly earnings, percentages of home ownership, level of schooling completed, of significantly lower percentage of participants receiving social services, and significantly fewer arrests.

Determined to increase the data supporting preschool attendance, the Perry researchers conducted yet another study. The age 40 follow-up report showed that:

> "As adults, the preschool group was employed at higher rates, had higher incomes, enjoyed more stable housing situations, owned more automobiles, and were more likely to have a savings account. Also, the preschool group received fewer social services than the no-preschool group, had considerably fewer arrests, and much less drug abuse was evident" (Manning & Patterson, 2006).

In more recent monographs, available from the High/Scope Press (2009), researchers have examined Perry Preschool data for effect size. They report effect sizes of school achievement tests to be 0.28 at age 8, 0.29 at age 9, 0.34 at age 10, and 0.49 at age 14 (Schweinhart L. J., 2009). These findings support the statistically significant findings in earlier studies by showing a practical significance ranging from small- to medium-effect size differences with the strength increasing over time. This data not only negates the idea of Head Start Fade, it shows that the influence measured by academic gains increased over time.

These data show why and how the Perry Preschool Program and the related follow-up studies have gained so much weight in education research. This program is

regarded as a strong experiment and its results are impressively positive and enduring. Nonetheless, many argue that the findings are not applicable to Head Start or to any publicly funded program because of Perry's experimental design, which is difficult, if not impossible, to replicate in the "real" world. The following excerpt was taken from literature which generally supported public preschool: "The randomized controlled trial shows major impact on education and life outcomes; we note, however, that this was a demonstration project, and it is not yet known if the results can be replicated on a broader scale in typical classroom settings" (Coalition for Evidence-Based Policy, 2002-2004, p. 1). Thus, further studies on public programs are warranted, necessary, and justified.

North Carolina Abecedarian Study

Like the Perry Preschool program in Michigan, the Abecedarian program in North Carolina has gained popularity and notoriety in the field of education research. Evidence of this claim can be found throughout the literature. The following quote is one representative sample: "The Perry Preschool Program in Ypsilanti, Michigan, and the Abecedarian Program in Chapel Hill, North Carolina, have been arguably the nation's best model programs" (Haskins & Rouse, 2005, p. 3). Although different in program delivery, purpose, and design than the Perry Preschool, the Abecedarian program has won as much acclaim for demonstrating the value of early childhood education.

In addition to studying preschool benefits to children, the Abecedarian program found that child care, along with early childhood education, benefits the mothers in several measurable and relevant ways: "The Abecedarian program is the only randomized trial of child care with a longitudinal follow-up to adulthood" (Barnett & Masse, 2007, p. 114). Abecedarian results not only support findings on early intervention

first shown in the Perry study, they add a significant body of evidence to the knowledge dynamic by providing data on child care.

The following comprehensive overview of the Abecedarian project was given by Star in 2002:

> "The most thorough study, called the Carolina Abecedarian Project, followed 111 disadvantaged North Carolina kids for 21 years. Half were enrolled in a high-quality educational program (full-day and year-round with low child/adult ratios ranging from 1-3to 1-6) from infancy to age five, while the control group got only nutritional supplements. All the children attended comparable public schools from kindergarten on. The result: Those who attended preschool were less likely to drop out of school, repeat grades, or bear children out of wedlock. By age 15, less than a third had failed a grade, vs. more than half of the control group. At age 21, the preschoolers were more than twice as likely to be attending a four year college" (p. 98).

Star's summation provides a broad and encompassing overview of the value of Abecedarian's findings.

In more detail, Barnett and Masse explained that "Early assessments indicated substantial early gains in IQ and achievement and the most recent assessment at age 21 found continued effects on IQ and achievement. Effects on school success include much lower levels of grade retention, placements in special education classes, reduced high school dropout, and a higher rate of attending a 4 yr. college at age 21" (2007, p. 116). Akin to the Perry Preschool study, researchers involved with the Abecedarian program

and subsequent follow-up studies found statistically significant gains in early cognition and school readiness as well as lasting lifelong benefits.

Along with finding benefits for the children who attended the Abecedarian program, researchers conducted a cost-benefit analysis with relation to the children's mothers and found that the day care provided through the program more than paid for itself. Barnett and Masse found that "The program passes a basic benefit-cost test at discount rates of 3-7%. Given the estimated net present value at 7% and the benefits we were not able to include in the analysis, the internal rate of return to the program could be considerably higher (2007, p. 122). Barnett and Masse explained that the children chosen for the Abecedarian program were from the area's neediest families and were deemed at high risk for school failure. Consequently, the children's mothers were also at high risk for living lives dependent on social services such as welfare and Medicaid. As a result of the Abecedarian participants' mothers being provided with full-day child care, along with free preschool, they were able to increase their own level of education and/or find work with higher wages. Therefore, the benefits greatly outweighed the cost. In essence, the study found that it is simply cheaper to pay for day care and preschool for several years than to pay for a lifetime of social services which these children and their mothers would have most likely required absent the Abecedarian program.

The cost-benefit findings detailed by Barnett and Masse are echoed throughout the related literature. Pungello explained that "The results of the Abecedarian study demonstrate that high-quality child care can have long-lasting benefits for children. The results also demonstrate that the provision of such care can have both educational and vocational benefits for teen mothers. Whereas the early intervention itself appears to

affect the developmental trajectories of the young children, having reliable full-time care appears to affect the developmental trajectories of teen mothers as well" (Pungello, et al., 2000 p. 3). These findings not only confirmed previous data on the value of preschool, but they also added increased evidence in terms of child care benefits and changed the conversation to include more services.

In addition to the previously explored findings, researchers studied the Abecedarian program to address the phenomenon of fade. In 1994, Campbell and Ramey found that "In contrast to the report by the Consortium for Longitudinal Studies that early IQ gains eroded within 3 years of school entry, and academic gains within 5-6 years, the intellectual and academic gains from the Abecedarian program persisted through 7 years of school. In fact, the Abecedarian preschools treatment/control IQ difference is slightly more pronounced at age 12 than at age 8. The critical point to be made from the Abecedarian longitudinal IQ results is that, from infancy through age 12, subjects having preschool treatment maintained an IQ advantage over those without the early treatment" (Campbell & Ramey, 1994, p. 694). These results contradicted the data found in earlier studies of Head Start programs. This led to studies of the "model" Abecedarian program to determine what conditions contributed to the retention of early gains. One of the most glaring conditions was small class size, which is discussed further later in this literature review.

According to Campbell and Ramey, "The most important policy implication of these findings is that early education intervention for impoverished children can have long-lasting benefits, in terms of improved cognitive performance" (1994, p. 695). This may indeed be true. But like the Perry Preschool project, the Abecedarian program was

delivered under strict control conditions, to a distinct population (over 90% African Americans) who were identified as extreme high-risks for school failure (IQ scores near or less than 70). These facts have led many to question the ability to generalize these findings and apply them to public-funded preschool or Head Start programs. Such is the case with "model" programs, as was evident in the review of the Perry Preschool project. This is another compelling reason to study a public preschool with heterogeneous groups of students who participated in a program under typical unadulterated conditions.

Chicago Child-Parent Centers

The Chicago Child-Parent Centers (CPCs) are center-based early intervention programs which provide comprehensive educational and family support services to Chicago's economically disadvantaged children and their parents. The program began in 1967 with funding from the Elementary and Secondary Education Act (ESEA) (P.L. 89-10) of 1965 and continued to be federally-funded in 2009. Much like the High/Scope Perry Preschool Project and the Carolina Abecedarian Project, the Chicago CPCs have become well-known early intervention programs which have produced positive enduring results for its participants. One distinction, however, is that the Chicago CPC program is a large scale federally funded program, whereas the other two were model programs.

An ongoing investigation of the Chicago CPCs, which has involved numerous researchers, has been taking place since 1985. The data are derived from "a complete cohort of 1,539 low-income children (93% African American) who participated in CPCs beginning in 1983 and 1984, and a comparison group of children the same age who enrolled in alternative kindergarten programs without CPC preschool experience. The 989 program participants and 550 comparison-group participants in this matched-group,

quasiexperimental design were born in 1980, resided in high-poverty neighborhoods, and attended Chicago public schools" (Reynolds et al., p.636). As the study progressed, researchers have collected and analyzed information on child and family well-being, standardized test scores, surveys and interviews, social service records, and justice system records. Data from this study, known as the Chicago Longitudinal Study (CLS), have been used to highlight the Chicago CPCs successes in terms of academic gains, long-term social benefits, and cost effectiveness. This is particularly significant because the Chicago CPCs are government-funded, the participants were not hand-selected, and the sample size in the study was large.

In 2003, Reynolds, Temple, and Ou examined data from the CLS and found significant evidence of positive gains. "Using data from 1,539 children in the Chicago Longitudinal Study (CLS), preschool participation was associated with higher levels of school readiness, achievement, and educational attainment, and with lower rates of child maltreatment, juvenile delinquency, special education placement, and grade retention (p. 633). In addition, the researchers found that "Every dollar invested in the preschool program returned $7.14 to society at large" (p. 633). In concluding the study, the researchers stated that their findings presented strong evidence that large-scale, public, early intervention can enhance children's well-being if it is offered effectively and use elements similar to the ones used in Chicago CPCs.

> In a 2003 study, researchers examined the CLS data and determined that: CPC preschool participation was associated with significantly higher levels of school readiness at kindergarten entry. About twice as many program participants as comparison participants scored at or above

> national norms on the cognitive composite of the Iowa Tests of Basic Skills. A similar pattern occurred for reading achievement over the school-age years (Reynolds et al., 2003, p. 641).

These data highlight the early cognitive gains which can be obtained from participating in an effectively delivered publicly funded early intervention program. The question that remains, however, is whether these findings can be translated to Head Start as it currently exists.

Additionally, researchers discovered from the CLS data that lasting impacts of effective early intervention programs can manifest themselves in a variety of ways and eliminate the need for later costly interventions.

(1) "Program participation was associated with significantly lower rates of grade retention and special education placement. Program participants also spent fewer years in remedial education" (Reynolds et al., 2003, p.641).

(2) "Children who participated in Child-Parent Center preschool had a significantly lower rate of special education placement (12.5%) than the comparison group (18.4%) who participated in an alternative all-day kindergarten program."

(3) "Preschool intervention was associated with reductions in the incidence, frequency, and severity of juvenile delinquency by age 18" (Mann & Reynolds, 2006, p. 153).

(4) "Preschool participation was linked to greater educational attainment by age 21. CPC participants had a 20% higher rate of graduating from high school or earning a GED. They also had a higher mean number of years of completed education" (Reynolds et al., 2003, p. 643).

These findings illustrate the enduring effects that can result from effectively delivered early intervention programs. The findings bring into question the "fade effect" especially

if success is measured in ways other than purely academic achievement and standardized test scores; such as affective gains for students and teachers.

Researchers also examined the CLS data on Chicago CPCs for cost effectiveness, and "The findings suggest that investments in preschool had substantially higher net benefits and benefit-cost ratios than several education, job training, and health service interventions (Temple & Reynolds, 2007, p. 142). "The estimated average cost per child for one year of child welfare services in the Chicago sample is $9,492 (in 1998 dollars), more than twice the cost of one year of preschool. Combined with the demonstrated effect of the program in reducing expenditures associated with special education and juvenile arrests and in increasing educational attainment, the program's cost effectiveness is high" (Reynolds et al., 2003, p. 633).

Along with tests of statistical significance, researchers have conducted tests for the practical significance, effect size, of the CPC's influence on academic achievement. They found that the CPC had an effect size of 0.61 on the school readiness cognitive composite at age 5. Scores for academic achievement were also reported in later grades and ranged from 0.20 - 0.30 (Reynolds, 2000). Using Cohen's guidelines (1988) for interpreting effect size, these results are practically significant and, although they did decrease over time, the data show that participation in the CPC remained to have an influence on the participants' academic achievement in later grades.

This review makes clear that the way in which the Chicago CPCs delivered early intervention and education to poverty-stricken children has been shown to be effective. Even though Head Start programs, however, are also federally funded, they do not meet the requirements of the Chicago CPCs. Therefore, the present researcher will examine if

a Head Start program which includes the elements of early intervention, small class sizes, heterogeneity, intensity, and duration, which are all a part of the effective Chicago CPCs, can yield some of the same positive results.

Michigan School Readiness Program (MSRP)

Like Chicago, cities in Michigan such as Detroit, Grand Rapids, Grayling, Kalamazoo, Muskegon, and Port Huron have provided publicly funded early interventions for children from impoverished families and/or neighborhoods. In 2002, studies were conducted for the Michigan State Board of Education on Michigan School Readiness Program (MSRP) participants to determine if the state-funded preschools were worthy of the over $100 million investment. The results showed that "24 percent more MSRP participants passed the Michigan Educational Assessment Program (MEAP) literacy test for grade four than nonparticipants, and that 16 percent more passed the mathematics test. Study results also indicated that, of the children who participated in the program, 35 percent fewer needed to repeat a grade level, compared to children who did not participate" (Schweinhart, 2002, p. 1).

In addition to the multiple successes found using "hard data", the study also discovered that:

> "Children in the study who completed the state-funded preschool program were significantly more advanced in key areas of development - language and literacy, creative representation, music and movement, initiative, and social relations than non-participants. From kindergarten through fourth grade they were found by their elementary teachers to be significantly more ready than their nonparticipant classmates were – more interested in school, more likely to have good attendance,

to take initiative, and to retain learning; stronger in reading, mathematics, thinking and problem solving skills; and better at working with others. Their parents were also more involved in their children's school activities and talked with their teachers more frequently" (Schweinhart, 2002, p.3).

Similar to the Chicago CPCs, the data on MSRPs show promise for Head Start programs. Like Head Start, the MSRP programs are publicly funded, albeit by the state and not the federal government, and are offered to the neediest children from low-income and single-parent homes. However, the MSRP programs all had class sizes of fewer than 16 students and were administered by a certified teacher and a trained assistant which is not always the case with Head Start programs. This adds further significance to the current study which examined a Head Start program which met most of the conditions of early intervention, duration, intensity, and heterogeneity to determine if similar results can be produced.

Class Size Studies and Pupil-to-Teacher Ratio

Akin to the model preschool studies previously discussed, numerous class-size studies have shown that children can and do succeed if educated under the proper conditions, regardless of their background. One such study is the Tennessee Student-Teacher Achievement Ratio (STAR) experiment which has come to be recognized as the landmark empirical study on class size. "Project STAR, a study of the education effects of class size in the state of Tennessee, is one of the greatest experiments in U.S. history," claimed Mosteller, Light, and Sachs (1996, p. 814).

The STAR study has earned these accolades because of the depth and breadth of reliable data which its participants produced and its significant longitudinal outcomes,

and also because the study included students from impoverished neighborhoods who are traditionally at a disadvantage to begin with. The researchers followed nearly 12,000 students and clearly demonstrated the value of small classes, especially among minorities and impoverished children, with both quantitative and qualitative data. Here is a list of findings from STAR and STAR-related studies:

- Pupils in small classes (S) outperform pupils in regular classes (R) and regular classes with an aide (RA) on all cognitive measures and the early treatment lasts at least into Grade 8 after the K-3 start.
- Pupils in S have relatively fewer examples of poor discipline.
- The S classes seem to reduce the known deleterious effects of big schools.
- Teachers have more "on task" time in S and this stays constant all year, but in R the behaviors decline over the year.
- Students in S are more engaged and participative in school than are students in R and RA.
- There are relatively fewer retentions in grade in S.
- The traditional test-score gap between white and nonwhite pupils does not open as much in S as in R and RA classes on criterion-referenced tests.
- Early identification of special needs in S seems to reduce later special education placements.
- Student scores in S are up in all tested areas, not just in targeted areas characteristic of special projects.
  (Achilles C. M., 1999, p. 28)

As Mosteller (1995) explained, "After four years, it was clear that smaller classes did produce substantial improvement in early learning and cognitive studies, and that the effect of small class size on the achievement of minority children was initially about double that observed for majority children" (Mosteller F. , p. 113). In a similarly succinct statement, Achilles (1999) explained, "We are able to show definitively what many parents and teachers have long known: Small is better, especially in the early years of schooling" (p. 27).

The second phase of STAR was called the Lasting Benefits Study (LBS) which began in 1989. This phase was an observational study of the original participants when

they returned to regular classes in fourth, fifth, sixth grades and beyond. The driving question was whether or not the children who started in smaller classes continued to perform better in later grades; i.e., were the results of early intervention self-sustaining? Results showed that "In the fourth and fifth grades, the children who had originally been in small classes scored higher than those who had been in regular-sized classes or in regular-sized classes with an aide (Mosteller F. , 1995, p. 121).

Project STAR has shown what many have long believed, even without the statistical data. Caulfield, a long time practitioner as Superintendent of Schools, shared what he learned from years of experience in his district: "Class size is the crucial element in instruction. Thus, it is imperative that class size be reasonable to insure opportunities for individualization and reinforcement. Class size is fundamental to success" (1989, p. 60). Data from STAR strengthen this proclamation and common sense argument. As Mosteller concluded from the STAR study,

> The evidence is strong that smaller class size at the beginning of the school experience does improve the performance of children on cognitive tests. Observations from the Lasting Benefits Study confirm that the effect continues into later grades when children are returned to regular-sized classes (1995, p. 123).

Not everyone who reviews class-size studies is convinced, however. Some critics argue that the evidence on class size is limited and highly selective, and that when the data are examined in less controlled environments the evidence is meager and unconvincing. For instance, Hanushek (1998) claimed that "The surprising fact is that the enormous amount of research devoted to studying class size has failed to make a very

convincing case that reducing class size is likely to improve student performance" (p.1). Hanushek also stated, "In order to support calls for class size reductions, there has been a tendency to pick and choose among available studies and evidence" (1998, p. 1).

> To counter Hanushek's claims, researchers have pointed out the flaw in his data: The studies summarized by Hanushek were not studies of class size but of a different construct: the pupil-teacher ratio of schools, districts, states, and countries. Aggregate pupil-teacher ratios do not describe the day-to-day setting in which students are learning; many districts have low pupil-teacher ratios, while most students spend the entire school day, every day, in crowded classrooms (Finn, Pannozzo, & Achilles, 2003, p. 321).

After conducting a careful reading of Hanushek's work, evidence to support these claims can be easily discovered in the writing. In an article where Hanushek directly attacked class-size studies, words which are allegedly used interchangeably with class size are evident and are underlined for emphasis in the following quote: "When combined with data on student performance, however, the wide discrepancies in pupil-teacher ratios show little relationship to achievement (1998, p. 2). It becomes clear that one must recognize the difference between the two terms, class size and pupil-teacher ratio (PTR), when examining relevant studies.

The Tennessee STAR study, although the most highly-acclaimed research on class sizes and their impact on students, is not the only study which has found positive and sustained results. Tennessee's Project Challenge (Achilles, Nye, & Zaharias, 1995) and Wisconsin's Project Student Achievement Guarantee in Education (SAGE) (Molnar,

Smith, & Zahorik, 2000) are additional examples of extensive large-scale randomized or matched studies which have contributed data. In addition:

> Researchers have studied class size in American schools for more than a century. Well over 100 studies were reviewed by Glass and Smith (1978) and Robinson (1990). These authors concluded that the preponderance of the evidence showed that reduced-size classes – below 20 pupils – were associated with improved academic performance. Effects were most pronounced in the early primary grades, and especially among students from low-income homes (Finn, Pannozzo, & Achilles, 2003, p. 321).

These class-size studies, when juxtaposed with the aforementioned studies of model preschool programs - where class sizes were always regulated - are particularly relevant to this paper. The parallel theories and practices which can be derived from the Perry Preschool Project, the Abecedarian study, the Chicago Parent-Child Centers, The Michigan School Readiness Program, and the STAR and STAR-like studies have formed the foundation for the theory of early intervention, duration, intensity, and heterogeneity which is the premise of the present study, or theory test.

Ineffective Class Size Reduction Initiatives

Numerous studies, such as Tennessee STAR and Wisconsin SAGE, have produced valid and reliable data to support the claim that class size reduction (CSR) produces a variety of positive results. The findings range from increased student achievement to decreased discipline problems among participants and include findings that last well beyond the treatment years. Not all attempts at CSR, however, have produced these same results. Many attempts at reducing class sizes have been

unsuccessful at attaining their goal because of improper implementation and/or lack of attention to the findings from the scientific-based research (SBR).

The state of Indiana funded an initiative to reduce class sizes in grades one through three in 1984. The intervention, which was known as Prime Time, took place over three years, beginning with grade one in 1984, adding grade two in 1985, and grade three or kindergarten (school's option) in 1986. The initiative aimed at reducing class sizes to an average of 18 pupils, or to 24 pupils if an instructional assistant, or aide, was in the room.

Indiana Prime Time data revealed that "results for academic achievement were mixed – at times, small classes were found to have superior outcomes and, at times, the large classes performed better" (Finn J. , 1998, p. 3). These findings resulted because of a lack in rigor and attention to detail in the implementation of the CSR initiative. As Finn (1998) explained:

> Prime Time did not implement a single, well defined, small-class intervention. While the average class size of 18 pupils was viewed as a target, actual class sizes ranged from 12 to 31; classes of 24 pupils with a teacher aide were considered to be small despite the number of pupils in the classroom." In addition, "small classes may not have been kept small for the entire school day (p. 3).

Thus, it becomes more evident that not all CSR programs are equal. Reducing class size, and especially class size averages, does not guarantee that achievement will improve. Unless the classes are significantly reduced to manageable sizes, less than 19,

as was evident in STAR and are maintained all day every day for a period of at least four years the data are inconclusive on the type of return that can be expected.

In 1996, a CSR program was adopted in California covering 1.8 million students in grades kindergarten through the third grade. Implementation lasted for at least three years. During this CSR initiative, class size averages were reduced from 29 to 20 students (note that this is an average and not an actual consistent class size number). According to Cobbold (2005), "an ongoing evaluation of the California class size reduction program has found no relationship between state-wide student achievement and class size reductions" (p. 1). Although the California program was inspired by the results of the Tennessee STAR experiment which produced significant achievement gains for all students from CSR, especially for low-income and minority students, the California project was not as successful.

As Cobbold explained, "the California project differs in significant respects from the STAR Project" (p.1). For instance, the STAR Project was a carefully controlled experiment with random assignment of both teachers and students to small and regular class sizes. The California program was implemented state-wide, and conditions varied depending on regional circumstances such as lack of adequate space, facilities, and enough qualified teachers and resources to implement the program as outlined by STAR. The STAR Project involved a reduction in class size from 22-26 to 13-17 while California classes were reduced from an average of 29 to 20. These differences can be attributed to the lack of sustainable results from STAR and its follow-up studies. The California initiative is one example of how simply implementing a CSR program does

not guarantee positive results. If not done correctly and according to theory and research-based practices, there is a chance that positive outcomes may prove elusive.

In a more recent and smaller-scale study conducted at the middle school level, Tienken and Achilles (2009) found that a properly implemented class size reduction initiative had a statistically and practically significant positive influence on achievement. The study also confirmed that implementing a CSR initiative without attending to the research base and knowledge dynamic produced non-significant results. The study which measured the influence of CSR on students' writing achievement found that "CSR had a statistically significant influence on the achievement of students who received CSR for 3 consecutive years" (Tienken & Achilles, 2009, p. 13). These findings were also shown to be practically significant when compared to students who did not receive CSR treatment for 3 consecutive years. Along with these findings, however, the researchers pointed out that study results "demonstrate that 1 year of CSR treatment is not enough to have a statistically significant influence on achievement" (Tienken & Achilles, 2009, p.22). These data, and the findings in this study, add to the existing body of information and illustrate how CSR initiatives in and of themselves may not provide gains in achievement. Only CSR initiatives that were implemented with duration and intensity have been found to have significant and positive results - which is a major part of the basis for the current study.

Theory of Early Intervention, Duration, Intensity, and Heterogeneity

After reviewing the organization of the aforementioned model preschool programs, as well as the relevant class-size studies, four common conditions stand out as conditions for success in raising achievement with low SES students. These conditions

are early intervention, duration, intensity, and heterogeneity. Among these four conditions, small class sizes are an integral part which must not be overlooked. Much like Clarke's study in 2005, it is the intention of this researcher to draw attention to the abundance of existing data and research which support these conditions and highlight the importance of all these conditions being implemented to maximize potential.

Modern class-size studies, such as STAR in Tennessee (1985) and the Wisconsin Sage Project (1996), have shown that reducing the amount of students in classes has a positive and significant impact on the attendees in terms of both achievement and affective measures. To support these findings, all model preschool programs reviewed in this chapter had small class sizes as one of their requirements. The High/Scope Perry Preschool Project had "Small groups to develop closer relationships between the teacher and the child" and "a child-staff ratio of no more than 10 children per adult" (Schweinhart, 2004, p. 135). In the Abecedarian program, "The preschool program was center-based with teacher/child ratios that ranged from 1:3 for infants/toddlers and 1:6 for older children" (Barnett & Masse, 2007, p. 116). In the Chicago Child-Parent Centers:

> To maximize individual learning opportunities, preschool class sizes are small, and each classroom has a teacher's aide in addition to a regular classroom teacher. The average teacher-to-child ratio is 1 to 8. The smaller class size allows for a child-centered, individualized approach to language development, cognitive development, and improving social relations (Griffin, 2009, p. 1).

And, similarly, in the Michigan School Readiness Program, "Each class has a certified teacher and a trained assistant to serve no more than 16 children" (Schweinhart, 2002, p.

1). After examining these data, it becomes clear that having students assigned to small class sizes plays at least some role, if not a very major part, in the success of early intervention programs.

In addition to providing the structure of small class sizes, the model early intervention programs and the class size studies, all provided services for a length of time beyond three years. These conditions are imperative for a successful program. Lee and Loeb explained that:

> No matter how beneficial the Head Start experience was initially for its participants, such benefits are likely to be undermined if these students are thereafter exposed to lower quality schooling. The particularly low quality of middle-grade schools attended by former Head Start participants explains, at least in part, why Head Start effects often fade out over time (1994, p. 1).

Effects of the STAR experiment, the SAGE project, the Perry Preschool, The Chicago CPCs, the MSRP, and the Abecedarian Project are all enduring and show little to no evidence of fade. Much of the credit for this success, in large part, has been attributed to the duration of treatment on the participants. The STAR experiment provided treatment-condition services for four years; the MSRP study began in preschool and continued through fourth grade; the Chicago CPCs begin in pre-kindergarten and continued through third grade; and select groups in the Abecedarian Project received services for eight years. These data undoubtedly give credence and/or justification to the argument for duration of program services beyond one year and beyond preschool.

Along with the evidence of duration, the model programs all provided services that meet the criteria for the condition of "intensity". The Perry Preschool has "Staff who are highly trained in early childhood education who work in a well-defined classroom operating at least 12½ hours per week" (Schweinhart, 2004, p. 3). The Abecedarian Program ran from 7:30am–5:30 pm 5 days a week, and the structured curriculum was delivered by certified teachers (Barnett & Masse, 2007). The Chicago CPCs "provided high-quality educational enrichment to at-risk children, with a focus on language and cognitive skills, delivered by well-qualified and well-paid teachers" (Temple & Reynolds, 2007, p. 129), and each class in the MSRP "has a certified teacher and a trained assistant to serve no more than 16 children" (Schweinhart, 2002, p. 4). The STAR experiment and the Wisconsin SAGE Project had instruction delivered by the same certified teacher, all day and every day, and also produced positive and significant results. Therefore, it becomes evident that intensity; that is, rigorous curriculum; taught consistently by a certified teacher each school day, is another key component of the theory for effective early childhood programs.

The final condition for the theory, which is based on data that shows positive influences on the effect of early intervention programs, is the condition of heterogeneity. Although the model preschool programs have been established to help impoverished and at-risk children, the existing body of data shows that when the programs are administered to a group of students with mixed backgrounds and abilities, the results are promising. This is especially true for the students who need the intervention the most. Several studies whose parameters met the condition of heterogeneity have found significant and positive results. The STAR experiment had pupils and the teachers assigned at random,

as did the Abecedarian Study and the High/Scope Perry Preschool Project. In addition to this evidence, Head Start requires programs to make spots available for students from moderate to high SES background which is another compelling reason to include it as a condition in the present study.

Several researchers have begun to accumulate data and assert that Head Start and other publicly funded programs can be successful, or can become more successful, if the programs are administered according to research and theory and if the sufficient follow-up conditions are met. In 2006, Lewis pointed out that Head Start fade can be avoided, or at least reduced, if certain conditions are met which support and extend preschool; i.e., class sizes; of around 15-17 students in subsequent years of schooling. Similarly, Reynolds et al. (2003) explained:

> Greater investments in effective programs like CPCs (Chicago Parent-Child Centers) can substantially reduce government expenditures for remediation and treatment services. CPCs provide a model for improving children's well being (p. 654).

Of note is that the CPCs met the conditions of early intervention (ages 3-8), small classes, duration (5 years), intensity (certified teachers all day and every day), and heterogeneity. In addition, Barnett and Schwienhart have repeatedly drawn conclusions using High/Scope data and have made suggestions on how Head Start could produce long-lasting results by following the conditions of the model programs, Perry Preschool and MSRP. Along with this, Clarke (2005) compiled theories from numerous programs and studies to create a coherent Big Theory which stated that early intervention when coupled with small class size, duration, intensity, and heterogeneity can have a lasting

impact on students and that meeting those conditions could eliminate or mitigate Head Start Fade. This theory, which was reexamined in 2008 by Clarke and Achilles, is the precedent and model for the current study.

Non-Role of Research on School Improvement

Despite the overwhelming amount of data and convincing evidence presented in this literature review from scientific based research and from sound theory for small classes, enduring programs, intense instruction, and heterogeneously-grouped classes, many publicly funded programs continue to operate according to the status quo. Head Start itself is loosely bound by federal regulations, and each program is delivered in a unique way by an independent agency or institution without regard for data from model programs and often without following suggestions put forth by research and theory. As one researcher pointed out:

> Despite the recent rhetoric about the importance of using research evidence to guide education policy and practice, the sad reality is that research has had less constructive influence during the past 10 to 15 years than it did in the previous decade or two. Most of our major policy and reform initiatives have either been launched without any prior research on their efficacy, and without provisions for assessing their intended or unintended effects, or have ignored or misused whatever relevant research evidence was available when they were launched (Schaps, 2008, p. 24).

A prime example of this is the passing of NCLB (2002). Emphasis was placed on high stakes testing and accountability regardless of the data and research which showed how ineffective that approach has been at affecting change. Perhaps the money spent on

implementing NCLB and its legislated requirements would have been better utilized creating effective preschools with small classes and certified teachers and reducing class sizes in the primary grades.

The research and theories examined in this study provide clear, cogent, and compelling evidence that Head Start and publicly funded preschools can (and at times does) produce sustainable results in both the cognitive and affective areas of child development when implemented correctly with attention to theory and scientific-based research as suggested by Haskins and Rouse (2005). The challenge is to convince educators and policy makers to sift through the poorly constructed studies pointed out by Barnett (2000) and realize the value of the existing statistically significant and practically significant data from quality programs and properly conducted-studies. As Hess explained in 2008, "Data-driven decision-making does not simply require good data; it also requires good decisions (p. 17). As is evident by material presented in Chapter 2, the good data are already available, so now is the time for education administrators and social policy makers to take action, prove the theory again, and make good decisions for our nation's children.

Chapter 3 presents the design and methods for this study in which existing data and theories are tested. In Chapter 4, the researcher examines the data and findings of the study. In Chapter 5, the researcher will present the "good decisions" which should ensure that early childhood education is being delivered in the most efficient and most effective way possible for our nation's children.

CHAPTER 3: DESIGN AND METHODOLOGY

This researcher conducted a nonexperimental quantitative research study (Johnson, 2001) to investigate the effectiveness of Head Start and the existence of the Head Start Fade effect. Using a longitudinal, explanatory design (Johnson 2001), the researcher examined grouped academic achievement data from the study's participants. In doing so, the researcher further tested the theory and treatment conditions proposed by Clarke in 2007. As shown in the literature review, decades of studies have yielded different results and different interpretations on the value of Head Start. Some researchers have gone as far as saying that the billions of dollars spent on Head Start is a waste of money and resources. However, policy-makers continue to fund the program with the hope that a relatively small public investment early in a child's life will result in benefits which greatly exceed the initial costs.

This researcher provided an analysis of Head Start attendees' academic performance. The data was analyzed to determine if preschool participation, when followed-up with participation in small classes, taught by the same teacher for the majority of the day, with a heterogeneous group of students, influenced academic performance in later years. Data from students who were members of the treatment group were compared to multiple sets of grouped data from students who did not meet the treatment conditions. The data from the treatment group was also compared to state and DFG averages, which were used as baselines. Results should provide additional data for researchers to evaluate the effectiveness of Head Start and lead to suggestions and sound reasoning for conducting a controlled experiment specifically designed to advance the present theory.

Sampling

The researcher identified four groups of students for participation in this study. The first two groups were students who participated in public preschool under Head Start conditions in the school year 2002-2003 and who continued their education in the same district for the following four years of their education. Group I was formed from these students using the number of participants who met most of the criteria for duration, intensity, and heterogeneity as proposed by Clarke (2007) in kindergarten (K), and in grades 1, 2, & 3. Group II was comprised of students from the same grade level who did not meet the established criteria set forth by the theorized treatment conditions. The researcher then repeated the methodology using groups of students from the same municipality's 2003-2004 preschool population of students. The resulting groups were Group III and Group IV.

The participants of Group I were chosen based on their participation in Head Start (early intervention) in 2002–2003 and meeting the further conditions of duration (five years of treatment conditions), intensity (same teacher all day and each school day), and heterogeneity (random assignment). The participants in Group II were chosen based on Head Start attendance in 2002–2003 (early intervention) and the failure to meet the subsequent criteria for duration, intensity, and heterogeneity. Groups III and IV were formed by applying the same criteria to students from the 2003-2004 school year. In addition to being compared to one another using reading, language arts, and math, Terra Nova and NJASK3 scores the groups' data were compared to NJASK3 state and DFG respective year averages to determine if they contribute to the research in any way.

The number of students chosen for the study was based on the number of general education students who attended Head Start in the township's school system during the 2002-2003 and 2003-2004 school years. All special education students and students who received English Language Learner (ELL) services were excluded from the study. Test results from the NJASK series of tests have never been validated for ELL and special education accommodations. Only students who remained in the district long enough to receive valid scores on the first and second grade Terra Nova tests and the NJASK3 were included.

The number of 2002-2003 preschool students who remained, after attrition was accounted for and students who received special services were extracted, who met the criteria established for duration, intensity, and heterogeneity, made up Group I (n=31). Group I is the first treatment group. Group II (n=103) was formed using the remaining general education students from the 2002-2003 preschool class whose data were available but who did not meet the criteria of the treatment conditions. This group was also separated into smaller randomly selected and matched pair subgroups to conduct further analyses.

The number of 2003-2004 preschool students who remained, after attrition was accounted for and students who received special services were extracted, who met the criteria established for duration, intensity, and heterogeneity, made up Group III (n=29). Group III is the second treatment group. Group IV (n=83) was formed using the remaining general education students from the 2003-2004 preschool class whose data were available but who did not meet the criteria of the treatment conditions. This group

was also separated into smaller randomly selected and matched pair subgroups to conduct further analyses.

As reported by the 2006 School Report Card for New Jersey, the township being studied is classified by SES as a 'DE' school district factor group (See Appendix B on DFGs). Using this scale, where 'A' districts are the poorest and 'J' districts are the wealthiest, the township is ranked on the lower middle half of the spectrum with 25.5% of the students qualifying for free or reduced lunch (NCES 2005). According to Standard and Poor's School Matters, 37.1% of the district's enrollment was white and 66.9% was non-white, 37.3% was black, 14.4% was Hispanic, and 11.3% was Asian during the 2005-2006 school year.

The town used in this study was a middle-class town with a measurable amount of diversity. Clearly, not all students whose scores were a part of this study were from equal backgrounds, especially with the assumption that Head Start attendees were the town's neediest children. However, using as large a group as possible, and considering that "typical" students were enrolled in the program with those who qualified for Head Start should result in a fair representation of various ethnic, SES, and racial groups. The researcher also conducted a descriptive analysis to check each group's characteristics relative to free/reduced lunch eligibility, race, and gender to determine comparability of the groups.

Instrumentation

This study is quantitative, mostly conducted using archived data. As such, the researcher was limited to the instruments which were available. Data for this study came from two instruments: norm-referenced and criterion-referenced standardized tests.

Grouped reading, language arts, math, and total scores from the Terra Nova standardized test in grades 1, 2, and 3 are analyzed and grouped New Jersey state-standardized test (NJASK3) language arts and math scores are examined.

The Terra Nova is designed by McGraw Hill and "has had the advantage of longstanding and thorough data on validity and reliability. In terms of content validity, it has proven to be on par with the best achievement test batteries. The test was developed through numerous revisions with a specific focus on relevance to actual curricular practice" (Sandhu, 2008). The Terra Nova measures concepts, processes, and objectives taught nationwide and common to many state standards. It meets the highest standards of accuracy and reliability (alpha near .90). However, its exact validity is difficult to determine because validity is determined through the establishment of a match between the objectives of a school, school district, or state and those assessed by the Terra Nova. Measurements of validity such as these are rarely, if ever, conducted by the independent groups who purchase the test (Cunningham, 1998).

The NJASK3 is a state-selected standardized test that was field tested in the spring of 2003 and administered operationally for the first time in the spring of 2004. The NJASK3 is administered to every third grade student in New Jersey who does not qualify for an exemption under strict guidelines. The NJDOE (2006, p.66) stated that "The validity of the NJASK scores is based on the alignment of the NJASK assessments to the Core Curriculum Content Standards and the knowledge and skills expected of third- and fourth-graders. In their technical report, NJDOE personnel summarized the NJASK's reliability and provided reliability coefficients in tables based on Cronbach's coefficient alpha measure of internal consistency (NJDOE, 2006, pp. 67-75). In 2005,

the NJASK3 Language Arts Literacy section had an alpha of .82 and the Mathematics section had an alpha of .85. (NJDOE, 2005) In the same year, the NJASK4 Language Arts Literacy Section had an alpha of .84 and the Mathematics section had an alpha of .89.

The validity of the instrumentation used in this study is strong. Data were collected ex post facto (after the fact) and have not been influenced, manipulated, or controlled by the researcher in any way. The procedures for administering, collecting, and grading state tests are clearly defined by state guidelines, and Terra Nova administration manuals and are consistently followed in participating schools. The validity is strengthened by the use of multiple assessments which can be compared and contrasted with each other, and which can be examined retrospectively over a five-year time period. The test items on all New Jersey state standardized tests are aligned to state standards which are public knowledge. A further discussion of internal and external validity is contained in the following section.

The reliability data for this study can be determined by comparing this study with other similar studies. Numerous Head Start and publicly funded preschool case studies have used standardized test scores as methods of assessment. Methods employed in this study are easily replicable in other New Jersey districts. The Terra Nova is a frequently administered standardized test nationwide and is state-mandated to administer the NJASK3 to all third grade students in New Jersey.

Internal and External Validity

The researcher attempted to reduce threats to internal and external validity. There were, however, several concerns regarding both internal and external validity in this

study. Therefore, the following discussion is presented to acknowledge and address those concerns.

Internal Validity

Threats to historical validity were present in this study because the researcher examined data from a period of 5 years under non-experimental conditions. The researcher had no control over what happened in those five years inside or outside of school. The study, however, included similar types of students based on race, SES, and gender from the same town who were in the school system for all 5 years of the study. Based on this, the students had similar historical experiences and experienced similar events. Thus, threats to historical validity were reduced. These conditions also accounted for maturation concerns. All students were in the same school system, for the same amount of time, and were generally the same age as one another throughout the study, which nullifies the maturation issue.

Testing validity was strengthened by using only general education students' test scores from standardized tests. All students who participated in this study took the same tests as the others, at the same time, and under the same conditions. By excluding special education and English Language Learners (ELL), the researcher eliminated test scores which may have been received as a result of modified testing conditions or with assistance. These conditions also accounted for equalizing instrumentation validity.

Attrition validity was not a factor in this study. The researcher included only those students who participated in the district's preschool program and who had valid scores on all three standardized tests used. This selection process mitigated attrition as a threat.

Threats to selection bias/differential selection validity were reduced by using multiple groups of students as comparison groups against the treatment groups. The first comparison group included all remaining students in the study who did not meet the treatment conditions. The students were from the same town as the students in the treatment group and attended the same preschool. The second and third comparison groups were equal-sized groups selected from the remaining students using a random selection process on SPSS statistical software. The fourth and fifth comparison groups were equal-sized matched pair groups selected to create proportionate groups with regard to race, SES, and gender. Although selection bias and matching bias could not be eliminated, the researcher made significant attempts to reduce threats by using multiple comparison groups that were selected using different parameters.

External Validity

The small sample sizes for each group in this study created population validity concerns and the setting of the study created ecological validity concerns. As a result, the ability to generalize the study's results beyond the school in this study is limited. The district in this study is categorized as a DFG D/E district by the NJDOE which is representative of the students' SES classification. Thus, the findings may be applicable to other D/E districts or schools with similar SES students' conditions nationally, but they may not be applicable to districts outside of this setting. Although these threats clearly exist, the researcher finds strength in that the study is based on a strong theoretical framework as well as previously conducted scientific-based research at various school sites across the country and internationally, whose students have characteristics different from those in this study.

Data Collection

Data for this study were collected by disaggregating historical standardized test scores from study participants' reading, language arts, and math scores on first and second grade Terra Nova and NJASK3 tests. The data was disaggregated based on the conditions of the preschool students' subsequent years in their respective elementary schools. The criteria by which students were placed into groups beyond their pre-kindergarten experience were based on their participation in smaller class sizes with heterogeneous grouping in kindergarten and grades 1, 2, & 3. The first treatment group consisted of students who attended the 2002-2003 pre-kindergarten program and who were placed in small classes in pre-kindergarten, kindergarten and grades 1, 2, and 3. Using multiple methods of data analysis, their data were compared to that of their peers who were not in small classes for all five years. The second treatment group consisted of students who attended the 2003-2004 pre-kindergarten program and who were placed in small classes in pre-kindergarten, K, and grades 1, 2, and 3. Using multiple methods of data analysis, their data were compared to that of their peers who were not in small classes for all five years. Permission was obtained from the participating district's Superintendent for the researcher to confidentially review all standardized tests results and school records for each of the study's participants.

Data Analysis

The researcher conducted non-experimental research (Kerlinger, 1968) with a longitudinal explanatory design (Johnson 2001). This means that the students' data being compared were not manipulated or affected in any way by the researcher. The data were collected after the students had already completed the predetermined factors required for

participation in the study and comparisons were made, using data collected from multiple time points, to test a theory.

Academic achievement data have been obtained from the students' first, second, and third grade standardized tests. Data were grouped according to the conditions of the study which were originally proposed by Clarke (2007). Data from students who participated in Head Start, remained in the district until third grade, and received valid scores on the first and second grade Terra Nova tests and the NJASK3 are included. This information was disaggregated into groups, depending on whether or not students met the conditions of the study in kindergarten, first grade, second grade, and third grade. The data were entered into an SPSS statistical program and analyzed for differences in mean group scores as disaggregated into the aforementioned groups. In addition, the mean group scores of each treatment group were compared to state and local averages for the same DFG as the township (See Appendix B).

To analyze the statistical significance of these data, SPSS statistical software was employed to conduct the data analyses. Calculations included, but were not limited to, correlation coefficients, independent samples t-tests, and calculated effect size using Cohen's d: $d = \frac{mean\ difference}{standard\ deviation}$ (Witte & Witte, 2007, p. 299).

All data were evaluated and used to determine if the null hypothesis for this study should be accepted or not accepted. Essentially, the analysis of data tested the theory of early intervention, duration, intensity, and heterogeneity as a unitary factor to explain the influence of its implementation on student academic achievement. If differences existed between students who have met the conditions of the theory (early intervention, duration,

intensity, and heterogeneity) and those who have not, the theory will have gained support and further recommendations for studies, including a controlled experiment, can be made.

CHAPTER 4: RESULTS OF DATA ANALYSIS

In this chapter, the researcher presents, in four sections, the results of the statistical analyses used to address the research questions developed for this study. The first section provides a description of the school district whose students' test scores were selected to be included in the study. The procedures used to collect the data for the study are presented in the second section. The third section provides the results and interpretations of quantitative statistical analyses that were used to test the research hypothesis and the theory. A qualitative analysis of research and theory related to this study's findings are presented in the final section. The researcher's purpose for the study was to explore if Head Start Fade could be eliminated or mitigated if theories and findings from scientific-based research were applied to test-score data of students who were in early intervention programs and primary grade classes under the conditions explicated in the theory.

Hypothesis

Based on the results from the review of literature, research, and established theory, the working hypothesis for this study was: Head Start students (early intervention) who experience heterogeneously-grouped, full-day, every-day K in small classes and in grades 1, 2, and 3 (duration and intensity) will not demonstrate the Head Start Fade effect on achievement tests completed in grades 1, 2, and 3. A non-experimental, longitudinal, explanatory study (Johnson, 2001) using non-equivalent control group design was conducted to test this hypothesis.

Description of the School District

The school district selected to participate in the study had six elementary schools (K-4) and was located in New Jersey. The New Jersey Department of Education (NJDOE) categorizes school districts into district factor groups (DFG) ranging from A to J, where "A" school districts are located in the poorest communities and "J" school districts are located in the state's wealthiest communities. In 2009, the state categorized the participating district as "DE" based on socio-economic status, property wealth, and other factors. The data presented in Table 6, which was obtained from the School Matters website (October, 2009), reflect the demographics of the community chosen for this study.

Table 6: School District Demographics 2009

State Tests	
District-wide Reading Proficiency	76.2%
District-wide Math Proficiency	70.6%
Classroom Profile	
Students per teacher	14.3
Enrollment	7,905
Economically Disadvantaged	26.5%
Breakdown by Ethnicity	
White	32.6%
Black	40.2%
Hispanic	15.9%
Asian/Pacific Islander	11.4%
District Household Characteristics	
Number of Households	19,392
Single-Parent Households	11.5%
Adults with at Least a High School Diploma	84.4%
Adults with at Least a Bachelor's Degree	29.0%

Spending & Revenue per Student	
Total Revenue	$12,848
Total Expenditures	$12,614

Student Groups

Data were obtained by collecting standardized test outcomes from 246 general education students who attended schools in the same school district and who met the criteria for the study. The 2002-2003 pre-k group of students who met the conditions (n = 134) attended pre-school in the district and remained in the district until at least the end of third grade. All students in this group were members of heterogeneously grouped classes that had the same teacher all day, every day. Students who qualified for the small-class group (n = 31) were in classes equal to, or less than, the district's median class size plus one for each grade (pre-k – grade 3). These students were compared to several groups of students: a) all of the remaining students who did not qualify for small classes (n = 103) b) two different random samples of students who did not qualify for small classes (both groups n = 31) and c) two different matched pair groups that were matched by gender, ethnicity and SES (both groups n = 31).

The 2003-2004 pre-k groups were assembled using the same methods that were applied to the 2002–2003 pre-k groups. There was one small class group (n = 29), one group comprised of the remaining students not in small classes (n = 83), two random sample groups taken from the students not in small classes (each group n = 29), and two matched pair groups (matched by gender, ethnicity, and SES) taken from the students not in small classes (each group n = 29). See Table 1 below for the selection criteria.

Table 1: Criteria for Treatment and Non-Treatment Groups

Treatment Groups	**Comparison Groups**
These students met the condition for smaller class sizes (median +1 or less) in all grades from preschool through grade 3.	These students did not meet the condition for smaller class sizes (median +1 or less) in all of the grades from preschool through grade 3.
I. 2002-2003 Preschool Participants n = 31	II. 2002-2003 Preschool Participants a. n = 103 b. n = 31 (random sample) c. n = 31 (random sample) d. n = 31 (matched pair) e. n = 31 (matched pair)
III. 2003-2004 Preschool Participants n = 29	IV. 2003-2004 Preschool Participants a. n = 83 b. n = 29 (random sample) c. n = 29 (random sample) d. n = 29 (matched pair) e. n = 29 (matched pair)

Tables 2 and 3 present class-size numbers (pre-k – 3) for every student who qualified for the 2002-2003 and the 2003–2004 small class-size groups. These tables are available below. Tables 4 and 5 contain the class-size numbers for all the remaining students involved in the study who did not meet the condition for small classes in this study. Because of the size of these tables, they are located in Appendix C and not within the body of text. Ideally, the group of small class-size students would have been comprised of students whose class sizes ranged from 13 -18, but since this study was conducted using archived data that was not possible.

Table 2: 2002-2003 Preschool Participants - Small Class Size Group
Ethnic Codes: 1=White 2=Black 3=Hispanic 4=American Indian 5=Asian 6=Pacific Islander 7=Multi-Ethnic

Student Number	Free Reduced Lunch (1=No 2=Yes)	Gender (1=Male 2=Female)	Ethnic Code	Class Sizes by Grades				
				Pre-K	K	1	2	3
1	2	2	1	19	24	21	21	19
2	1	1	3	21	22	18	21	21
3	1	1	2	20	25	21	19	21
4	2	2	2	17	22	20	21	21
5	2	2	3	17	22	19	21	21
6	1	1	5	21	23	19	22	21
7	1	2	3	21	24	21	18	20
8	1	1	5	17	24	20	21	21
9	1	1	3	17	23	19	21	21
10	2	2	2	17	23	19	22	19
11	1	1	2	21	23	20	21	21
12	1	1	1	21	23	18	21	19
13	1	2	2	21	23	19	21	20
14	2	1	2	19	24	21	21	19
15	2	2	5	21	23	20	21	21
16	1	1	3	19	23	21	18	20
17	1	2	5	20	25	21	20	22
18	1	1	3	20	23	21	20	21
19	1	1	3	21	23	19	21	20
20	1	1	1	18	25	21	21	22
21	1	1	1	19	24	21	21	19
22	1	1	1	21	22	19	21	20
23	1	1	3	21	23	20	21	21
24	1	1	1	21	22	18	21	21
25	1	1	1	21	22	20	21	21
26	2	2	2	21	23	19	21	20
27	1	2	3	18	25	21	19	21
28	1	1	2	21	24	21	21	20
29	1	2	1	19	24	21	18	20
30	1	1	1	18	25	21	22	22
31	1	1	2	21	23	18	21	21

Table 3: 2003-2004 Preschool Participants - Small Class Size Group
Ethnic Codes: 1=White 2=Black 3=Hispanic 4=American Indian 5=Asian 6=Pacific Islander 7=Multi-Ethnic

Student	Free Reduced Lunch (1=No 2=Yes	Gender (1=Male 2=Female)	Ethnic Code	Class Sizes by Grades				
				Pre-K	K	1	2	3
1	2	1	3	15	20	18	19	19
2	2	1	2	15	20	19	20	18
3	1	2	2	10	21	18	18	19
4	2	1	5	15	20	18	20	19
5	1	1	1	16	21	18	18	19
6	1	2	2	15	19	17	19	19
7	1	1	3	16	21	18	18	19
8	1	1	1	15	19	21	20	18
9	1	2	1	15	20	21	19	20
10	1	1	5	16	21	21	21	21
11	1	1	1	16	20	20	18	19
12	1	1	3	15	19	21	19	20
13	2	1	1	10	20	18	18	20
14	1	2	2	10	20	20	18	19
15	2	2	2	10	20	18	18	19
16	1	2	3	16	21	20	18	19
17	2	2	3	16	21	18	18	20
18	1	1	1	16	21	20	18	20
19	1	2	1	16	21	18	18	19
20	1	1	1	16	18	21	22	21
21	1	1	1	16	20	20	18	20
22	1	2	1	16	21	21	22	21
23	1	1	2	15	20	21	19	18
24	2	2	3	16	20	18	18	20
25	1	2	3	10	21	18	18	20
26	1	2	1	16	21	20	18	19
27	2	1	2	16	21	20	18	19
28	1	2	2	16	18	21	21	21
29	1	2	2	16	21	20	18	19

Instruments

The superintendent of the school district selected for the study was contacted to arrange for data collection. The researcher met with the superintendent and a district designee to review the research proposal and establish procedures for data collection. The superintendent agreed to allow the district's data to be used in the study, with the stipulations that names and identifying information for the included students had to be eliminated except for race and gender.

Achievement data provided came from three pre-existing instruments: (a) Terra Nova Grade 1, (b) Terra Nova grade 2, and (c) NJASK3. The 1st and 2nd Grades Terra Nova tests are norm-referenced and provided scores in reading, language arts, math, and a total combined score. The NJASK3 is a criterion-referenced test that provided scores in the area of language arts and math. This study is quantitative, mostly conducted using archived data. As such the researcher was limited to the instruments which were available.

The Terra Nova is designed by McGraw Hill, and "has had the advantage of longstanding and thorough data on validity and reliability. In terms of content validity, it has proven to be on a par with the best achievement test batteries. The test was developed through numerous revisions with a specific focus on relevance to actual curricular practice" (Sandhu, 2008). The Terra Nova measures concepts, processes, and objectives taught nationwide and common to many state standards. It meets the highest standards of accuracy and reliability (alpha near .90). However, its exact validity is difficult to determine because validity is determined through the establishment of a match between the objectives of a school, school district, or state and those assessed by the

Terra Nova. Measurements of validity such as these are rarely, if ever, conducted by the independent groups who purchase the test (Cunningham, 1998).

The NJASK3 is a state-selected standardized test that was field tested in the spring of 2003 and administered operationally for the first time in the spring of 2004. The NJASK3 is administered to every third grade student in New Jersey who does not qualify for an exemption under strict guidelines. The NJDOE (2006, p.66) stated that "The validity of the NJASK scores is based on the alignment of the NJASK assessments to the Core Curriculum Content Standards and the knowledge and skills expected of third- and fourth-graders. In their technical report, NJDOE personnel summarized the NJASK's reliability and provided reliability coefficients in tables based on Cronbach's coefficient alpha measure of internal consistency (NJDOE, 2006, pp. 67-75). In 2005, the NJASK3 Language Arts Literacy section had an alpha of .82 and the Mathematics section had an alpha of .85 (NJDOE, 2005). In the same year, the NJASK4 Language Arts Literacy Section had an alpha of .84 and the Mathematics section had an alpha of .89.

Data Collection

The researcher examined historical achievement data from the results of the Terra Nova for students in grades 1 and 2, and for the NJASK3. The data were disaggregated by groups, tested using Pearson correlation tests, and compared using Independent t-tests (to measure statistical significance) and Cohen's d (to measure effect size or practical significance). In addition, the grouped student data (NJASK3 scores) were compared to state and DFG averages to determine if relevant and significant differences existed.

Results of Data Analyses

Data obtained from two groups of students who began pre-kindergarten in the district one year apart were used in this study. The type of data and statistical measures used to assess them were the same but the data were from different school years. For the purpose of clarity, the results of the data analyses are reported in separate sections. The interpretation of the data analyses, however, is presented collectively in a third section. In addition, only the first set of tables is shown for each pre-k group's independent t-tests. The remaining tables are located in Appendix C. This was done intentionally by the researcher because the amount and size of the tables was disruptive to the text.

2002-2003 Preschool Group

Although the focus of this study is actual class sizes over time, not average class size, the researcher conducted correlation tests using average class size. This was done as a preliminary test to determine if there was a significant correlation between the average class sizes of all the students included in the study and their academic achievement in grades 1, 2, and 3. Scores from all three standardized tests, for all students in the 2002-2003 preschool groups, were used to determine the correlation among average class size and each score on the tests. Scores used were Terra Nova (TN) Grades 1 and 2 reading, language arts (LA), math and total and NJASK3 LA and math outcomes. In this first analysis, there were no significant Pearson correlations ($p \geq .05$) among average class size and any of the standardized test scores. See Table 7 for results.

Table 7: 2002 - 2003 Preschool Participants Average Class Size Correlations for All Standardized Tests (TN = Terra Nova)

[DataSet1] F:\SPSS COHORT INFO\Co1NoSpEdLEP.sav

Correlations

		TNR1	TNLA1	TNM1	TNTOT1	TN2R	TN2LA	TN2M	TN2TOTAL	NJASK3LA	NJASK3M	AvgClassSize
TNR1	Pearson Correlation	1	.516**	.585**	.846**	.520**	.413**	.424**	.523**	.507**	.325**	-0.032
	Sig. (2-tailed)		0	0	0	0	0	0	0	0	0	0.712
	N	134	134	134	134	134	134	134	134	134	134	134
TNLA1	Pearson Correlation	.516**	1	.534**	.804**	.378**	.425**	.392**	.464**	.473**	.412**	0.133
	Sig. (2-tailed)	0		0	0	0	0	0	0	0	0	0.126
	N	134	134	134	134	134	134	134	134	134	134	134
TNM1	Pearson Correlation	.585**	.534**	1	.814**	.621**	.459**	.672**	.676**	.533**	.522**	-0.069
	Sig. (2-tailed)	0	0		0	0	0	0	0	0	0	0.428
	N	134	134	134	134	134	134	134	134	134	134	134
TNTOT1	Pearson Correlation	.846**	.804**	.814**	1	.590**	.495**	.576**	.642**	.576**	.485**	0.038
	Sig. (2-tailed)	0	0	0		0	0	0	0	0	0	0.667
	N	134	134	134	134	134	134	134	134	134	134	134
TN2R	Pearson Correlation	.520**	.378**	.621**	.590**	1	.517**	.629**	.827**	.600**	.482**	0.029
	Sig. (2-tailed)	0	0	0	0		0	0	0	0	0	0.74
	N	134	134	134	134	134	134	134	134	134	134	134
TN2LA	Pearson Correlation	.413**	.425**	.459**	.495**	.517**	1	.547**	.789**	.533**	.453**	-0.031
	Sig. (2-tailed)	0	0	0	0	0		0	0	0	0	0.72
	N	134	134	134	134	134	134	134	134	134	134	134

Table 7 Continues

Table 7 Continued

TN2M	Pearson Correlation	.424**	.392**	.672**	.576**	.629**	.547**	1	.869**	.506**	.502**	0.029
	Sig. (2-tailed)	0	0	0	0	0	0		0	0	0	0.735
	N	134	134	134	134	134	134	134	134	134	134	134
TN2TOTAL	Pearson Correlation	.523**	.464**	.676**	.642**	.827**	.789**	.869**	1	.629**	.549**	0.034
	Sig. (2-tailed)	0	0	0	0	0	0	0		0	0	0.694
	N	134	134	134	134	134	134	134	134	134	134	134
NJASK3LA	Pearson Correlation	.507**	.473**	.533**	.576**	.600**	.533**	.506**	.629**	1	.602**	0.135
	Sig. (2-tailed)	0	0	0	0	0	0	0	0		0	0.121
	N	134	134	134	134	134	134	134	134	134	134	134
NJASK3M	Pearson Correlation	.325**	.412**	.522**	.485**	.482**	.453**	.502**	.549**	.602**	1	0.148
	Sig. (2-tailed)	0	0	0	0	0	0	0	0	0		0.087
	N	134	134	134	134	134	134	134	134	134	134	134
AvgClassSize	Pearson Correlation	-0.032	0.133	-0.069	0.038	0.029	-0.031	0.029	0.034	0.135	0.148	1
	Sig. (2-tailed)	0.712	0.126	0.428	0.667	0.74	0.72	0.735	0.694	0.121	0.087	
	N	134	134	134	134	134	134	134	134	134	134	134

**. Correlation is significant at the 0.01 level (2-tailed).

After determining that there were no statistically significant correlations for average class sizes, the grouped data were compared using multiple independent t-tests. The first t-test for the 2002-2003 preschool group was conducted using the test scores of students who were grouped into small classes under the conditions of the study (n = 31) and the remaining students not in small classes under the conditions of the study (n = 103). Achievement differences between the small-class group and the not-small class

group were not statistically significant ($p \geq .05$) for any of the ten standardized test outcomes. Several achievement differences were found to be practically significant ($\sigma \geq 0.30$), using the following guidelines: small = 0.2σ, medium = 0.5σ , and large 0.8σ or greater (Cohen, 1988). The TN2 Reading section was practically significant with a -0.37 effect size, the TN2 Math section was practically significant with a -.031 effect size, and the NJASK3 LA section was practically significant with a -.0.32 effect size. The effect size difference between the two groups was small and negatively impacted the small class group (or favored the not-small class group). See Tables 8, 8a, and 8b for results.

Table 8: 2002 - 2003 Preschool Students
Independent t-test – Small Classes and Not Small Classes - All Students

[DataSet1] F:\SPSS COHORT INFO\Co1NoSpEdLEP.sav

Group Statistics

	ClSzCode	N	Mean	Std. Deviation	Std. Error Mean
TNR1	Small Classes	31	79.42	20.288	3.644
	Not Small Classes	103	75.21	22.693	2.236
TNLA1	Small Classes	31	76.9	21.618	3.883
	Not Small Classes	103	81.99	16.738	1.649
TNM1	Small Classes	31	76.39	21.279	3.822
	Not Small Classes	103	76.98	19.886	1.959
TNTOT1	Small Classes	31	79.84	19.308	3.468
	Not Small Classes	103	81.37	17.556	1.73
TNR2	Small Classes	31	66.74	25.95	4.661
	Not Small Classes	103	75.07	21.129	2.082
TNLA2	Small Classes	31	80.32	18.375	3.3
	Not Small Classes	103	81.71	18.078	1.781
TNM2	Small Classes	31	73.65	23.919	4.296
	Not Small Classes	103	80.19	20.077	1.978
TNTOT2	Small Classes	31	76.65	22.463	4.034
	Not Small Classes	103	82.76	17.459	1.72
NJASK3LA	Small Classes	31	218.48	14.697	2.64
	Not Small Classes	103	223.34	15.326	1.51
NJASK3M	Small Classes	31	234.68	26.844	4.821
	Not Small Classes	103	240.2	30.076	2.964

Table 8a: Independent Samples Test for Small Classes and Not Small Classes – All Students
(Equal Variances Not Assumed)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TNR1	Equal variances assumed	1.275	0.261	0.926	132	0.356	4.206	4.541	-4.778	13.189
	Equal variances not assumed			0.984	54.572	0.33	4.206	4.275	-4.363	12.775
TNLA1	Equal variances assumed	4.546	0.035	-1.382	132	0.169	-5.087	3.68	-12.366	2.192
	Equal variances not assumed			-1.206	41.405	0.235	-5.087	4.219	-13.604	3.43
TNM1	Equal variances assumed	0.479	0.49	-0.143	132	0.886	-0.593	4.14	-8.784	7.597
	Equal variances not assumed			-0.138	46.893	0.891	-0.593	4.295	-9.234	8.047
TNTOT1	Equal variances assumed	1.205	0.274	-0.416	132	0.678	-1.53	3.681	-8.812	5.751
	Equal variances not assumed			-0.395	45.949	0.695	-1.53	3.875	-9.331	6.271
TNR2	Equal variances assumed	1.206	0.274	-1.821	132	0.071	-8.326	4.572	-17.369	0.717
	Equal variances not assumed			-1.631	42.667	0.11	-8.326	5.105	-18.623	1.971
TNLA2	Equal variances assumed	0.001	0.979	-0.373	132	0.71	-1.386	3.717	-8.739	5.967
	Equal variances not assumed			-0.37	48.807	0.713	-1.386	3.75	-8.923	6.151

Table 8a Continued

TNM2	Equal variances assumed	3.833	0.052	-1.521	132	0.131	-6.549	4.304	-15.064	1.966
	Equal variances not assumed			-1.385	43.497	0.173	-6.549	4.73	-16.084	2.986
TNTOT2	Equal variances assumed	3.517	0.063	-1.594	132	0.113	-6.112	3.834	-13.696	1.471
	Equal variances not assumed			-1.394	41.498	0.171	-6.112	4.386	-14.966	2.742
NJASK3LA	Equal variances assumed	1.729	0.191	-1.561	132	0.121	-4.856	3.111	-11.01	1.298
	Equal variances not assumed			-1.597	51.238	0.116	-4.856	3.041	-10.96	1.249
NJASK3M	Equal variances assumed	0.334	0.564	-0.918	132	0.36	-5.526	6.017	-17.429	6.376
	Equal variances not assumed			-0.977	54.656	0.333	-5.526	5.659	-16.87	5.817

Table 8b: 2002 -2003 Preschool Students Test for Practical Significance

2002 - 2003- All Students - Cohen's d		Test for Practical Significance		
Standardized Test	**Mean of Smaller Classes**	**Mean of Larger Classes**	**Pooled Standard Deviation**	**Cohen's d (Effect Size)**
TerraNova1 Reading	79.42	75.21	22.16	0.19
TerraNova1 LangArts	76.90	81.99	18.03	(0.28)
TerraNova1Math	76.39	76.98	20.14	(0.03)
TerraNova1Total	79.84	81.37	17.91	(0.09)
TerraNova2Reading	66.74	75.07	22.51	(0.37)
TerraNova2 LangArts	80.32	81.71	18.09	(0.08)
TerraNova2 Math	73.65	80.19	21.12	(0.31)
TerraNova2 Total	76.65	82.76	18.82	(0.32)
NJASK3 LangArts	218.48	223.34	15.27	(0.32)
NJASK3 Math	234.68	240.20	29.36	(0.19)

After determining that there were no statistically significant data for the small class group compared to the remaining students group (not-small classes) and that the only results that were practically significant favored the not-small class group, the researcher decided to further disaggregate the data and conduct additional analyses. The grouped data from the 2002-2003 preschool group were compared a second time using the students who were grouped into small classes under the conditions of the study (n = 31) and a random sample (created using SPSS statistical software) of the remaining

students who were not in small classes under the conditions of the study (n = 31). Achievement differences between the small class group and the equal-sized, random sample, not-small class group were not statistically significant ($p \geq .05$) for any of the ten standardized test outcomes. One achievement difference was found to be practically significant ($\sigma \geq 0.30$), using the following guidelines: small = 0.2σ, medium = 0.5σ , and large 0.8σ or greater (Cohen, 1988). The TN 1 LA section was practically significant with a -0.35 effect size. The effect size difference between the two groups was small and negatively impacted the small class group (or favored the not-small class group). See Tables 9, 9a, and 9b located in Appendix C.

After determining that there were no statistically significant differences ($p \leq 0.05$) for the small-class group compared to the random sample of the remaining students group, the researcher compared the grouped data from the 2002–2003 group a third time using the students grouped into small classes under the conditions of the study (n = 31) and a second random sample (selected using SPSS statistical software) of the remaining students who were not in small classes under the conditions of the study (n = 31). Achievement differences between the small class group and the second equal-sized, random sample, not-small class group were not statistically significant ($p \geq .05$) for any of the ten standardized test outcomes. Two achievement differences were found to be practically significant ($\sigma \geq 0.30$), using the following guidelines: small = 0.2σ, medium = 0.5σ , and large 0.8σ or greater (Cohen, 1988). The TN 2 Reading section was practically significant with a -0.40 effect size and the NJASK3 LA section was practically significant with a -0.46 effect size. The effect size differences between the two groups were small/medium in both instances and they negatively impacted the small

class group (or favored the not-small class group). See Tables 10, 10a, and 10b located in Appendix C.

After determining that there were no statistically significant differences ($p \leq 0.05$) for the small class group compared to the second random sample of remaining students group, the grouped data from the 2002-2003 preschool group were compared a fourth time using the students who were grouped into small classes under the conditions of the study (n = 31) and a matched pair sample (using SES, gender, and ethnicity) of the remaining students who were not in small classes under the conditions of the study (n = 31). Achievement differences between the small class group and the equal-sized, matched pair sample, not-small class group were not statistically significant ($p \geq .05$) for 9 of the 10 standardized test outcomes. One achievement difference was statistically significant ($p \leq .05$). The TN Reading 1 scores were significant with p=.048, df=60, and calculated t=2.019 where critical t=2.000. The positive value of calculated t indicated that there was a statistically significant impact of small class participation on the TN 1 Reading section. In addition, several achievement differences were found to be practically significant ($\sigma \geq 0.30$), using the following guidelines: small = 0.2σ, medium = 0.5σ , and large 0.8σ or greater (Cohen, 1988). The TN 1 Reading section was practically significant with a 0.50 effect size (medium). The TN 1 Math section was practically significant with a 0.42 effect size (small/medium) and the TN 2 Language Arts section was practically significant with a 0.36 effect size (small/medium). The effect size differences among the groups were positive, which suggests that being a member of the small class group positively influenced the outcomes on these standardized test sections. See tables 11, 11a, and 11b located in Appendix C.

After analyzing the t-test data from the small class group and the first matched pair, the grouped data from the 2002–2003 preschool group were compared a fifth time to determine if the findings (or similar findings) would be repeated. The researcher conducted another independent t-test using the students who were grouped into small classes under the conditions of the study (n = 31) and a second matched-pair group (using SES, gender, and ethnicity) from the remaining students who were not in small classes under the conditions of the study (n = 31). Achievement differences between the small class group and the second equal-sized, matched pair, not-small class group were not statistically significant ($p \geq .05$) for any of the 10 standardized test outcomes. Five achievement differences were practically significant ($\sigma \geq 0.30$), using the following guidelines: small = 0.2σ, medium = 0.5σ , and large 0.8σ or greater (Cohen, 1988). The TN 1 LA section was practically significant with a -0.46 effect size (medium). The TN 2 Reading section was practically significant with a -0.36 effect size (small), as was the TN 2 Math section with an effect size of -0.32 (small) and the TN 2 Total score with an effect size of -0.31 (small). In addition, the NJASK3 Language Arts scores were practically significant with an effect size of -0.34 (small). The effect size differences between all groups were negative which suggests that being a member of the small class group negatively influenced the outcomes on these standardized tests. See Tables 12, 12a, and 12b located in Appendix C.

2003-2004 Preschool Group

As with the 2002–2003 preschool participants, the researcher acknowledges that actual class sizes over time, not average class size, are the primary focus of this study. The researcher conducted correlation tests using average class sizes, however, to

determine if any significant correlations existed among the average class sizes of all the students included in the study and their academic achievement in grades 1, 2, and 3. Scores of all three standardized tests for all students in the 2003–2004 preschool groups were analyzed to determine if there existed a significant correlation among average class size and each score on the tests. The scores used were TN Grades 1 and 2 reading, LA, math, and total and NJASK3 LA and math outcomes. In this test, there was one significant Pearson correlation ($p \leq .05$) between average class size and the NJASK3 math scores. The correlation was small in strength ($r = .238$) but significant ($p = .011$) and positive. This means that there was causation suggesting that being in a larger class positively impacted performance on the NJASK3 math test. 6% of the variance can be explained by these data ($r^2 = .0566$). See Table 13 below.

Table 13: 2003 – 2004 – Correlations: Mean Class Size with All Standardized Tests

[DataSet1] F:\SPSS COHORT INFO\Co2NoSpEdLEP.sav

Correlations

	TN1R	TN1LA	TN1M	TN1TOTAL	TN2R	TN2LA	TN2M	TN2TOTAL	NJASK3LA	NJASK3MA	Mean Class Size
TN1R Pearson Correlation	1	.598**	.568**	.840**	.523*	.512**	.429*	.587**	.383**	.347**	0.022
(2-Tailed)											
N	0	0	0	0	0	0	0	0	0	0	0.819
	112	112	112	112	112	112	112	112	112	112	112
TN1LA Pearson Correlation	.598*	1	.608**	.846**	.626*	.580**	.455*	.662**	.406**	.462**	-0.04
(2-Tailed)											
N	0		0	0	0	0	0	0	0	0	0.672

	112	112	112	112	112	112	112	112	112	112	112
TN1M Pearson Correlation (2-Tailed) N	.568** .000 112	.608** .000 112	1 112	.835** .000 112	.508** .000 112	.526** .000 112	.611** .000 112	.658** .000 112	.440** .000 112	.597** .000 112	0.13 0.171 112
TN1TOTAL Pearson Correlation (2-Tailed) N	.840** .000 112	.846** .000 112	.835** .000 112	1 112	.611** .000 112	.622** .000 112	.562** .000 112	.723** .000 112	.442** .000 112	.532** .000 112	0.038 0.693 112
TN2R Pearson Correlation (2-Tailed) N	.523** .000 112	.626** .000 112	.508** .000 112	.611** .000 112	1 112	.564** .000 112	.625** .000 112	.829** .000 112	.487** .000 112	.475** .000 112	-0.014 0.883 112
TN2LA Pearson Correlation (2-Tailed) N	.512** .000 112	.580** .000 112	.526** .000 112	.622** .000 112	.564** .000 112	1 112	.513** .000 112	.839** .000 112	.536** .000 112	.527** .000 112	-0.052 0.583 112
TN2M Pearson Correlation (2-Tailed) N	.429** .000 112	.455** .000 112	.611** .000 112	.562** .000 112	.625** .000 112	.513** .000 112	1 112	.838** .000 112	.599** .000 112	.654** .000 112	-0.019 0.838 112
TN2TOTAL Pearson Correlation (2-Tailed) N	.587** .000 112	.662** .000 112	.658** .000 112	.723** .000 112	.829** .000 112	.839** .000 112	.838** .000 112	1 112	.635** .000 112	.670** .000 112	-0.032 0.736 112
NJASK3LA Pearson Correlation (2-Tailed) N	.383** .000	.406** .000	.440** .000	.442** .000	.487** .000	.536** .000	.599** .000	.635** .000	1 	.640** .000	0.17

Table 13 Continued

	0 1 1 2	0 112	0 112	0 112	0 1 1 2	0 11 2	0 1 1 2	0 112	 112	0 112	0.072 112
NJASK3MA Pearson Correlation (2-Tailed) N	.3 4 7*. 0 1 1 2	.462** 0 112	.597** 0 112	.532**	.4 7 5*. 0 1 1 2	.5 27** 0 11 2	.6 5 4*. 0 1 1 2	.670** 0 112	.640** 0 112	1 112	.238* 0.011 112
MeanClassSize Pearson Correlation (2-Tailed) N	0. 0 2 2 0. 8 1 9 1 1 2	- 0.04 0.67 2 112	0.1 3 0.1 71 112	0.038 0.693 112	- 0. 0 1 4 0. 8 8 3 1 1 2	- 0. 05 2 0. 58 3 11 2	- 0. 0 1 9 0. 8 3 8 1 1 2	- 0.0 32 0.7 36 112	0.1 7 0.0 72 112	.238* 0.0 11 112	1 112

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

After calculating the correlations for the 2003–2004 preschool group, the grouped data were compared using an independent t-test. Students were grouped into two groups: small classes under the study's conditions (n = 29) and not-small classes using the remaining students who were not in small classes under the conditions of the study (n = 83). Achievement differences between the small class group and the not-small class group were not statistically significant ($p \geq .05$) for 9 of the 10 standardized test outcomes. One achievement difference was found to be statistically significant ($p \leq .05$). The significant difference was found on the NJASK3 math scores ($p = 0.046$, $df = 110$, and calculated $t = -2.018$ where critical $t = 2.000$). The negative t score suggested that being a member of the small-class group negatively influenced performance on the NJASK3. In addition to this, two achievement differences were found to be practically significant ($\sigma \geq 0.30$), using the

following guidelines: small = 0.2σ, medium = 0.5σ , and large 0.8σ or greater (Cohen, 1988). The TN 1 math section was practically significant with a -0.37 effect size and the NJASK3 math section was practically significant with a -0.43 effect size. The effect size differences between the two groups were small/medium in both instances and they negatively impacted the small class group (or favored the not-small class group). See Tables 14, 14a, and 14b below.

Table 14: 2003 – 2004 - Independent T-test – All Students

[DataSet2] F:\SPSS COHORT INFO\Co2NoSpEdLEP.sav

Group Statistics

	ClsSzCode	N	Mean	Std. Deviation	Std. Error Mean
TN1R	Small Classes	29	80	21.414	3.977
	Not Small Classes	83	80.7	22.823	2.505
TN1LA	Small Classes	29	83	18.476	3.431
	Not Small Classes	83	83.76	17.362	1.906
TN1M	Small Classes	29	70.97	25.777	4.787
	Not Small Classes	83	78.99	19.882	2.182
TN1TOTAL	Small Classes	29	80.93	21.426	3.979
	Not Small Classes	83	84.28	18.294	2.008
TN2R	Small Classes	29	69.66	21.036	3.906
	Not Small Classes	83	67.81	23.72	2.604
TN2LA	Small Classes	29	74.45	17.185	3.191
	Not Small Classes	83	75.63	24.015	2.636
TN2M	Small Classes	29	71.48	22.823	4.238

	Not Small Classes	83	71.29	23.762	2.608
TN2TOTAL	Small Classes	29	74.28	16.533	3.07
	Not Small Classes	83	74.88	22.194	2.436
NJASK3LA	Small Classes	29	221.55	14.669	2.724
	Not Small Classes	83	222.63	15.762	1.73
NJASK3MA	Small Classes	29	236.1	22.732	4.221
	Not Small Classes	83	245.72	21.873	2.401

Table 14a: 2003 -2004 Independent Samples Test – All Students

(Equal Variance Not Assumed)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TN1R	Equal variances assumed	0.151	0.698	-0.144	110	0.886	-0.699	4.848	-10.306	8.908
	Equal variances not assumed			-0.149	51.847	0.882	-0.699	4.7	-10.13	8.733
TN1LA	Equal variances assumed	0.31	0.579	-0.199	110	0.842	-0.759	3.808	-8.305	6.787
	Equal variances not assumed			-0.193	46.435	0.847	-0.759	3.925	-8.657	7.139
TN1M	Equal variances assumed	3.601	0.06	-1.727	110	0.087	-8.022	4.646	-17.229	1.184
	Equal variances not assumed			-1.525	40.256	0.135	-8.022	5.261	-18.653	2.608
TN1TOTAL	Equal variances assumed	0.698	0.405	-0.81	110	0.419	-3.346	4.129	-11.528	4.836

Table 14a Continued

	Equal variances not assumed			-0.751	43.125	0.457	-3.346	4.457	-12.333	5.641
TN2R	Equal variances assumed	1.439	0.233	0.371	110	0.711	1.848	4.976	-8.013	11.708
	Equal variances not assumed			0.394	54.715	0.695	1.848	4.694	-7.561	11.257
TN2LA	Equal variances assumed	4.341	0.04	-0.243	110	0.808	-1.178	4.848	-10.786	8.429
	Equal variances not assumed			-0.285	68.374	0.777	-1.178	4.139	-9.437	7.08
TN2M	Equal variances assumed	0.52	0.472	0.038	110	0.97	0.194	5.075	-9.864	10.251
	Equal variances not assumed			0.039	50.741	0.969	0.194	4.976	-9.798	10.185
TN2TOTAL	Equal variances assumed	5.085	0.026	-0.134	110	0.894	-0.604	4.508	-9.538	8.331
	Equal variances not assumed			-0.154	65.494	0.878	-0.604	3.919	-8.43	7.223
NJASK3LA	Equal variances assumed	0	0.992	-0.322	110	0.748	-1.075	3.342	-7.697	5.548
	Equal variances not assumed			-0.333	52.245	0.74	-1.075	3.227	-7.55	5.4
NJASK3MA	Equal variances assumed	0.009	0.924	-2.018	110	0.046	-9.619	4.766	-19.065	-0.174
	Equal variances not assumed			-1.981	47.354	0.053	-9.619	4.856	-19.387	0.148

Table 14b: 2003 -2004 Preschool Group Test for Practical Significance

2003 – 2004 - All Students - Cohen's d		Test for Practical Significance		
Standardized Test	Mean of Smaller Classes	Mean of Larger Classes	Pooled Standard Deviation	Cohen's d (Effect Size)
TerraNova1 Reading	80.00	80.70	22.37	(0.03)
TerraNova1 LangArts	83.00	83.76	17.58	(0.04)
TerraNova1Math	70.97	78.99	21.73	(0.37)
TerraNova1Total	80.93	84.28	19.11	(0.18)
TerraNova2Reading	69.66	67.81	22.98	0.08
TerraNova2 LangArts	74.45	75.63	22.38	(0.05)
TerraNova2 Math	71.48	71.29	23.42	0.01
TerraNova2 Total	74.28	74.88	20.81	(0.03)
NJASK3 LangArts	221.55	222.63	15.43	(0.07)
NJASK3 Math	236.10	245.72	22.40	(0.43)

After analyzing the data for the small class group compared to the remaining students group, the grouped data from the 2003–2004 preschool group were compared a second time using the students who were grouped into small classes under the conditions of the study (n = 29) and a random sample of the remaining students who were not in small classes under the conditions of the study (n = 29). Achievement differences between the small class group and the equal-sized, random sample, not-small class group were not statistically significant ($p \geq .05$) for any of the ten standardized test outcomes. Four achievement differences, however, were found to be practically significant ($\sigma \geq 0.30$), using the following guidelines: small = 0.2σ, medium = 0.5σ , and large 0.8σ or greater (Cohen, 1988). The TN 1 math section was practically significant with a -0.36

effect size. The effect size difference between the two groups was small and negatively impacted the small class group (or favored the not-small class group). The TN 2 reading section was practically significant with a 0.34 effect size. The effect size difference between the two groups was small and positively impacted the small class group. The TN 2 math section was practically significant with a 0.40 effect size. The effect size difference between the two groups was small/medium and positively impacted the small class group. The TN 2 total score was practically significant with a 0.31 effect size. The effect size difference between the two groups was small and positively impacted the small class group. See Tables 15, 15a, and 15b located in Appendix C.

After determining that there were practically significant data for the small class group compared to the random sample of the remaining students group, the grouped data from the 2003 -2004 preschool group were analyzed a third time to determine if the findings (or similar results) would be replicated. Using the students who were grouped into small classes under the conditions of the study (n = 29) and a second random sample of the remaining students who were not in small classes under the conditions of the study (n = 29) another independent t-test was conducted. Achievement differences between the small class group and the second equal-sized, random sample, not-small class group were not statistically significant ($p \geq .05$ and/or calculated $t < 2.000$) for any of the ten standardized test outcomes. Two achievement differences, however, were found to be practically significant ($\sigma \geq 0.30$), using the following guidelines: small = 0.2σ, medium = 0.5σ , and large 0.8σ or greater (Cohen, 1988). The TN 1 math section was practically significant with a -0.44 effect size and the NJASK3 math section was practically significant with a -0.51 effect size. The effect size differences between the two groups

were medium in both instances and favored the not-small class group. See Tables 16, 16a, and 16b located in Appendix C.

After determining that there were no statistically significant data for the small class group compared to the second random sample of remaining students group, the findings became increasingly inconsistent. Thus, the grouped data from the 2003-2004 preschool group were compared a fourth time using the students who were grouped into small classes under the conditions of the study (n = 29) and a matched pair sample of the remaining students who were not in small classes under the conditions of the study (n = 29). Achievement differences between the small class group and the equal-sized, matched pair sample (taken from the not-small class group) were not statistically significant ($p \geq .05$) for all ten standardized test outcomes. Several achievement differences, however, were found to be practically significant ($\sigma \geq 0.30$), using the following guidelines: small = 0.2σ, medium = 0.5σ , and large 0.8σ or greater (Cohen, 1988). The TN 2 LA section was practically significant with a 0.30 effect size (small) and the TN 2 Total scores were practically significant with a 0.32 effect size (small). The effect size differences among the groups were positive which suggests that being a member of the small class group positively influenced the outcomes on these standardized test sections. See Tables 17, 17a, and 17b located in Appendix C.

After analyzing the data from the t-test with the small class group and the first matched pair, the grouped data from the 2003 – 2004 preschool group were compared a fifth time using the students who were grouped into small classes under the conditions of the study (n = 29) and a second matched pair group from the remaining students who were not in small classes under the conditions of the study (n = 29). Achievement

differences between the small class group and the second equal-sized, matched pair, not-small class group were not statistically significant ($p \geq .05$ and/or critical $t < 2.000$) for any of the ten standardized test outcomes. Two achievement differences, however, were found to be practically significant ($\sigma \geq 0.30$), using the following guidelines: small = 0.2σ, medium = 0.5σ , and large 0.8σ or greater (Cohen, 1988). The TN 1 math section was practically significant with a -0.50 effect size (medium) and the NJASK3 math section was practically significant with a -0.48 effect size (medium). The effect size differences between the groups were negative which suggests that being a member of the small class group negatively influenced the outcomes on these standardized tests. See Tables 18, 18a, and 18b located in Appendix C.

2002-2003 and 2003–2004 Preschool Groups Compared to State and DFG Averages

The researcher compared both preschool groups' and small class size groups' NJASK3 scores to the readily available state and matching DFG general education averages which were used as baselines. The 2002-2003 small class group, who took the NJASK3 in 2007, scored an average 218.48 on the NJASK3 language arts section compared to a state general education average of 222 and a DFG (D/E) general education average of 222.9. Thus, the 2002 – 2003's small class group's language arts scores were lower than the state and DFG averages by roughly 4 points. The 2002-2003 small class group, who took the NJASK3 in 2007, scored an average 234.68 on the NJASK3 math section compared to a state general education average of 236 and a DFG (D/E) general education average of 237.7. Thus, the 2002–2003 small class group's math scores were lower than the state and DFG averages by roughly 1 and 3 points respectively.

The 2003–2004 small class group, who took the NJASK3 in 2008, scored an average 221.55 on the NJASK3 language arts section compared to a state general education average of 221.4 and a DFG (D/E) general education average of 221.9. Thus, the 2003–2004 small class group's language arts scores were basically equal to the state and DFG averages. The 2003–2004 small class group, who took the NJASK3 in 2008, scored an average 236.1 on the NJASK3 math section compared to a state general education average of 235.9 and a DFG (D/E) general education average of 237.4. Thus, the 2003–2004 small class group's math scores were nearly equal to the state and DFG averages.

Quantitative Results Regarding Head Start Fade

The statistics in the previous section were related to achievement outcomes of two different groups of students who attended school one year apart. The tests used were divided into ten sub-categories, and each group's scores were analyzed five different times with varied class size groupings. The results of the tests indicated that participating in smaller classes relative to peers, but not small by SBR standards, does not eliminate a fade effect. The following analyses support these findings.

The correlation tests which were conducted on both groups of students showed only one significant correlation between average class size and performance on any of the ten standardized tests sections. The 2002-2003 preschool group's test produced no significant correlations at all. The 2003-2004 preschool group's test resulted in the only significant correlation. The result was a small correlation ($r = .238$) on the NJASK3 math section and it was positive which means that the larger the class sizes became the higher

the achievement results were. These data do not support the elimination or diminution of the fade effect by reducing class size averages.

The independent t-tests conducted on both groups of students produced results that also appeared inconsistent with research on class-size reduction (CSR). In total, there were 50 t-tests conducted on the data for each group - equaling a total of 100 t-tests for the study. Of the 2002–2003 preschool groups' 50 t-tests, only one produced statistically significant results. In the first matched pairs group, the TN Reading scores for first graders produced a statistically significant result ($p = 0.048$, $t = 2.019$, and $df = 60$) that favored participation in smaller classes. These findings suggest that participating in reduced class size, by this study's standards, does not eliminate the Head Start fade. This is evident because: (a) only one of the 50 tests was significant, (b) the only statistically significant result occurred in first grade, and (c) the statistically significant finding occurred in one matched pair group but did not occur again with the second matched pair group, either of the random groups' tests, or the whole group comparisons.

Of the 2003-2004 preschool groups' 50 t-tests, only one produced statistically significant results. Comparing data from the small class size group to the remaining students who were not in reduced sized classes resulted in a statistically significant finding on the NJASK3 math test ($p = 0.046$, $t = -2.018$, and $df = 110$). The negative t result favored participation in larger classes. These findings suggest that participating in reduced class sizes, by this study's standards, does not eliminate or mitigate Head Start fade. This is evident because: (a) only one of the 50 tests was significant, and (b) the only significant result favored being a member of the larger class size group.

In addition to the correlation tests and the independent t-tests, the tests for practical significance, or effect size, also produced results that were inconsistent with the body of research on CSR. Of the 50 tests conducted on the 2002-2003 preschool group of students, using Cohen's d, 14 of the 50 tests produced practically significant findings. The majority of the findings were small ($\sigma < .50$) and the majority of the findings favored participation in the larger class-size group. Only 3 of the 14 significant effect sizes favored participation in the small class-size group. All 3 of the significant findings were in the first matched pair group. The findings were not replicated among any of the 4 other groups and none of the 3 practically significant findings extended into third grade. As a result of these data, participating in small class sizes by this study's standards, was not shown to not eliminate the Head Start Fade effect.

Of the 50 tests conducted on the 2003–2004 preschool group of students, using Cohen's d, 12 of the 50 tests produced practically significant findings. The majority of the findings were small ($\sigma < .50$) and favored participation in the larger class-size group. Only 5 of the 12 significant effect sizes favored participation in the small class-size group. 2 of the practically significant findings occurred in the first matched pair group and 3 of the findings occurred in the first random sample group. The findings were not replicated in the three other groups and none of the 5 practically significant findings extended into third grade. Thus, participating in small class sizes by this study's standards was not shown to eliminate the Head Start Fade effect.

Along with these statistical findings are the comparisons of each group's NJASK scores to the state and DFG averages that were used as baselines. The 2002-2003 preschool group of students in the small class-size group scored slightly lower than the

general education average for the state and the general education average for the matching DFG (D/E) on both the LA section and match section of the NJASK3. The 2003–2004 preschool group of students in the small class-size group had outcomes that were nearly equal to the general education state average and the same DFG (D/E) averages on both the language arts and math sections of the NJASK3. Participating in the CSR treatment group did not result in higher outcomes than state or DFG test averages. These findings suggest that participating in reduced class size classes by this study's standards does not eliminate Head Start Fade.

Although this study's findings appear inconsistent with the existing body of research and the knowledge dynamic on the cumulative influence of CSR in the early grades, this is not the case. The data produced by the tests on academic achievement outcomes in this study confirm the SBR and existing theory base. Class-size studies such as STAR (Word et al., 1990), SAGE (Molnar, Smith, Zahorik, Palmer, Halbach, & Ehrle, 1999), and CSPAR (Blatchford, et al., 2003), along with early intervention studies such as the Perry Preschool Study (Schweinhart et al., 2005) and the Abecedarian study (Campbell & Ramey, 2007), clearly defined standards for small class sizes as 13 through 18 students. Although the researcher was able to assemble a group of students who participated in classes that were smaller than their peers' for five consecutive years, none of the participants in this study participated in classes of the size established by SBR for the duration of the study. In fact, few participants were members of small classes by SBR standards in any of their first five years of school.

The effect of CSR is one of the key components, if not the integral component, to the theory being tested in this study. While the study tested the impact, or influence, of

CSR, none of the data from the participants in this study met the criteria of coming from small classes established by the existing SBR. Therefore, the quantitative findings from this study confirm the theory proposed by Clarke (2007), which stated that to eliminate or mitigate the effects of Head Start Fade all of the conditions of 1) early intervention, 2) participation in small classes, 3) duration, 4) intensity, and 5) heterogeneity must be met. By not meeting the criteria for one of the major components of the theory (CSR), the desired results to support a reduction in Head Start Fade continued to be elusive. Thus, the findings from this study provide confirmatory support for Clarke's argument and strengthen the theory proposed in 2007. A further discussion on the qualitative nature of these findings is in the following section.

Qualitative Results Regarding Head Start Fade

In 2007, Clarke tested a theory which proposed that an early intervention program with small class sizes (13-18) taught by a certified teacher, of considerable duration (4-5 years) and intensity (all day every day) with heterogeneously grouped classes would eliminate or mitigate the effects of Head Start Fade. Results of the study showed that, by meeting the criteria for all of the components of the study, an impact on Head Start Fade could be measured, especially in the neediest populations. The purpose of the present study was to investigate if the application of Clarke's theory in another setting produced significant and/or practical results to reduce Head Start Fade effects in the participating school district. The present study's data confirmed Clarke's theory, not by finding many significant or practical results, but by showing that the improper implementation of one of the theory's conditions, specifically CSR, influenced the finding of a negative or non-significant effect on the reduction of Head Start fade. By not meeting a major component

of the theoretic construct (CSR), the study resulted in unfavorable outcomes. The following qualitative information establishes one hypothesis for why this occurred.

As early as the 1920s, Piaget developed theories and conducted research that concentrated on the importance of the early years of a child's life. His work on the sensory-motor, pre-operational, and concrete-operational stages of children's cognitive development laid the groundwork for the constructivist movement. These principles were later built upon by the likes of well-known scientists such as Vygotsky and Bruner. The constructivist theories which have come from their work have provided support for the argument that effective early intervention programs are paramount in a child's development. Bloom (1964) put forth theories that stressed the importance of developmental learning, and his work highlighted the importance of starting early when the rate of growth is strongest for children in education programs. This component of the theoretic construct (early intervention) was implemented effectively in the present study.

In 1998, Ramey and Ramey created a framework, in the Abecedarian study, for early intervention programs with an emphasis on structure and efficacy. The three major conditions proposed by Ramey and Ramey (1989) were (a) early intervention – begin schooling at young age, (b) duration – provide small classes for three to four years(preferable), and (c) intensity – maintain the small class with the same teacher all day and every day. This theoretical construct has been developed further by the body of SBR such as the STAR study (Word et al.,1990), and class size research conducted by Finn and Achilles (1999); Finn, Gerber, Achilles, and Boyd-Zaharias (2001); Krueger and Whitmore (2000); Nye, Hedges, and Kanstantopoulos (1999); and Tienken and Achilles (2009) . In addition, research from studies such as STAR (1990), Chicago Child

Parent Centers (CPCs) and Perry Preschool have added to this theory the significance of parental involvement, participation in the cohort experience, or Sarason's (1974) Psychological Sense of Community (PSOC), and heterogeneity or random class assignment.

The foundation for the theoretical framework is supported by the work of Deming (2000), who explained that structural changes account for 80% or more of an organization's effectiveness. If administrators subscribe to this proposal and structure early intervention programs to affect working conditions, such as implementing CSR for the first five years of schooling, then one can, and should, expect to find significant results. As Achilles (1999) explained, "The contexts in which teachers must teach (i.e., structure and organization) influence greatly what teachers can do to teach and teach well" (p.11). With this understanding, one can assume that an intense intervention, aimed at children in the early years, and sustained for a considerable amount of time can and should produce significant and practical results as measured by student achievement. The structural component of the theoretic construct, however, was only partially implemented in this study. While the program did meet the criteria for heterogeneity, intensity, and duration, it did not meet the criteria for small classes.

The theory-testing construct employed in this study was designed to determine if the theoretic position of Ramey and Ramey (1989), which later became the theoretic base of the STAR study (Word et al, 1990) and a basis for Clarke's study (2007), would decrease or eliminate Head Start fade when applied to the district's early intervention program. Results did not show significant or practical effects of marginally reducing class sizes on Head Start fade. These outcomes, however, confirm that the theoretical

construct developed over the years in the Abecedarian Program, Perry Preschool Program, Chicago Child Parent Centers, the STAR study (1990), the SAGE study (1996), and the Burke County Studies (1991), must be implemented precisely and consistently to be effective. Evidence similar to that discovered in this study, from a much larger data group, is present in the outcomes of the Indiana Prime Time (1984-1986) class size reduction initiative. Under this initiative, 286 districts participated in a CSR intervention that did not produce consistent outcomes. The most commonly cited reason for the inconsistent outcomes is that the CSR implementation did not follow well-defined guidelines evident in the knowledge dynamic, as was the case in this study's analysis of archived data.

Student members of the small class size group in this study were randomly assigned to a Head Start class and to classes for the subsequent four years. In addition, their classes were taught by the same teacher all day and every day. They met the conditions of the theoretic construct for this study in all areas except one. Although they were in "smaller" classes than their peers, the students were not in small classes as measured by the standards established by SBR. As such, the effects of their program were not enduring and did not show a significant reduction in fade when compared to their peers although the students did achieve comparable test scores when compared to state and DFG (D/E) averages. If the major component of the theoretic construct - class size reduction - was implemented according to the recommendations in addition to the other components, the expectation of the researcher is that study findings would have been different. These conclusions are based on the body of research that supports CSR's

positive and significant influence on student achievement when implemented as prescribed by the theoretic construct.

Summary

In this study, the researcher compared grouped student outcomes on standardized tests to determine if CSR had an effect on Head Start Fade. The students in the small class group began their tenure in the district as members of the Head Start program. They attended full-day kindergarten and were members of "smaller" class sizes when compared to their peers for the first five years of schooling. The results of the data analyses did not provide support for the elimination of Head Start Fade in the participating district's small class groups. The data, however, do support the theoretic construct of early intervention, duration, intensity, and heterogeneity by showing that without proper implementation of the interventions, specifically CSR in this study, consistent influence on achievement cannot be expected. Conclusions and recommendations based on these findings are presented in Chapter 5.

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

For decades, empirical research has supported the existence of a phenomenon known as Head Start Fade, where students who make initial gains in an early intervention program lose those gains by the third or fourth grade. Education theorists and researchers have proposed theories and have found evidence that, by properly structuring early intervention programs and the programs that are delivered in the subsequent years of schooling, Head Start Fade can be eliminated and/or greatly reduced. In this nonexperimental study, the researcher employed retrospective data analysis using a longitudinal explanatory design to test the influence of small class sizes, intensity (all day and every day), duration (five years), and heterogeneity (random class assignment) on the fade effect.

At first review, this study's findings appear to be inconsistent with the theory. Upon closer scrutiny, however, findings from this study are consistent with the theory and research on the influence of: early intervention, reduced class size, intensity, duration, and heterogeneity. The study provided data that confirmed what is stated in the theory - that, when not implemented correctly, class size reduction (CSR) initiatives are generally ineffective at eliminating the fade. In this study, the researcher also tracked the influence that effective programs such as the Perry Child Development Center, the Abecedarian Program, and the Chicago Parent Centers had on eliminating or moderating the fade effect. The researcher compared and contrasted these model programs with the

one offered in the study to highlight the importance of consistency when implementing the theory to achieve a significant influence on Head Start Fade.

Chapter 5 begins with a statement of the issue addressed in the study, a study of the literature on successful pre-k programs and class size research, the methodology and hypothesis of the study, and the findings of the study, as well as how they compare with the knowledge dynamic and prior research. The chapter ends with a conclusion and recommendations. The recommendations are for (a) policy; (b) practice; and (c) advancing the theory of early intervention, intensity, duration, and heterogeneity in future research.

Summary

An extensive review of the literature on successful early intervention programs and research on class size revealed several parallel findings that were keys to reducing, or eliminating, Head Start Fade. The findings from aggregating this information suggest that an early intervention program with small class sizes that continues for multiple years with randomly assigned students should have a positive and statistically significant influence on the program's participants. Evidence for this can be found in the largest, longest, and most well-respected empirical class-size study to date, the Tennessee STAR study, as well as in numerous successful early intervention programs; e.g., the Abecedarian Program, the Chicago Child Parent Centers, and the Perry Childhood Development Centers (which continue to be monitored by Ypsilanti, Michigan's well-respected High Scope Foundation). These elements are evident in all of the aforementioned programs, and as such they have become the foundation for the theoretical design of an effective early education program that is tested by this study.

Researchers for the STAR study conducted in Tennessee found several significant outcomes that are relevant to this study: (1) Early intervention prevented an achievement gap, (2) there was no fade effect among the study's participants, (3) students in small classes had stronger gains than did their peers in larger classes with or without an aide, and (4) the study's participants had a lower referral rate for special education services and in-grade retention. These findings from the original study were strengthened by findings in STAR follow-up studies such as the Lasting Benefits Study (LBS), Challenge, and Enduring Effects. Also, studies similar to STAR, the DuPont Study, SAGE, Challenge, and Burke County, North Carolina supported the outcomes and findings of the STAR data. In addition to these class-size studies, several extensive early intervention studies, many with ongoing longitudinal studies, found similar results. The Abecedarian Program, the Chicago Child Parent Centers, and the Perry Childhood Development Centers all have added to the body of evidence that well-structured early intervention programs with the proper follow-up interventions lessen the achievement gap for needy children and produce enduring results well beyond the implementation years.

The Purpose, Design, Method, Hypothesis, and Outcome of the Study

In this study, a theory which has evolved over time from class-size and early intervention research findings was tested in an actual school setting. The theory proposed that implementing an early intervention program with small class sizes (13-18) for a significant duration (pre-k – 3), with intensity (same teacher all day every day) and heterogeneity (random class assignment) would eliminate or reduce the Head Start Fade effect. The theory was tested using archived data from two different groups of students who attended school one year apart in the same New Jersey school district. The specific

purpose of the study was to determine if meeting the conditions of the theory produced academic gains as measured by three different standardized tests, administered at different grade levels, and if any initial gains persisted beyond four years.

The researcher used a nonexperimental, longitudinal, retrospective explanatory design (Johnson, 2001) to structure the theory-testing study. The study was conducted on two groups of students who attended the same school one year apart, and the duration for each group was five years (pre-k through grade 3). The specific years of the study spanned 2002-2007 for the first group of students and 2003-2008 for the second group of students. Also, the specific numbers of students in the study were: N= 134 for the first group and N= 112 for the second group.

The method employed was as follows: Two groups of students (Group I and Group II) who participated in the districts 2002-2003 and 2003–2004 pre-kindergarten programs were identified and separated into subgroups based on their respective class sizes for the first five years of their schooling. The subgroups within each pre-k group consisted of one group of students who participated in classes that were considered small by the study's parameters (less than the mean class size plus one) and five groups of students using the remaining students who were not in small classes. These groups were: 1) all of the not-small class students combined, (2) an equal-sized randomly selected group of not-small class students, (3) a second equal-sized randomly selected group of not-small class students, (4) an equal-sized matched pair group of not-small class students, and (5) a second equal-sized matched pair group of not small class students. The researcher examined the outcomes of these students first and second grade Terra

Nova scores and their scores on the NJASK3 using several tests for statistical and practical significance.

Study outcomes provided data that were inconsistent with the theory proposed throughout the study and the existing research on class size and early intervention. This was most likely the case because of improper/inconsistent application of the theory. Specifically, the class size reductions (CSR) were not implemented correctly. The class sizes of the small class groups in this study were not consistent with the recommended class sizes of 13-18 students for all five years. As a result, the outcomes did not support a reduction or elimination of Head Start Fade among the small class participants. These data confirmed that CSR initiatives must be implemented with acute attention to prior research. If and when CSR initiatives are not implemented according to the parameters set forth by studies such as STAR, SAGE, Burke County, Abecedarian, and the Perry Preschool Study, the outcomes will be unpredictable at best.

In this study, there was one statistically significant finding (out of 100 t-tests) that suggested students in moderately reduced class sizes performed better than their peers in larger classes on standardized tests. There was one statistically significant finding that suggested that students in the larger classes performed better. The data which supported the moderately reduced classes resulted from an independent t-test when the small class subgroup from the 2003-2004 preschool group was compared to the first equal-sized matched pair subgroup using the first grade Terra Nova. The data which favored the larger classes resulted from an independent t-test conducted on the smaller class subgroup from the 2003-2004 preschool group and the remaining students who were not in smaller classes using the NJASK3 math scores. These statistical findings were not significant

enough to provide support for the elimination or reduction of Head Start Fade by meeting the conditions of the study's theory.

There were 19 outcomes out of 100 Cohen's d tests that produced practically significant effect sizes on the first and second grade Terra Nova and the NJASK3 tests. Only eight of those 19 tests favored participation in small class sizes, and the vast majority of them resulted in small effect sizes. Also, none of the outcomes that produced practically significant effect sizes extended into third grade. Both the 2002-2003 preschool group and the 2003–2004 preschool group mean NJASK3 math and language arts test scores were nearly equal to or less than state and District Factor Group D/E scores. Thus, these additional data did not support the elimination of Head Start Fade among students who met the conditions of the study's theory.

Unexpected Results

Study findings were unexpected, based on the theoretical construct and the existing literature and research on the topics of early intervention and class-size research. A further and more extensive review of research and theoretic construct, however, highlighted several possible explanations for the results. First, the students who were members of the small-class subgroups did not meet the class size recommendations (13–18 students) established by the body of research for the duration of the study and, in many cases, the students did not meet that class size requirement for any one year of the study. In addition, the researcher was unable to account for two key contributing factors which have been associated with academic success in the majority of class-size and early intervention studies. Parental involvement has been shown in numerous studies; e.g., Edmondson (2009), Campbell and Ramey (1994), Reynolds (2000); to have significant

influence on achievement. Unfortunately, there was no way to measure parental involvement using the archived quantitative data available for this study. Likewise, membership in cohorts has been associated with creating a psychological sense of community – PSOC (Sarason, 1974) and (Finn et al. 2001) - and has shown significant influence on achievement. The students who attended this school district were not assigned to "true" cohorts for any of the years involved in the study. These missing components, especially the absence of small classes by research standards, provide a viable explanation for the unexpected results and confirm the need for rigorous implementation of early intervention and class size initiatives.

Conclusions

Study findings were not statistically significant enough to have further implications. The resulting data, however, did confirm the need for rigorous implementation of the theory's conditions to achieve success. The theoretic construct originally conceived in the Abecedarian study and further developed by the STAR study and the study design conceived by Clarke (2007) has been further developed by this study. This study's design could and should be replicated by other researchers on a larger scale and under different conditions. The value of this study, therefore, is to bring additional attention to an area of research that is ripe for exploration. By testing a theory using a carefully designed and thought-out research process, this researcher has shed light on a structural change (Deming, 2000) that promises to eliminate or greatly reduce Head Start Fade, especially for the nation's neediest children.

Recommendations for Policy, Practice and Future Research

Although the data from this study were not statistically significant enough to generalize or to make claims about causation, the extensive review of literature and research does provide evidence for the implementation of several policies. The findings from the Tennessee STAR study, the DuPont Study, SAGE, Challenge, and the Burke County study, with regard to CSR, are consistent and significant enough to suggest that states adopt policies that ensure small class sizes for all students in grades pre-k through four. In addition, the review of early intervention studies such as the Perry Child Development Centers, the Chicago Child-Parent Centers, and the Abecedarian Program suggest that early intervention should be offered as a part of the public education system. The findings from this study are significant because they illustrate the importance of implementing changes with rigor and attention to detail. Educators, administrators, and legislators must not adopt policies blindly or implement initiatives half-heartedly. It is imperative that the body of scientific-based research is examined closely and that changes are implemented correctly. The consequences of doing otherwise are too costly. As was the case in this study, and on a larger scale in the implementation of Indiana's Prime Time CSR initiative (Finn, 1998) and California's CSR initiative (Cobbold, 2005), the improper implementation of well-intentioned programs can lead to inconsistent or inconsequential outcomes.

Outcomes of this study have brought attention to the importance of implementing interventions with acute attention to research and theoretic construct. Based on the review of literature and research conducted in this study, the researcher's recommendations for practitioners are to implement the conditions of early intervention

(pre-k), duration (small classes of 13-18 students for five years), intensity (same teacher all day and every day), and heterogeneity (random assignment) in all elementary schools. Also, if possible, practitioners should integrate the concept of cohorts (psychological sense of community) and maximize parental involvement in education. This researcher stresses the importance of implementing all of the conditions proposed by the theoretic construct exactly as recommended to avoid inconsistent or non-significant results as was the case in this study.

Outcomes of this study did not produce enough statistically significant results to support the theory that meeting the conditions of early intervention, duration, intensity, and heterogeneity reduces or eliminates Head Start Fade. The research that went into the development of the study, however, did show promising benefits for the theory. Based on the review of literature and research the researcher can make specific recommendations for future researchers. Researchers should ensure that all conditions of the study are met by the population before beginning the study, especially the implementation of small class sizes as proposed by existing research and theory (n = 13-18). Researchers also need to have access to reliable and valid data for all of the study's participants. Conducting a study where the Head Start program is offered directly by a school district will most likely provide more reliable data than outsourced private programs. In addition, the optimal condition for testing this theory would be an experimental setting. Researchers should look for an environment where that can be created or where the existing setting is as close to experimental conditions as possible. If possible, future researchers should also test for outcomes other than academic achievement. Several researchers whose studies were examined by this literature review

have provided examples of how to do this; i.e., social outcomes, cost vs. benefit tests, classroom behavior, teacher morale, etc. Benefits might also be found in a similar study that is conducted in a different environment. For example, future researchers could replicate this study and use data obtained from students who have lower SES backgrounds but who reside in wealthy neighborhoods and/or school districts.

This study has shown the importance of implementing interventions with careful regard for the existing body of empirical research and attention to theoretic constructs. Educators should not only know the knowledge dynamic on the theory, they must use the research and implement it correctly. Conducting interventions for a short time or implementing reform in ways that are convenient (and not necessarily rigorous) may not, and most likely will not, produce the desired results. Policies and practices need to be formulated based on best practices with guidance from scientific-based research, sound theory, and the knowledge dynamic. If done effectively, future studies that test the theory of early intervention, duration, intensity, and heterogeneity should provide additional evidence for the cumulative positive influences (academically, socially, emotionally, and in terms of costs vs. benefits) of meeting these conditions for all students.

REFERENCES

Achilles, C. M. (2005). *Financing Class Size Reduction.* Greensboro, North Carolina: The SERVE Educational Laboratory.

Achilles, C. M. (1999). *Let's Put Kids First, Finally: Getting Class Size Right.* Thousand Oaks, California : Corwin Press Inc.

Achilles, C., & Finn, J. (2002). *The Role of School Leadership in Reform: A Case of Mistaken Identity.* New Orleans, LA: Paper presented at the annual meeting of the American Educational Research Association.

American Psychological Association (2006). *Publication Manual of the American Psychological Association; Fifth Edition.* Washington D.C: American Psychological Association.

Barnett, S. W. (1993). Does Head Start Fade Out? *Education Week* , 40-41.

Barnett, S. W. (1998). Long-Term Cognitive and Academic Effects of Early Childhood Education on Children in Poverty. *Preventive Medicine, 27* , 204-207.

Barnett, S. W. (1995). Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes. *The Future of Children, 5, 3*, 25-50.

Barnett, S. W. (2008). Preschool Education and Its Lasting Effects: Research and Policy Implications. *Education Policy Research Unit* , 1-2.

Barnett, S. W. (2003, March). Quality Preschool is a Good Investment. *Techniques: Front and Center* , 12-13.

Barnett, S. W. (2008). Report's Attack on Federal Preschool Proposals is Poorly Grounded and Misleading. *Education and the Public Interest Center* , 1-3.

Barnett, S. W. (2002). The Battle Over Head Start: What the Research Shows. *National Institute for Early Education Research Working Papers* , 1-5.

Barnett, S. W., & Masse, L. N. (2007). Comparative Benefit-Cost Analysis of the Abcedarian Program and its Policy Imlpications. *Economics of Education Review 26* , 113-125.

Belfield, C., & Levin, H. M. (2007). *The Price We Pay: Economic and Social Consequences of Inadequate Education.* University of Michigan: Brookings Institution Press.

Berliner, D. C. (2006). Our Impoverished View of Educational Research. *Teachers College Record; 108, 6*, 949-995.

Borman, G. D., & Hewes, G. M. (2002, Winter). The Long-term Effects and Cost-effectiveness of Success for All. *Educational Evaluation and Policy Analysis; 24, 4*, 243-262.

Boyd-Zaharias, J., & Pate-Bain, H. (2008). Class Matters - In and Out of School: Closing Gaps Requires Attention to Issues of Race and Poverty. *Phi Delta Kappan 90(1)* , 40-44.

Bracey, G. (1999). Poverty and Achievement. *Phi Delta Kappan, 77(7)* , 510-512.

Bracey, G. W., & Stellar, A. (2003, June). Long-term Studies of Preschool: Lasting Benefits Far Outweigh Costs. *Phi Delta Kappan; 84, 10* , 780-785.

Campbell, F. A., & Ramey, C. T. (1994). Effects of Early Intervention on Intellectual and Academic Achievement: A Follow-up Study of Children from Low-Income Families. *Child Development; 65*, 684-698.

Caputo, R. K. (2004). The Impact of Intergenerational Head Start Participation on Success Measures Among Adolescent Children. *Journal of Family and Economic Issues, 25, 2*, 199-223.

Caulfield, J. A. (1989). *The Role of Leadership in the Administration of Public Schools.* Union, New Jersey: Nevfield Press.

Clarke, D. T. (2007). *Exploration into the Head Start Fade Phenomenon.* Ypsilanti, Michigan: Eastern Michigan University - Dissertation.

Clarke, D. T., & Achilles, C. M. (2008). Testing Theory to Explain the Head Start Fade (The Fade Phenomenon). *Leadership on the Frontlines: Changes in Preparation and Practice* . Lancaster, PA: DEStech Publications, Inc.

Coalition for Evidence-Based Policy. (2002-2004). Perry Preschool Project (High Quality Preschool for Children from Disadvantaged Backgrounds). *Social Programs That Work* . Washington, D.C.: The Council for Excellence in Government.

Cobbold, T. (2005). *California Reduction* . Retrieved October 19, 2009, from ACT Council of Parents & Citizens Associations: http://www.schoolparents.canberra.net.au/California%20Reduction.pdf

Colvin, R. (2009, 3 10). *News-US-Obama-Education.* Retrieved 3 10, 2009, from Comcast.net: http://www.comcast.net/articles/news-general/20090310/NEWS-US-OBAMA-EDUCATION/

Conyers, L. M., Reynolds, A. J., & Ou, S.-R. (2003). The Effect of Early Childhood Intervention and Subsequent Special Education Services: Findings from the Chicago Child-Parent Centers. *Educational Evaluation and Policy Analysis; (2003, Spring) 25, 1*, 75-90.

Council, S. S. (2006). *Initial Report of the Start Strong Council.* Virginia.

Cunningham, G. K. (1998). *Assessment in the classroom: constructing and interpreting texts.* Falmer Press : London.

Currie, J. (2001). Early Childhood Education Programs. *The Journal of Economic Perspectives; Spring 2001, Vol. 15, Iss. 2* , 213-239.

Currie, J., & Duncan, T. (1994). *Does Head Start Make a Difference?: Labor and Population Program Working Paper.* New York: Alfred P. Sloan Foundation.

Demarest, E., Reisner, E., Anderson, L., Humphrey, D., Farquhar, E., & Stein, S. (1993). *Review of research on achieving the nation's readiness goal.* Retrieved April 13, 2009, from Washington, DC: U.S. Department of Education: http://www.ncrel.org/sdrs/areas/issues/students/earlycld/ea7lk5.htm

Edmondson, B. P. (2004). *The Effects of Parental Involvement and Small Class Sizes on the Academic Achievement Gap* . Ypsilanti: Eastern Michigan University Dissertation.

Egelson, P., Achilles, C. M., Harman, P., & Hood, A. (2005). *How Class Size Makes a Difference.* Greensboro, North Carolina: The SERVE Education Laboratory.

Fantuzzo, J. W., Rouse, H. L., McDermott, P. A., & Sekino, Y. (2005). Early Childhood Experiences and Kindergarten Success: A Population-Based Study of a Large Urban Setting. *School Psychology Review; 34, 4* , 571-590.

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "Why's" of Class Size: Student Behavior in Small Classes. *Review of Educational Research; 73(3)* , 321-350.

Finn, J. (1998). *Research on the Academic Effects of Small Class Size.* U.S. Department of Education: National Institute on the Education of At-Risk Students.

Garces, E., Thomas, D., & Currie, J. (2002, September). Longer-term Effects of Head Start. *The American Economic Review; 92, 4* , 999-1013.

Gayl, C. L. (2007). Expanding Voluntary Preschool Education: The Federal Role. *Policy Research Brief, 27, 3* , 9-10.

Goodman, A., & Sianesi, B. (2005, December). Early Education and Children's Outcomes: How Long do the Impacts Last? *Fiscal Studies; December 2005, 26, 4* , 513-549.

Griffin, S. (n.d.). *Promising Practices Network.* Retrieved July 22, 2009, from Chicago Public Schools: http://www.promisingpractices.net/program.asp?programid=98

Gromley, W. T., & Gayer, T. (2002). *Promoting School Readiness in Oklahoma: An Evaluation of Tulsa'a Pre-K Program* . Washington D.C.: Center for Research on Children in the U.S. (CROCUS).

Haller, E. J., & Knapp, T. R. (1985). Problems and Methodology in Educational Administration. *Educational Administration Quarterly 23, 3*, 157-168.

Hanushek, E. A. (1998). *The Evidence on Class Size.* Rochester, NY: W. Allen Wallis Institute of Political Economy.

Haskins, R., & Rouse, C. (2005). Colsing Achievement Gaps. *The Future of Children: Policy Brief* , 1-7.

Hellerich-Tuttle, L., & al., e. (1996). The Nebraska Head Start / Public School Early Childhood Transition Demonstration Project. *Head Start National Research Conference*, (pp. 1-21). Washington D.C.

Henry, G. T., Gordon, C. S., & Rickman, D. K. (2006). Early Education Policy Alternatives: Comparing Quality and Outcomes of Head Start and State Prekindergarten. *Educational Evaluation and Policy Analysis; 28, 1* , 77-88.

Hess. (2008). The New Stupid. *Educational Leadership - Decemeber/January* , 12-17.

Hodges, W., & Cooper, M. (1981). Head Start and Follow Through: Influences on Intellectual Development. *The Journal of Special Education; 15(2)* , 221- 238.

Hood, J. (1992, December). Caveat Emptor: The Head Start Scam: *Policy Analysis; 187*, 1-11.

Johnson, B. (2001). Toward a New Classification of Nonexperimental Quantitative Research. *Educational Researcher, 30* , 3-13.

Kafer, K. (2004, May). A Head Start for Poor Children? . *Backgrounder; 1755* , 1-4.

Kerlinger, F. N. (1968). The Doctoral Training of Research Specialists. *Teachers College Record, 69, 5*, 477-483.

Kozol, J. (1992). *Savage Inequalities.* New York : Harper Collins.

Lai, M. K., & York, S. E. (2008). A More Appropriate Determination of the Effectiveness of a Prekindergarten Initiative in Hawaiian Communities. *Curriculum Research and Development Group, University of Hawaii* , 1-24.

Lee, V. E. (2002). *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School.* Retrieved July 10, 2008 , from Washington, D.C.: Economic: http://www.epinet.org/books/starting_gate.html

Lee, V. E., & Loeb, S. (1994). *Where Do Head Start Attendees End Up? One Reason Why Preschool Effects Fade Out.* Ann Arbor, Michigan: American Educational Research Association 1992-1993 Grants Program.

Leedy, P. D., & Ormond, J. E. (2005). *Practical Research: Planning and Design, 8th Edition.* Upper Saddle River, New Jersey: Pearson Education, Inc. .

Lewis, A. C. (2005, November). More Than Just Cute Kids. *Phi Delta Kappan; 87, 3,* 179-180.

Lewis, A. (2006). We Think We Can . *Washington Commentary 88(1)* , 3-4.

Lewit, E. M., & Baker, L. S. (1997). Class Indicators: Class Size. *Future of Children,* 112-121.

Ludwig, J., & Phillips, D. (2007). The Benefits and Costs of Head Start. *Social Policy Report; 21(3)*, 1-20.

Madsen, D. (1992). *Successful Dissertations and Theses: A Guide to Graduate Student Research from Proposal to Completion, Second Edition.* San Francisco: Jossey Bass.

Magidson, J., & Sorbom, D. (1982). Adjusting for Confounding Factors in Quasi-experiments: Another Reanalysis of the Westinghouse Head Start Evaluation. *Educational Evaluation and Policy Analysis Fall, Vol. 4, #3* , 321-329.

Mann, E. A., & Reynolds, A. J. (2006, September). Early Intervention and Juvenille Deliquency Prevention: Evidence from the Chicago Longitudinal Study . *Social Work Research; 30(3)* , 153-168.

Manning, M., & Patterson, J. (2006). Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40. *Childhood Education; 83(2)* .

McGroder, S. M. (1990). *Head Start: What Do We Know About What Works?* . Washington D.C: Office of the Assistant Secretary for Planning and Evaluation U.S. Department of Health and Human Services.

McMillan, D., & Chavis, P. (1986). Sense of Community: A definitioin and Theory . *Journal of Community Psychology (14)* , 6-23.

McRobbie, J., Finn, J., & Harman, P. (1998). *Class size reduction: Lessons learned from experience.* Retrieved April 13, 2009, from Policy Brief #23, WestEd: http://www.wested.org/pub/docs/policy/class_red.htm

Milesi, C., & Gamoran, A. (2006). Effects of Class Size and Instruction on Kindergarten Achievement. *Educational Evaluation and Policy Analysis; 28(4)* , 287-314.

Moore, E., & Yzaguirre, R. (2004). Head Start's National Reporting System Fails Our Children. Here's Why. . *Education Week , June 9* , 40-41.

Mosteller, F. (1995). The Tennessee Study of Class Size in the Early Grades. *The Future of Children: Critical Issues for Children and Youths; 5(2)* , 113-127.

Mosteller, F., Light, R., & Sachs, J. (1996). Sustained Inquiry in Education . *Harvard Educational Review*, 66, 4, 797-828.

NJDOE. (2006). *Grades 3 and 4 New Jersey Assessment of Skills and Knowledge - Technical Report.* Trenton, NJ: New Jersey Department of Education .

NJDOE. (2005). *Grades 3 and 4 New Jersey Assessment of Skills and Knowledge Technical Report* . Trenton, NJ: New Jersey Department of Education .

NJDOE. (2008). *New Jersey School Report Card - 2008.* Retrieved February 13, 2009, from New Jersey Department of Education: http://education.state.nj.us/rc/rc08/index.html

NJDOE. (2009). *State of New Jersey: Department of Education.* Retrieved April 13, 2009, from NJ Department of Education District Factor Groups (DFG) for School Districts : http://www.state.nj.us/education/finance/sf/dfgdesc.shtml

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, Vol. 27, #3* , 245-261.

OHS. (2008). *U.S. Department of Health and Human Services: Administration for Children and Families*. Retrieved May 20, 2008, from Office of Head Start: http://www.acf.hhs.gov/programs/ohs/

Ou, S.-R. (2005). Pathways of Long-Term Effects of an Early Intervention Program on Educational Attainment: Findings from the Chicago Longitudinal Study . *Applied Developmental Psychology* , 578-611.

Pajares, F. (2007). Elements of a Proposal . *Available from the author* .

Pungello, E., Campbell, F. A., & Miller-Johnson, S. (2000). Benefits of High Quality Childcare for Low-Income Mothers: The Abecedarian Study. *Head Start National Research Conference* (p. 6). Washington, D.C.: Health Resources and Services Administration.

Pyrczak, F., & Bruce, R. R. (2005). *Writing Empirical Research Reports, Fifth Edition* . Glendale, California : Pyrczak Publishing.

Quinn, K. (2005). Beliefs About Parenting and How It Affects a Child's Development. *Eastern Michigan UniversityDissertation* .

Reynolds, A. J. (2000). *Success in Early Intervention: The Chicago Child-Parent Centers.* Lincoln, NE: University of Nebraska Press.

Reynolds, A. J., Temple, J. A., & Ou, S.-R. (2003). School-Based Early Intervention and Child Well-Being in the Chicago Longitudinal Study. *Child Welfare League of America; LXXII, 5*, 634-656.

Sandhu, I. K. (2008, December 30). *Make Your Child Smarter*. Retrieved December 30, 2008, from Brainy-Child.com: http://www.brainy-child.com/expert/terra-nova-achievement-test.shtml

Sarason, S. (1974). *The PSOC: Prospects for a Community Psychology.* San Francisco, CA: Jossey Bass.

Schaps, E. (2008, November). Missing in Action: The Non-Role of Research in Policy and Practice. *Education Week;* 24-26.

Schweinhart, L. J. (2003). *Benefits, Costs, and Explanation of the High/Scope Perry Preschool Program.* Tampa, Florida: Meeting of the Society for Research in Child Development.

Schweinhart, L. J. (2002, June). How the High/Scope Perry Preschool Study Grew: A Researcher's Tale. *Phi Delta Kappan; 32,* 19-26

Schweinhart, L. J. (2009). *Programs that work* . Retrieved Novemeber 13, 2009 , from Promising Practices Network : www.promisingpractices.net/program.asp?programid=128

Schweinhart, L. J. (2004). Significant Benefits: The High Scope Perry Preschool Study Through Age 27. *Texas Youth Commisssion: Prevention Summary* , 1-3.

Schweinhart, L. J. (2002). State-Funded Preschool Program Works for Children at Risk. *High/Scope Educational Research Foundation* , 1-4.

Schweinhart, L. J. (2000). The High/Scope Perry Preschool Study: A Case Study in Random Assignment. *Evaluation and Research in Education; 14, 3&4* , 136-147.

Schweinhart, L. J. (2000). The High/Scope Perry Preschool Study: A Case Study in Random Assignment. *Evaluation and Research in Education; 14, 3*, 136-147.

Sirin. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, 75, 3, 417-453.

Sirin, S. (2005). Socioeconomic Status and Academic Achievement: A Meta-analytic review of research. *Review of Educational Research* , 417-463.

Slaby, R., Loucks, S., & Stelwagon, P. (2005). Why is Preschool Essential in Closing the Achievement Gap? . *Educational Leadership and Administration; Fall 2005, 17(47)* , 47-60.

Soriano, D., Duenas, M., & LeBlanc, P. (2006, August). The Short and Long Term Effects of Head Start Education and No Child Left Behind. *August 2006 Meeting of theAssociation of Teacher Educators*, Philadelphia, PA, 1-14

Starr, A. (2002, April). Does Universal Preschool Pay? *Business Week; 3780* , 98-99.

Temple, J. A., & Reynolds, A. J. (2007). Benefits and Costs of Investments in Preschool Education: Evidence from the Child-Parent Centers and Related Programs. *Economics of Education Review (26)* , 126-144.

Texas Youth Commission. (1993). Significant Benefits: The High/Scope Perry Preschool Study Through Age 27. *Monographs of the High/Scope Educational Research Foundation, 10* , 134-136.

Tienken, C. (2008, April). Rankings of International Achievement Test Performance and Economic Strength: Correlation ot Conjecture? *International Journal of Education Policy & Leadership*, 3, 4, 1-15.

Tienken, C. H., & Achilles, C. M. (2009). Relationship Between Class Size and Students' Opportunity to Learn Writing in Middle School. *Research in the Schools 16, 1* , 13-24.

Westinghouse Learning Corporation. (1969). The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Afective Development (Executive Summary). *Clearinghouse for Federal Scientific and Technical Information* , 12.

Wilkins, C. (. (2007). *Office of Head Start* . Retrieved 8 20, 2007, from About the Office of Head Start: Program Fact Sheets : www.acg.hhs.gov/programs/hsb/about/index.html

Witte, R. S., & Witte, J. S. (2007). *Statistics* . Hoboken, NJ: John Wiley & Sons, Inc. .

Zill, N., Resnick, G., Kim, K., O'Donnell, K., Sorongon, A., McKey, R. H., et al. (2003). *Head Start FACES 2000: A Whole Child Perspective on Program Performance. Fourth Progress Report.* Alexandria, Virginia: Head Start Information and Publication Center.

APPENDIX A: Institutional Review Board – Approval for Research

OFFICE OF INSTITUTIONAL REVIEW BOARD
SETON HALL UNIVERSITY

September 30, 2009

Christopher Huss
12 Caro Court
Red Bank, NJ 07701

Dear Mr. Huss,

The Seton Hall University Institutional Review Board has reviewed your research proposal entitled "The Influence of Small Class Size, Duration, Intensity, and Heterogeneity on Head Start Fade" and has approved it as submitted under exempt status.

Enclosed for your records is the signed Request for Approval form.

Please note that, where applicable, subjects must sign and must be given a copy of the Seton Hall University current stamped Letter of Solicitation or Consent Form before the subjects' participation. All data, as well as the investigator's copies of the signed Consent Forms, must be retained by the principal investigator for a period of at least three years following the termination of the project.

Should you wish to make changes to the IRB approved procedures, the following materials must be submitted for IRB review and be approved by the IRB prior to being instituted:

- Description of proposed revisions;
- *If applicable,* any new or revised materials, such as recruitment fliers, letters to subjects, or consent documents; and
- *If applicable,* updated letters of approval from cooperating institutions and IRBs.

At the present time, there is no need for further action on your part with the IRB.

In harmony with federal regulations, none of the investigators or research staff involved in the study took part in the final decision.

Sincerely,

Mary F. Ruzicka, Ph.D.

Mary F. Ruzicka, Ph.D.
Professor
Director, Institutional Review Board

cc: Dr. Charles M. Achilles

REQUEST FOR APPROVAL OF RESEARCH, DEMONSTRATION OR RELATED ACTIVITIES INVOLVING HUMAN SUBJECTS

All material must be typed.

PROJECT TITLE: The Influence of Small Class Size, Duration, Intensity, and Heterogeneity on Head Start Fade

CERTIFICATION STATEMENT:

In making this application, I(we) certify that I(we) have read and understand the University's policies and procedures governing research, development, and related activities involving human subjects. I (we) shall comply with the letter and spirit of those policies. I(we) further acknowledge my(our) obligation to (1) obtain written approval of significant deviations from the originally-approved protocol BEFORE making those deviations, and (2) report immediately all adverse effects of the study on the subjects to the Director of the Institutional Review Board, Seton Hall University, South Orange, NJ 07079.

Christopher Huss — 9/05/09
RESEARCHER(S) OR PROJECT DIRECTOR(S) — DATE

**Please print or type out names of all researchers below signature.
Use seperate sheet of paper, if necessary.**

My signature indicates that I have reviewed the attached materials and consider them to meet IRB standards.

Charles M. Achilles Ed D. Dr. Charles Achilles — 9/9/09
RESEARCHER'S ADVISOR OR DEPARTMENTAL SUPERVISOR — DATE

Please print or type out name below signature

The request for approval submitted by the above researcher(s) was considered by the IRB for Research Involving Human Subjects Research at the mtg Sept 8 2009 meeting.

The application was approved ✓ not approved ___ by the Committee. Special conditions were ___ were not ✓ set by the IRB. (Any special conditions are described on the reverse side.)

Nancy F. Ruzicka, Ph.D. — 9/30/09
DIRECTOR,
SETON HALL UNIVERSITY INSTITUTIONAL
REVIEW BOARD FOR HUMAN SUBJECTS RESEARCH — DATE

Seton Hall University
3/2005

Township of Union Public Schools
The Presidential Model School District

Theodore A. Jakubowski, Ed.D.
Superintendent of Schools

September 10, 2009

To Whom It May Concern:

As Superintendent of the Township of the Union Public School District, I hereby grant permission for Christopher Huss to access information related to this district's groups of students who attended pre-kindergarten in 2002-2003 and in 2003-2004. I also grant permission for Mr. Huss to have access to these students' information through the 2007-2008 school years, including NJASK and Terra Nova scores. Mr. Huss has agreed to have a third party Union Township employee provide him with coded information which will be stripped of identifiers to ensure each child's confidentiality. Mr. Huss has agreed to use this archived information solely for the purpose of conducting his study on *The Influence of Small Class Size, Duration, Intensity, and Heterogeneity on Head Start Fade* unless granted permission otherwise. In addition, Mr. Huss has agreed to share the results of his study with the Township of Union Public School District Board of Education upon completion of his research.

Sincerely,

Theodore A. Jakubowski

THEODORE A. JAKUBOWSKI, Ed.D.
Superintendent

TAJ:ps

2369 Morris Avenue Union, New Jersey 07083-5712 Tel: (908) 851-6420
Fax: (908) 851-6421 E-mail: tjakubowski@twpunionschools.org

APPENDIX B: New Jersey District Factor Group Information

The New Jersey State Department of Education's definition of District Factor Grouping (DFG) for school districts:

The New Jersey Department of Education introduced the District Factor Grouping system (DFG) in 1975. This system provides a means of ranking school districts in New Jersey by their socioeconomic status (SES). The first DFG was based on data from the 1970 decennial Census. A revision was made in 1984 to take into account new data from the 1980 Census and to slightly change the theoretical model of socioeconomic status. Following is a description of the work undertaken in the construction of the third DFG, reflecting data from the 1990 Census.

The DFG was motivated by research conducted in the late 1960's and early 1970's that showed a strong relationship between socioeconomic status and educational outcomes. The creators of the DFG were concerned that educational policymakers, after reviewing the educational outcomes obtained in different circumstances, would make unjustified inferences about the importance of various, school-based inputs to the educational process. Because the research showed that students (i.e. what students bring to school, including socialization that takes place before they step inside the school building) are the most important determinant of educational outcomes, the effectiveness of school systems cannot be sensibly judged without reference to the socioeconomic background of their students.

The DFG is an index of socioeconomic status that is created using data for several "indicators" available in the decennial Census of Population. Socioeconomic status cannot be measured directly. Rather, the literature holds that it is a function of other, measurable quantities (traditionally, the basic three are income, occupation, and education). Therefore, the DFG is a composite statistical index created using statistical procedures, a "model" of socioeconomic status, and input data for various socioeconomic traits. Seven indices were developed from the census data as follows: 1) Percent of population with no high school diploma 2) Percent with some college 3) Occupation 4) Population density 5) Income 6) Unemployment 7) Poverty.

These seven indices were utilized in a principal components analysis to produce a statistical score which was used to rank the districts. Districts were then grouped so that each group would consist of districts having factor scores within an interval of one tenth of the distance between the highest and lowest scores (NJDOE, 2009).

Visit http://www.nj.gov/education/finance/sf/dfg.shtml for a complete list of New Jersey's DFGs.

APPENDIX B (Continued)

(Data obtained from: http://www.nj.gov/education/finance/sf/dfg.shtml)

Table 19
Average District Factor Group Assessment Scores by 2000

	ESPA		GEPA			HSPA	
	Lang Arts	Math	Lang Arts	Math	Science	Lang Arts	Math
A	208.9	199.4	201.0	191.3	201.4	209.9	197.4
B	214.1	210.3	213.4	206.4	217.4	221.0	212.3
CD	218.3	219.0	217.2	208.7	224.0	224.7	216.2
DE	221.8	224.8	221.9	214.6	228.6	228.3	220.5
FG	224.1	229.3	224.9	220.5	232.3	230.9	226.2
GH	226.1	233.4	227.8	225.7	235.2	234.6	231.2
I	230.6	240.4	233.4	231.8	240.1	240.1	239.6
J	233.8	247.1	238.5	238.6	244.0	244.1	244.8

Correlations

[DataSet2]

		DisFacGroup	ESPALA	ESPAMATH
DisFacGroup	Pearson Correlation	1	.993**	.990**
	Sig. (2-tailed)		.000	.000
	N	8	8	8
ESPALA	Pearson Correlation	.993**	1	.999**
	Sig. (2-tailed)	.000		.000
	N	8	8	8
ESPAMATH	Pearson Correlation	.990**	.999**	1
	Sig. (2-tailed)	.000	.000	
	N	8	8	8

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix C: Tables and Statistical Data

Table 4: 2002-2003 Preschool Participants - Class Sizes All Students (n = 134)
Ethnic Codes: 1=White 2=Black 3=Hispanic 4=American Indian 5=Asian 6=Pacific Islander 7=Multi-Ethnic

Ethnic Code	Class Sizes by Grades					
	Pre-K	K	1	2	3	Mean
1	20	27	23	27	23	24.0
2	22	29	20	27	22	24.0
2	22	27	20	27	23	23.8
1	22	29	20	27	23	24.2
1	22	29	23	27	23	24.8
1	20	27	23	27	23	24.0
1	22	29	23	27	22	24.6
1	20	25	23	27	22	23.4
3	21	25	20	27	23	23.2
1	20	25	23	27	23	23.6
1	20	27	23	27	22	23.8
1	21	22	19	26	20	21.6
1	22	27	20	26	22	23.4
5	22	29	23	26	23	24.6
3	21	22	20	26	20	21.8
1	17	24	20	26	20	21.4
5	17	25	20	26	22	22.0
1	20	25	22	26	23	23.2
1	20	25	23	26	23	23.4

1	20	25	23	26	23	23.4
1	20	27	22	26	22	23.4
1	22	29	20	26	23	24.0
5	20	27	20	26	23	23.2
1	22	27	22	26	23	24.0
2	17	24	21	26	21	21.8
1	21	24	19	26	20	22.0
1	20	29	23	26	22	24.0
1	22	29	20	26	23	24.0
3	17	23	19	26	20	21.0
3	25	20	18	25	23	22.2
1	25	21	18	25	21	22.0
3	21	23	21	25	20	22.0
1	25	21	18	25	23	22.4
2	20	21	19	25	24	21.8
1	25	20	19	25	24	22.6
1	20	21	21	25	23	22.0
1	21	24	21	25	20	22.2
1	25	21	21	25	23	23.0
2	17	23	20	25	21	21.2
1	25	20	21	25	24	23.0
1	25	21	19	25	23	22.6
1	25	21	21	25	23	23.0
2	20	21	18	25	24	21.6

3	20	21	19	25	24	21.8
1	25	20	18	25	23	22.2
2	20	21	21	25	21	21.6
2	21	24	18	25	23	22.2
5	25	20	19	25	23	22.4
1	25	21	18	25	21	22.0
1	20	21	18	25	21	21.0
1	25	21	18	25	24	22.6
2	20	20	19	25	21	21.0
2	20	20	18	25	22	21.0
3	25	21	21	25	23	23.0
1	25	21	21	25	24	23.2
1	20	21	21	25	23	22.0
1	25	21	19	25	23	22.6
1	20	27	23	24	22	23.2
1	20	27	20	24	22	22.6
2	20	27	23	24	23	23.4
2	20	29	20	24	23	23.2
1	22	27	20	24	22	23.0
1	22	27	23	24	23	23.8
5	21	29	21	22	21	22.8
2	20	29	24	22	23	23.6
5	21	23	19	22	21	21.2
5	20	25	24	22	23	22.8
2	17	23	19	22	19	20.0

2	18	28	24	22	22	22.8
2	21	23	24	22	22	22.4
5	20	28	22	22	21	22.6
1	18	25	21	22	22	21.6
1	19	24	21	21	19	20.8
1	25	24	21	21	21	22.4
3	21	22	18	21	21	20.6
1	25	23	21	21	20	22.0
2	17	22	20	21	21	20.2
5	20	28	24	21	22	23.0
1	25	28	21	21	17	22.4
5	17	24	20	21	21	20.6
3	17	23	19	21	21	20.2
1	25	23	21	21	20	22.0
2	21	23	20	21	21	21.2
1	21	23	18	21	19	20.4
1	25	24	21	21	19	22.0
6	25	24	18	21	17	21.0
5	20	28	23	21	21	22.6
1	25	28	18	21	21	22.6
2	21	23	19	21	20	20.8
2	19	24	21	21	19	20.8
5	19	28	18	21	21	21.4
5	18	28	24	21	21	22.4

3	25	23	21	21	20	22.0
3	21	23	19	21	20	20.8
3	18	28	21	21	22	22.0
1	25	23	21	21	17	21.4
1	18	25	21	21	22	21.4
1	19	24	21	21	19	20.8
3	21	23	20	21	21	21.2
1	21	22	18	21	21	20.6
1	21	22	20	21	21	21.0
1	25	24	18	21	17	21.0
5	18	28	18	21	17	20.4
2	21	23	19	21	20	20.8
2	21	24	21	21	20	21.4
2	21	23	18	21	21	20.8
3	17	22	19	20	21	19.8
1	25	24	21	20	19	21.8
5	21	29	21	20	22	22.6
2	21	29	22	20	21	22.6
5	18	28	23	20	22	22.2
5	21	23	20	20	21	21.0
5	25	23	21	20	17	21.2
5	25	28	21	20	20	22.8
5	20	25	21	20	22	21.6
3	20	23	21	20	21	21.0

1	18	28	23	20	22	22.2
2	20	25	21	19	21	21.2
1	18	29	23	19	21	22.0
3	19	28	21	19	23	22.0
3	20	28	24	19	22	22.6
5	20	29	23	19	23	22.8
3	18	25	21	19	21	20.8
3	20	29	21	19	21	22.0
5	20	28	21	19	22	22.0
3	21	24	21	18	20	20.8
1	25	28	21	18	19	22.2
1	25	23	21	18	17	20.8
2	25	23	21	18	21	21.6
3	19	23	21	18	20	20.2
1	25	28	18	18	17	21.2
3	17	28	21	18	21	21.0
1	19	24	21	18	20	20.4
1	21	22	19	18	20	20.0

Table 5: 2003 – 2004 Preschool Participants - Class Sizes All Students (n = 112)
Ethnic Codes: 1=White 2=Black 3=Hispanic 4=American Indian 5=Asian 6=Pacific Islander 7=Multi-Ethnic

Gender	Ethnic Code	Class Sizes by Grade					
		Pre-K	K	1	2	3	Mean
M	2	20	22	20	25	20	21.4
M	2	17	22	21	20	23	20.6
F	2	13	26	22	22	23	21.2
M	3	15	20	18	19	19	18.2
M	1	18	20	18	19	19	18.8
F	3	13	22	22	24	24	21.0
F	2	15	20	19	20	18	18.4
F	5	17	27	23	22	21	22.0
M	3	20	26	19	25	21	22.2
F	2	10	21	18	18	19	17.2
M	2	17	26	21	22	22	21.6
M	5	15	20	18	20	19	18.4
M	1	16	21	18	18	19	18.4
F	3	17	24	20	22	21	20.8
F	1	20	26	20	26	21	22.6
M	1	17	29	20	21	21	21.6
M	1	16	21	22	21	21	20.2
F	2	15	19	17	19	19	17.8
F	5	17	25	21	22	22	21.4
M	2	20	22	20	25	21	21.6
F	3	17	22	19	23	24	21.0
M	2	16	21	21	22	23	20.6
M	2	20	22	21	26	21	22.0
M	3	16	21	18	18	19	18.4
M	5	17	27	20	21	21	21.2
F	3	12	18	21	21	23	19.0
M	1	15	19	21	20	18	18.6
M	1	20	22	17	25	21	21.0
F	1	15	20	21	19	20	19.0
M	1	16	18	22	21	23	20.0
M	1	17	27	16	19	20	19.8
M	5	16	21	21	21	21	20.0

Table 5 Continued

Gender	Ethnic Code	Class Sizes					
		Pre-K	K	1	2	3	Mean
F	5	13	29	20	21	21	20.8
F	1	16	18	22	22	21	19.8
F	1	17	22	22	24	23	21.6
M	2	18	20	18	20	20	19.2
F	5	17	22	21	24	22	21.2
M	1	16	18	22	21	21	19.6
M	1	20	26	19	25	20	22.0
F	5	17	27	20	22	21	21.4
M	1	16	20	20	18	19	18.6
M	5	17	22	20	24	24	21.4
M	3	15	19	21	19	20	18.8
F	5	16	19	22	21	21	19.8
M	2	13	24	20	23	21	20.2
M	1	10	20	18	18	20	17.2
F	1	12	21	21	21	23	19.6
M	1	17	26	22	20	24	21.8
M	1	20	22	20	26	20	21.6
M	1	20	26	17	25	21	21.8
F	1	20	26	20	25	21	22.4
M	1	20	22	17	26	21	21.2
F	2	10	20	20	18	19	17.4
F	2	10	20	18	18	19	17.0
M	1	17	22	20	23	23	21.0
F	1	17	22	20	23	23	21.0
M	2	20	22	17	26	21	21.2
F	3	16	21	20	18	19	18.8
M	1	17	29	20	21	19	21.2
M	2	16	21	21	21	23	20.4
F	3	16	21	18	18	20	18.6
F	1	16	21	20	18	20	19.0
M	1	16	21	18	18	19	18.4
F	5	16	21	21	22	23	20.6
M	1	16	18	21	22	21	19.6
M	2	16	18	22	21	21	19.6
F	3	13	22	22	23	23	20.6
M	1	20	22	20	26	20	21.6
F	2	18	19	19	19	18	18.6

Table 5 Continued

Gender	Ethnic Code	Class Sizes					
		Pre-K	K	1	2	3	Mean
M	1	20	22	21	26	21	22.0
M	2	12	19	22	21	21	19.0
M	2	18	20	17	20	18	18.6
F	3	20	26	21	25	21	22.6
M	5	17	24	16	19	21	19.4
F	5	17	22	22	24	23	21.6
F	5	20	22	20	26	21	21.8
F	2	18	20	17	20	18	18.6
M	1	16	20	20	18	20	18.8
M	5	18	20	17	20	18	18.6
F	1	16	21	21	22	21	20.2
M	2	20	22	20	26	20	21.6
M	2	15	20	21	19	18	18.6
F	1	17	27	20	22	21	21.4
M	1	20	22	19	26	21	21.6
F	3	20	22	19	26	21	21.6
F	2	15	20	21	20	20	19.2
F	3	16	20	18	18	20	18.4
F	2	17	27	20	23	21	21.6
F	3	10	21	18	18	20	17.4
M	5	12	21	22	22	21	19.6
M	5	13	24	20	21	21	19.8
M	3	13	22	20	22	24	20.2
F	5	13	27	20	23	21	20.8
F	1	16	21	20	18	19	18.8
F	2	17	26	22	23	24	22.4
F	5	20	22	20	26	21	21.8
F	5	17	29	20	21	21	21.6
M	5	13	29	20	21	21	20.8
F	2	17	25	19	23	22	21.2
M	2	16	21	20	18	19	18.8
F	1	17	22	21	22	24	21.2
F	2	18	20	21	19	18	19.2
F	1	12	19	21	21	23	19.2
M	1	17	24	20	23	21	21.0
F	3	18	20	17	19	18	18.4
M	1	20	22	20	25	21	21.6

Table 5 Continued

Gender	Ethnic Code	Class Sizes					
		Pre-K	K	1	2	3	Mean
M	5	18	20	19	19	20	19.2
M	1	17	27	20	22	21	21.4
F	2	16	18	21	21	21	19.4
F	2	16	21	20	18	19	18.8
F	2	18	20	21	19	18	19.2
F	1	20	26	17	26	21	22.0

Table 9: 2002 – 2003 – Independent T-test: First Random Sample

[DataSet1] F:\SPSS COHORT INFO\Co1RandomSample.sav

Group Statistics

	ClSz Code	N	Mean	Std. Deviation	Std. Error Mean
TNR1	Small Classes	31	79.42	20.288	3.644
	Not Small Classes	31	74.65	23.609	4.24
TNLA1	Small Classes	31	76.9	21.618	3.883
	Not Small Classes	31	83.32	13.415	2.409
TNM1	Small Classes	31	76.39	21.279	3.822
	Not Small Classes	31	75.77	19.209	3.45
TNTOT1	Small Classes	31	79.84	19.308	3.468
	Not Small Classes	31	81.42	15.813	2.84

TNR2	Small Classes	31	66.74	25.95	4.661
	Not Small Classes	31	72.74	21.544	3.869
TNLA2	Small Classes	31	80.32	18.375	3.3
	Not Small Classes	31	78.65	20.009	3.594
TNM2	Small Classes	31	73.65	23.919	4.296
	Not Small Classes	31	76.68	22.152	3.979
TNTOT2	Small Classes	31	76.65	22.463	4.034
	Not Small Classes	31	80.97	18.444	3.313
NJASK3LA	Small Classes	31	218.48	14.697	2.64
	Not Small Classes	31	221.26	15.095	2.711
NJASK3M	Small Classes	31	234.68	26.844	4.821
	Not Small Classes	31	237.94	22.187	3.985

Table 9a: 2002-2003 Independent Samples Test with First Random Sample

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper

TNR1	Equal variances assumed	2.267	0.137	0.854	60	0.397	4.774	5.591	-6.409	15.957
	Equal variances not assumed			0.854	58.672	0.397	4.774	5.591	-6.414	15.963
TNLA1	Equal variances assumed	6.732	0.012	-1.405	60	0.165	-6.419	4.57	-15.56	2.721
	Equal variances not assumed			-1.405	50.12	0.166	-6.419	4.57	-15.597	2.758
TNM1	Equal variances assumed	0.558	0.458	0.119	60	0.906	0.613	5.149	-9.686	10.912
	Equal variances not assumed			0.119	59.383	0.906	0.613	5.149	-9.688	10.914
TNTOT1	Equal variances assumed	1.941	0.169	-0.353	60	0.726	-1.581	4.482	-10.547	7.386
	Equal variances not assumed			-0.353	57.757	0.726	-1.581	4.482	-10.554	7.393
TNR2	Equal variances assumed	0.723	0.398	-0.99	60	0.326	-6	6.058	-18.117	6.117
	Equal variances not assumed			-0.99	58.036	0.326	-6	6.058	-18.125	6.125
TNLA2	Equal variances assumed	0.463	0.499	0.344	60	0.732	1.677	4.879	-8.082	11.437

	Equal variances not assumed			0.344	59.569	0.732	1.677	4.879	-8.084	11.439
TNM2	Equal variances assumed	1.231	0.272	-0.518	60	0.606	-3.032	5.855	-14.744	8.68
	Equal variances not assumed			-0.518	59.65	0.606	-3.032	5.855	-14.746	8.681
TNTOT2	Equal variances assumed	1.326	0.254	-0.828	60	0.411	-4.323	5.22	-14.764	6.119
	Equal variances not assumed			-0.828	57.81	0.411	-4.323	5.22	-14.772	6.127
NJASK3LA	Equal variances assumed	0.348	0.557	-0.733	60	0.466	-2.774	3.784	-10.343	4.795
	Equal variances not assumed			-0.733	59.957	0.466	-2.774	3.784	-10.343	4.795
NJASK3M	Equal variances assumed	0.431	0.514	-0.521	60	0.604	-3.258	6.255	-15.77	9.254
	Equal variances not assumed			-0.521	57.946	0.604	-3.258	6.255	-15.779	9.263

Table 9b Cohen's d 2002 -2003 Test for Practical Significance

2002 - 2003 – First Random Sample – Cohen's d	Test for Practical Significance			
Standardized Test	Mean of Smaller Classes	Mean of Larger Classes	Pooled Standard Deviation	Cohen's d (Effect Size)
TerraNova1 Reading	79.42	74.65	21.96	0.22
TerraNova1 LangArts	76.90	83.32	18.13	(0.35)
TerraNova1Math	76.39	75.77	20.11	0.03
TerraNova1Total	79.84	81.42	17.52	(0.09)
TerraNova2Read ing	66.74	72.74	23.85	(0.25)
TerraNova2 LangArts	80.32	78.65	19.07	0.09
TerraNova2 Math	73.65	76.68	22.91	(0.13)
TerraNova2 Total	76.65	80.97	20.50	(0.21)
NJASK3 LangArts	218.48	221.26	14.84	(0.19)
NJASK3 Math	234.68	237.94	24.48	(0.13)

Table 10: 2002 – 2003 Independent T-test – Second Random Sample

[DataSet2] F:\SPSS COHORT INFO\Co1RandSamp(2).sav

Group Statistics

	ClSzCode	N	Mean	Std. Deviation	Std. Error Mean
TNR1	Small Classes	31	79.42	20.288	3.644
	Not Small Classes	31	76.84	22.426	4.028
TNLA1	Small Classes	31	76.9	21.618	3.883

	Not Small Classes	31	81.16	13.54	2.432
TNM1	Small Classes	31	76.39	21.279	3.822
	Not Small Classes	31	80.32	18.054	3.243
TNTOT1	Small Classes	31	79.84	19.308	3.468
	Not Small Classes	31	82.58	16.399	2.945
TNR2	Small Classes	31	66.74	25.95	4.661
	Not Small Classes	31	76.65	23.289	4.183
TNLA2	Small Classes	31	80.32	18.375	3.3
	Not Small Classes	31	81.87	19.105	3.431
TNM2	Small Classes	31	73.65	23.919	4.296
	Not Small Classes	31	76.26	24.873	4.467
TNTOT2	Small Classes	31	76.65	22.463	4.034
	Not Small Classes	31	81.87	20.401	3.664
NJASK3LA	Small Classes	31	218.48	14.697	2.64
	Not Small Classes	31	225.06	13.604	2.443
NJASK3M	Small Classes	31	234.68	26.844	4.821
	Not Small Classes	31	238.39	22.08	3.966

Table 10a: 2002 – 2003 Independent t-test Second Random Sample

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TNR1	Equal variances assumed	0.293	0.59	0.475	60	0.636	2.581	5.431	-8.284	13.445

	Equal variances not assumed			0.475	59.407	0.636	2.581	5.431	-8.286	13.447
TNLA1	Equal variances assumed	6.156	0.016	-0.929	60	0.356	-4.258	4.582	-13.422	4.906
	Equal variances not assumed			-0.929	50.398	0.357	-4.258	4.582	-13.458	4.942
TNM1	Equal variances assumed	1.224	0.273	-0.785	60	0.435	-3.935	5.012	-13.961	6.09
	Equal variances not assumed			-0.785	58.45	0.436	-3.935	5.012	-13.967	6.096
TNTOT1	Equal variances assumed	1.872	0.176	-0.603	60	0.549	-2.742	4.55	-11.843	6.359
	Equal variances not assumed			-0.603	58.468	0.549	-2.742	4.55	-11.848	6.364
TNR2	Equal variances assumed	0.125	0.725	-1.581	60	0.119	-9.903	6.262	-22.43	2.624
	Equal variances not assumed			-1.581	59.311	0.119	-9.903	6.262	-22.433	2.627
TNLA2	Equal variances assumed	0.01	0.92	-0.325	60	0.746	-1.548	4.761	-11.071	7.975
	Equal variances not assumed			-0.325	59.909	0.746	-1.548	4.761	-11.072	7.975
TNM2	Equal variances assumed	0.223	0.639	-0.422	60	0.675	-2.613	6.198	-15.01	9.784
	Equal variances not assumed			-0.422	59.908	0.675	-2.613	6.198	-15.011	9.785
TNTOT2	Equal variances assumed	0.67	0.416	-0.959	60	0.341	-5.226	5.45	-16.127	5.676
	Equal variances not assumed			-0.959	59.452	0.342	-5.226	5.45	-16.129	5.678

NJASK 3LA	Equal variances assumed	0.35	0.556	-1.83	60	0.072	-6.581	3.597	-13.775	0.614
	Equal variances not assumed			-1.83	59.645	0.072	-6.581	3.597	-13.776	0.615
NJASK 3M	Equal variances assumed	1.499	0.226	-0.594	60	0.555	-3.71	6.243	-16.197	8.778
	Equal variances not assumed			-0.594	57.846	0.555	-3.71	6.243	-16.207	8.787

Table 10b: 2002 – 2003 Cohen's d Test for Practical Significance

2002 - 2003 – Second Random Sample – Cohen's d		Test for Practical Significance		
Standardized Test	**Mean of Smaller Classes**	**Mean of Larger Classes**	**Pooled Standard Deviation**	**Cohen's d (Effect Size)**
TerraNova1 Reading	79.42	76.84	21.25	0.12
TerraNova1 LangArts	76.90	81.16	18.02	(0.24)
TerraNova1Math	76.39	80.32	19.67	(0.20)
TerraNova1Total	79.84	82.58	17.82	(0.15)
TerraNova2Reading	66.74	76.65	24.96	(0.40)
TerraNova2 LangArts	80.32	81.87	18.61	(0.08)
TerraNova2 Math	73.65	76.26	24.24	(0.11)
TerraNova2 Total	76.65	81.87	21.44	(0.24)
NJASK3 LangArts	218.48	225.06	14.43	(0.46)
NJASK3 Math	234.68	238.39	24.45	(0.15)

Table 11: 2002 – 2003 - Independent T-Test – First Matched Pair

[DataSet2] F:\SPSS COHORT INFO\Cohort1HandPicked.sav

Group Statistics

	ClSzCode	N	Mean	Std. Dev.	Std. Error Mean
TNR1	Small Classes	31	79.42	20.288	3.644
	Not Small Classes	31	68.71	21.454	3.853
TNLA1	Small Classes	31	76.9	21.618	3.883
	Not Small Classes	31	75.74	18.724	3.363
TNM1	Small Classes	31	76.39	21.279	3.822
	Not Small Classes	31	67.1	22.176	3.983
TNTOT1	Small Classes	31	79.84	19.308	3.468
	Not Small Classes	31	74.19	18.941	3.402
TNR2	Small Classes	31	66.74	25.95	4.661
	Not Small Classes	31	65.1	23.526	4.225
TNLA2	Small Classes	31	80.32	18.375	3.3
	Not Small Classes	31	73.13	21.041	3.779
TNM2	Small Classes	31	73.65	23.919	4.296
	Not Small Classes	31	69.61	24.981	4.487
TNTOT2	Small Classes	31	76.65	22.463	4.034
	Not Small Classes	31	72.61	21.227	3.812
NJASK3LA	Small Classes	31	218.48	14.697	2.64
	Not Small Classes	31	215.35	16.337	2.934

NJASK3M	Small Classes	31	234.68	26.844	4.821
	Not Small Classes	31	229.32	23.379	4.199

Table 11a: 2002 -2003 Independent Samples Test – First Matched Pair

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TNR1	Equal variances assumed	0.306	0.582	2.019	60	0.048	10.71	5.303	0.102	21.318
	Equal variances not assumed			2.019	59.813	0.048	10.71	5.303	0.101	21.318
TNLA1	Equal variances assumed	1.574	0.214	0.226	60	0.822	1.161	5.137	-9.114	11.436
	Equal variances not assumed			0.226	58.802	0.822	1.161	5.137	-9.118	11.441
TNM1	Equal variances assumed	0.08	0.779	1.683	60	0.098	9.29	5.52	-1.751	20.332
	Equal variances not assumed			1.683	59.898	0.098	9.29	5.52	-1.751	20.332
TNTOT1	Equal variances assumed	0.517	0.475	1.162	60	0.25	5.645	4.858	-4.072	15.362
	Equal variances not assumed			1.162	59.978	0.25	5.645	4.858	-4.072	15.362

TNR2	Equal variances assumed	0.017	0.897	0.262	60	0.795	1.645	6.291	-10.939	14.229
	Equal variances not assumed			0.262	59.432	0.795	1.645	6.291	-10.941	14.232
TNLA2	Equal variances assumed	0.667	0.417	1.434	60	0.157	7.194	5.017	-2.842	17.229
	Equal variances not assumed			1.434	58.931	0.157	7.194	5.017	-2.846	17.233
TNM2	Equal variances assumed	0.027	0.871	0.649	60	0.519	4.032	6.212	-8.393	16.458
	Equal variances not assumed			0.649	59.887	0.519	4.032	6.212	-8.394	16.458
TNTOT2	Equal variances assumed	0.081	0.777	0.726	60	0.47	4.032	5.551	-7.071	15.135
	Equal variances not assumed			0.726	59.809	0.47	4.032	5.551	-7.072	15.136
NJASK3LA	Equal variances assumed	1.56	0.217	0.793	60	0.431	3.129	3.947	-4.766	11.024
	Equal variances not assumed			0.793	59.34	0.431	3.129	3.947	-4.768	11.026
NJASK3M	Equal variances assumed	0.465	0.498	0.838	60	0.406	5.355	6.394	-7.434	18.144
	Equal variances not assumed			0.838	58.889	0.406	5.355	6.394	-7.439	18.149

Table 11b: 2002 – 2003 Cohen's d Test for Practical Significance

2002 – 2003 – First Matched Pair – Cohen's d		Test for Practical Significance		
Standardized Test	**Mean of Smaller Classes**	**Mean of Larger Classes**	**Pooled Standard Deviation**	**Cohen's d (Effect Size)**
TerraNova1 Reading	79.42	68.71	21.40	0.50
TerraNova1 LangArts	76.90	75.74	20.07	0.06
TerraNova1Math	76.39	67.10	22.06	0.42
TerraNova1Total	79.84	74.19	19.18	0.29
TerraNova2Reading	66.74	65.10	24.58	0.07
TerraNova2 LangArts	80.32	73.13	19.92	0.36
TerraNova2 Math	73.65	69.61	24.34	0.17
TerraNova2 Total	76.65	72.61	21.77	0.19
NJASK3 LangArts	218.48	215.35	15.49	0.20
NJASK3 Math	234.68	229.32	25.11	0.21

Table 12: 2002 – 2003 Independent t-test – Second Matched Pair

[DataSet8] F:\SPSS COHORT INFO\Cohort1MatchedPairs(2).sav

Group Statistics

	ClSzCode	N	Mean	Std. Deviation	Std. Error Mean
TNR1	Small Classes	31	79.42	20.288	3.644
	Not Small Classes	31	75.13	25.778	4.63

TNLA1	Small Classes	31	76.9	21.618	3.883
	Not Small Classes	31	85.06	11.759	2.112
TNM1	Small Classes	31	76.39	21.279	3.822
	Not Small Classes	31	79.61	17.402	3.126
TNTOT1	Small Classes	31	79.84	19.308	3.468
	Not Small Classes	31	82.87	17.469	3.138
TNR2	Small Classes	31	66.74	25.95	4.661
	Not Small Classes	31	75.39	21.404	3.844
TNLA2	Small Classes	31	80.32	18.375	3.3
	Not Small Classes	31	80.65	20.47	3.677
TNM2	Small Classes	31	73.65	23.919	4.296
	Not Small Classes	31	80.94	20.743	3.726
TNTOT2	Small Classes	31	76.65	22.463	4.034
	Not Small Classes	31	82.97	18.082	3.248
NJASK3 LA	Small Classes	31	218.48	14.697	2.64
	Not Small Classes	31	223.42	14.463	2.598
NJASK3 M	Small Classes	31	234.68	26.844	4.821
	Not Small Classes	31	241.39	19.268	3.461

Table 12a: 2002 – 2003 Independent Samples Test Second Matched Pair

	Levene's Test for Equality of Variances		t-test for Equality of Means					
	F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference

									Lower	Upper
TNR1	Equal variances assumed	1.183	0.281	0.728	60	0.469	4.29	5.892	-7.495	16.076
	Equal variances not assumed			0.728	56.859	0.469	4.29	5.892	-7.508	16.089
TNLA1	Equal variances assumed	11.027	0.002	-1.846	60	0.07	-8.161	4.42	-17.003	0.68
	Equal variances not assumed			-1.846	46.322	0.071	-8.161	4.42	-17.057	0.734
TNM1	Equal variances assumed	1.567	0.215	-0.653	60	0.516	-3.226	4.937	-13.101	6.65
	Equal variances not assumed			-0.653	57.727	0.516	-3.226	4.937	-13.109	6.658
TNTOT1	Equal variances assumed	1.401	0.241	-0.648	60	0.519	-3.032	4.677	-12.387	6.322
	Equal variances not assumed			-0.648	59.409	0.519	-3.032	4.677	-12.389	6.324
TNR2	Equal variances assumed	0.345	0.559	-1.431	60	0.158	-8.645	6.042	-20.73	3.44
	Equal variances not assumed			-1.431	57.904	0.158	-8.645	6.042	-20.739	3.449
TNLA2	Equal variances assumed	0.305	0.583	-0.065	60	0.948	-0.323	4.94	-10.205	9.56
	Equal variances not assume			-0.065	59.313	0.948	-0.323	4.94	-10.207	9.562

	d									
TNM2	Equal variances assumed	2.072	0.155	-1.282	60	0.205	-7.29	5.686	-18.665	4.084
	Equal variances not assumed			-1.282	58.822	0.205	-7.29	5.686	-18.669	4.089
TNTOT2	Equal variances assumed	1.637	0.206	-1.221	60	0.227	-6.323	5.179	-16.682	4.037
	Equal variances not assumed			-1.221	57.382	0.227	-6.323	5.179	-16.692	4.047
NJASK3 LA	Equal variances assumed	0.994	0.323	-1.333	60	0.188	-4.935	3.703	-12.343	2.472
	Equal variances not assumed			-1.333	59.985	0.188	-4.935	3.703	-12.343	2.473
NJASK3 M	Equal variances assumed	3.514	0.066	-1.131	60	0.263	-6.71	5.935	-18.581	5.162
	Equal variances not assumed			-1.131	54.427	0.263	-6.71	5.935	-18.606	5.187

Table 12b: 2002 -2003 Cohen's d Test for Practical Significance

2002 - 2003 – Second Matched Pairs – Cohen's d		Test for Practical Significance		
Standardized Test	**Mean of Smaller Classes**	**Mean of Larger Classes**	**Pooled Standard Deviation**	**Cohen's d (Effect Size)**
TerraNova1 Reading	79.42	75.13	23.10	0.19

TerraNova1 LangArts	76.90	85.06	17.74	(0.46)
TerraNova1Math	76.39	79.61	19.35	(0.17)
TerraNova1Total	79.84	82.87	18.32	(0.17)
TerraNova2Reading	66.74	75.39	23.99	(0.36)
TerraNova2 LangArts	80.32	80.65	19.29	(0.02)
TerraNova2 Math	73.65	80.94	22.51	(0.32)
TerraNova2 Total	76.65	82.97	20.47	(0.31)
NJASK3 LangArts	218.48	223.42	14.67	(0.34)
NJASK3 Math	234.68	241.39	23.42	(0.29)

Table 15: 2003 – 2004 Independent T-tests – First Random Sample

[DataSet3] F:\SPSS COHORT INFO\Co2RandomSample.sav

Group Statistics

	ClsSzCode	N	Mean	Std. Deviation	Std. Error Mean
TN1R	Small Classes	29	80	21.414	3.977
	Not Small Classes	29	82.1	22.264	4.134
TN1LA	Small Classes	29	83	18.476	3.431
	Not Small Classes	29	85.07	16.686	3.099
TN1M	Small Classes	29	70.97	25.777	4.787

	Not Small Classes	29	78.72	16.002	2.971
TN1TOTAL	Small Classes	29	80.93	21.426	3.979
	Not Small Classes	29	85.17	17.378	3.227
TN2R	Small Classes	29	69.66	21.036	3.906
	Not Small Classes	29	62.34	22.342	4.149
TN2LA	Small Classes	29	74.45	17.185	3.191
	Not Small Classes	29	72.45	22.516	4.181
TN2M	Small Classes	29	71.48	22.823	4.238
	Not Small Classes	29	61.69	25.363	4.71
TN2TOTAL	Small Classes	29	74.28	16.533	3.07
	Not Small Classes	29	68.24	21.38	3.97
NJASK3LA	Small Classes	29	221.55	14.669	2.724
	Not Small Classes	29	224.41	16.819	3.123
NJASK3MA	Small Classes	29	236.1	22.732	4.221
	Not Small Classes	29	242.28	24.335	4.519

Table 15a: 2003 – 2004 Independent Samples Test – First Random Sample

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TN1R	Equal variances assumed	0.508	0.479	-0.367	56	0.715	-2.103	5.736	-13.595	9.388
	Equal variances not assumed			-0.367	55.915	0.715	-2.103	5.736	-13.595	9.388
TN1LA	Equal variances assumed	0.921	0.341	-0.448	56	0.656	-2.069	4.623	-11.33	7.192

	Equal variances not assumed			-0.448	55.428	0.656	-2.069	4.623	-11.332	7.194
TN1M	Equal variances assumed	5.343	0.025	-1.377	56	0.174	-7.759	5.634	-19.045	3.528
	Equal variances not assumed			-1.377	46.79	0.175	-7.759	5.634	-19.094	3.577
TN1TOTAL	Equal variances assumed	1.414	0.239	-0.828	56	0.411	-4.241	5.123	-14.504	6.021
	Equal variances not assumed			-0.828	53.712	0.411	-4.241	5.123	-14.513	6.031
TN2R	Equal variances assumed	0.127	0.722	1.283	56	0.205	7.31	5.698	-4.105	18.726
	Equal variances not assumed			1.283	55.798	0.205	7.31	5.698	-4.106	18.727
TN2LA	Equal variances assumed	0.717	0.401	0.38	56	0.705	2	5.26	-8.537	12.537
	Equal variances not assumed			0.38	52.357	0.705	2	5.26	-8.553	12.553
TN2M	Equal variances assumed	1.636	0.206	1.546	56	0.128	9.793	6.336	-2.899	22.486
	Equal variances not assumed			1.546	55.388	0.128	9.793	6.336	-2.902	22.489
TN2TOTAL	Equal variances assumed	2.23	0.141	1.202	56	0.234	6.034	5.019	-4.019	16.088
	Equal variances not assumed			1.202	52.667	0.235	6.034	5.019	-4.033	16.102
NJASK3LA	Equal variances assumed	0.064	0.801	-0.691	56	0.493	-2.862	4.144	-11.164	5.44
	Equal variances not assumed			-0.691	54.984	0.493	-2.862	4.144	-11.167	5.443

NJASK3MA	Equal variances assumed	0.152	0.698	-0.998	56	0.322	-6.172	6.184	-18.56	6.215
	Equal variances not assumed			-0.998	55.742	0.323	-6.172	6.184	-18.561	6.217

Table 15b: Cohen's d Test for Practical Significance

2003 – 2004 – First Random Sample – Cohen's d		Test for Practical Significance		
Standardized Test	**Mean of Smaller Classes**	**Mean of Larger Classes**	**Pooled Standard Deviation**	**Cohen's d (Effect Size)**
TerraNova1 Reading	80.00	82.10	21.68	(0.10)
TerraNova1 LangArts	83.00	85.07	17.48	(0.12)
TerraNova1Math	70.97	78.72	21.62	(0.36)
TerraNova1Total	80.93	85.17	19.45	(0.22)
TerraNova2Reading	69.66	62.34	21.82	0.34
TerraNova2 LangArts	74.45	72.45	19.88	0.10
TerraNova2 Math	71.48	61.69	24.42	0.40
TerraNova2 Total	74.28	68.24	19.19	0.31
NJASK3 LangArts	221.55	224.41	15.71	(0.18)
NJASK3 Math	236.10	242.28	23.55	(0.26)

Table 16: 2003 – 2004 - Independent T-test – Second Random Sample

[DataSet5] F:\SPSS COHORT INFO\Cohort2RandomSample(2).sav

Group Statistics

	ClsSzCode	N	Mean	Std. Deviation	Std. Error Mean
TN1R	Small Classes	29	80	21.414	3.977
	Not Small Classes	29	78.72	24.678	4.583
TN1LA	Small Classes	29	83	18.476	3.431
	Not Small Classes	29	80.28	21.091	3.917
TN1M	Small Classes	29	70.97	25.777	4.787
	Not Small Classes	29	81.07	19.241	3.573
TN1TOTAL	Small Classes	29	80.93	21.426	3.979
	Not Small Classes	29	83.28	20.601	3.826
TN2R	Small Classes	29	69.66	21.036	3.906
	Not Small Classes	29	68.48	23.368	4.339
TN2LA	Small Classes	29	74.45	17.185	3.191
	Not Small Classes	29	76.38	24.362	4.524
TN2M	Small Classes	29	71.48	22.823	4.238
	Not Small Classes	29	75.66	17.086	3.173
TN2TOTAL	Small Classes	29	74.28	16.533	3.07
	Not Small Classes	29	76.86	19.572	3.634
NJASK3LA	Small Classes	29	221.55	14.669	2.724
	Not Small Classes	29	222.17	15.175	2.818

NJASK3MA	Small Classes	29	236.1	22.732	4.221
	Not Small Classes	29	247.38	20.507	3.808

Table 16a: 2003 -2004 Independent Samples Test – Second Random Sample

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TN1R	Equal variances assumed	0.068	0.795	0.21	56	0.834	1.276	6.067	-10.878	13.43
	Equal variances not assumed			0.21	54.91	0.834	1.276	6.067	-10.884	13.436
TN1LA	Equal variances assumed	0.73	0.397	0.523	56	0.603	2.724	5.207	-7.706	13.155
	Equal variances not assumed			0.523	55.046	0.603	2.724	5.207	-7.71	13.159
TN1M	Equal variances assumed	1.83	0.182	-1.691	56	0.096	-10.103	5.973	-22.069	1.862
	Equal variances not assumed			-1.691	51.81	0.097	-10.103	5.973	-22.09	1.884
TN1TOTAL	Equal variances assumed	0	0.991	-0.425	56	0.673	-2.345	5.52	-13.402	8.712
	Equal variances not assumed			-0.425	55.914	0.673	-2.345	5.52	-13.402	8.712
TN2R	Equal variances assumed	0.486	0.489	0.201	56	0.842	1.172	5.839	-10.524	12.868
	Equal variances not assumed			0.201	55.393	0.842	1.172	5.839	-10.526	12.871
TN2LA	Equal variances assumed	3.118	0.083	-0.349	56	0.729	-1.931	5.536	-13.022	9.159

	Equal variances not assumed			-0.349	50.335	0.729	-1.931	5.536	-13.049	9.187
TN2M	Equal variances assumed	1.234	0.271	-0.788	56	0.434	-4.172	5.294	-14.778	6.433
	Equal variances not assumed			-0.788	51.884	0.434	-4.172	5.294	-14.797	6.452
TN2TOTAL	Equal variances assumed	1.217	0.275	-0.544	56	0.589	-2.586	4.758	-12.117	6.944
	Equal variances not assumed			-0.544	54.478	0.589	-2.586	4.758	-12.123	6.95
NJASK3LA	Equal variances assumed	0.193	0.662	-0.158	56	0.875	-0.621	3.919	-8.472	7.231
	Equal variances not assumed			-0.158	55.936	0.875	-0.621	3.919	-8.472	7.231
NJASK3MA	Equal variances assumed	0.166	0.685	-1.983	56	0.052	-11.276	5.685	-22.664	0.113
	Equal variances not assumed			-1.983	55.416	0.052	-11.276	5.685	-22.667	0.115

Table 16b: Cohen's d Test for Practical Significance

2003 – 2004 – Second Random Sample – Cohen's d		Test for Practical Significance		
Standardized Test	**Mean of Smaller Classes**	**Mean of Larger Classes**	**Pooled Standard Deviation**	**Cohen's d (Effect Size)**
TerraNova1 Reading	80.00	78.72	22.91	0.06
TerraNova1 LangArts	83.00	80.28	19.70	0.14
TerraNova1Math	70.97	81.07	23.11	(0.44)
TerraNova1Total	80.93	83.28	20.87	(0.11)
TerraNova2Reading	69.66	68.48	22.05	0.05

TerraNova2 LangArts	74.45	76.38	20.92	(0.09)
TerraNova2 Math	71.48	75.66	20.09	(0.21)
TerraNova2 Total	74.28	76.86	18.00	(0.14)
NJASK3 LangArts	221.55	222.17	14.80	(0.04)
NJASK3 Math	236.10	247.38	22.20	(0.51)

Table 17: 2003 -2004 Independent t-test – First Matched Pair

[DataSet1] F:\SPSS COHORT INFO\Cohort2HandPicked.sav

Group Statistics

	ClsSzCode	N	Mean	Std. Deviation	Std. Error Mean
TN1R	Small Classes	29	80	21.414	3.977
	Not Small Classes	29	74.34	28.605	5.312
TN1LA	Small Classes	29	83	18.476	3.431
	Not Small Classes	29	77.31	20.436	3.795
TN1M	Small Classes	29	70.97	25.777	4.787
	Not Small Classes	29	71.31	21.319	3.959
TN1TOTAL	Small Classes	29	80.93	21.426	3.979
	Not Small Classes	29	76.72	23.281	4.323
TN2R	Small Classes	29	69.66	21.036	3.906
	Not Small Classes	29	64.1	27.142	5.04
TN2LA	Small Classes	29	74.45	17.185	3.191

	Not Small Classes	29	67.31	28.676	5.325
TN2M	Small Classes	29	71.48	22.823	4.238
	Not Small Classes	29	64.41	26.508	4.922
TN2TOTAL	Small Classes	29	74.28	16.533	3.07
	Not Small Classes	29	67.14	26.788	4.974
NJASK3LA	Small Classes	29	221.55	14.669	2.724
	Not Small Classes	29	217.55	17.836	3.312
NJASK3MA	Small Classes	29	236.1	22.732	4.221
	Not Small Classes	29	234.59	26.981	5.01

Table 17a: 2003 -2004 Independent Samples Test – First Matched Pair

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TN1R	Equal variances assumed	1.168	0.285	0.852	56	0.398	5.655	6.635	-7.637	18.947
	Equal variances not assumed			0.852	51.883	0.398	5.655	6.635	-7.66	18.971
TN1LA	Equal variances assumed	0.383	0.538	1.112	56	0.271	5.69	5.116	-4.559	15.938
	Equal variances not assumed			1.112	55.44	0.271	5.69	5.116	-4.561	15.94
TN1M	Equal variances assumed	0.754	0.389	-0.056	56	0.956	-0.345	6.212	-12.788	12.099
	Equal variances not assumed			-0.056	54.096	0.956	-0.345	6.212	-12.798	12.108
TN1TOTAL	Equal variances assumed	0.194	0.661	0.716	56	0.477	4.207	5.875	-7.563	15.977

	Equal variances not assumed			0.716	55.618	0.477	4.207	5.875	-7.565	15.978
TN2R	Equal variances assumed	2.229	0.141	0.871	56	0.388	5.552	6.377	-7.222	18.326
	Equal variances not assumed			0.871	52.72	0.388	5.552	6.377	-7.24	18.343
TN2LA	Equal variances assumed	12.696	0.001	1.15	56	0.255	7.138	6.208	-5.298	19.574
	Equal variances not assumed			1.15	45.815	0.256	7.138	6.208	-5.359	19.635
TN2M	Equal variances assumed	1.956	0.167	1.088	56	0.281	7.069	6.496	-5.943	20.081
	Equal variances not assumed			1.088	54.791	0.281	7.069	6.496	-5.95	20.088
TN2TOTAL	Equal variances assumed	12.423	0.001	1.221	56	0.227	7.138	5.846	-4.572	18.848
	Equal variances not assumed			1.221	46.628	0.228	7.138	5.846	-4.624	18.9
NJASK3LA	Equal variances assumed	0.2	0.657	0.933	56	0.355	4	4.288	-4.591	12.591
	Equal variances not assumed			0.933	53.989	0.355	4	4.288	-4.598	12.598
NJASK3MA	Equal variances assumed	1.552	0.218	0.232	56	0.818	1.517	6.551	-11.607	14.641
	Equal variances not assumed			0.232	54.432	0.818	1.517	6.551	-11.615	14.65

Table 17b: Cohen's d Test for Practical Significance

2003 – 2004 – First Matched Pair – Cohen's d		Test for Practical Significance		
Standardized Test	Mean of Smaller Classes	Mean of Larger Classes	Pooled Standard Deviation	Cohen's d (Effect Size)
TerraNova1 Reading	80.00	74.34	25.21	0.22
TerraNova1 LangArts	83.00	77.31	19.52	0.29
TerraNova1Math	70.97	71.31	23.45	(0.01)
TerraNova1Total	80.93	76.72	22.28	0.19
TerraNova2Reading	69.66	64.10	24.23	0.23
TerraNova2 LangArts	74.45	67.31	23.71	0.30
TerraNova2 Math	71.48	64.41	24.77	0.29
TerraNova2 Total	74.28	67.14	22.36	0.32
NJASK3 LangArts	221.55	217.55	16.31	0.25
NJASK3 Math	236.10	234.59	24.74	0.06

Table 18: 2003 -2004 - Independent T-Test – Second Matched Pair Group

[DataSet3] F:\SPSS COHORT INFO\Cohort2MatchedPairs(2).sav

Group Statistics

	ClsSzCode	N	Mean	Std. Deviation	Std. Error Mean
TN1R	Small Classes	29	80	21.414	3.977
	Not Small Classes	29	81.41	17.783	3.302
TN1LA	Small Classes	29	83	18.476	3.431

	Not Small Classes	29	83	15.991	2.969
TN1M	Small Classes	29	70.97	25.777	4.787
	Not Small Classes	29	82.07	16.266	3.02
TN1TOTAL	Small Classes	29	80.93	21.426	3.979
	Not Small Classes	29	85.9	12.712	2.361
TN2R	Small Classes	29	69.66	21.036	3.906
	Not Small Classes	29	69.62	20.67	3.838
TN2LA	Small Classes	29	74.45	17.185	3.191
	Not Small Classes	29	77.1	22.991	4.269
TN2M	Small Classes	29	71.48	22.823	4.238
	Not Small Classes	29	74.48	20.181	3.747
TN2TOTAL	Small Classes	29	74.28	16.533	3.07
	Not Small Classes	29	77.55	18.193	3.378
NJASK3LA	Small Classes	29	221.55	14.669	2.724
	Not Small Classes	29	222.48	12.872	2.39
NJASK3MA	Small Classes	29	236.1	22.732	4.221
	Not Small Classes	29	246.52	19.22	3.569

Table 18a: 2003- 2004 Independent Samples Test – Second Matched Pair Group

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TN1R	Equal variances assumed	1.367	0.247	-0.274	56	0.785	-1.414	5.169	-11.768	8.941

	Equal variances not assumed			-0.274	54.172	0.785	-1.414	5.169	-11.776	8.949
TN1LA	Equal variances assumed	0.879	0.352	0	56	1	0	4.537	-9.09	9.09
	Equal variances not assumed			0	54.871	1	0	4.537	-9.094	9.094
TN1M	Equal variances assumed	5.938	0.018	-1.962	56	0.055	-11.103	5.66	-22.442	0.235
	Equal variances not assumed			-1.962	47.246	0.056	-11.103	5.66	-22.488	0.281
TN1TOTAL	Equal variances assumed	2.806	0.1	-1.073	56	0.288	-4.966	4.626	-14.233	4.302
	Equal variances not assumed			-1.073	45.539	0.289	-4.966	4.626	-14.28	4.349
TN2R	Equal variances assumed	0.008	0.93	0.006	56	0.995	0.034	5.476	-10.936	11.005
	Equal variances not assumed			0.006	55.983	0.995	0.034	5.476	-10.936	11.005
TN2LA	Equal variances assumed	4.546	0.037	-0.498	56	0.62	-2.655	5.33	-13.333	8.023
	Equal variances not assumed			-0.498	51.844	0.62	-2.655	5.33	-13.352	8.041
TN2M	Equal variances assumed	0.202	0.655	-0.53	56	0.598	-3	5.657	-14.333	8.333
	Equal variances not assumed			-0.53	55.173	0.598	-3	5.657	-14.337	8.337
TN2TOTAL	Equal variances assumed	1.9	0.174	-0.718	56	0.476	-3.276	4.565	-12.421	5.869
	Equal variances not assumed			-0.718	55.495	0.476	-3.276	4.565	-12.422	5.871

NJASK3LA	Equal variances assumed	0.486	0.489	-0.257	56	0.798	-0.931	3.624	-8.191	6.329
	Equal variances not assumed			-0.257	55.07	0.798	-0.931	3.624	-8.194	6.331
NJASK3MA	Equal variances assumed	0.444	0.508	-1.884	56	0.065	-10.414	5.528	-21.487	0.66
	Equal variances not assumed			-1.884	54.494	0.065	-10.414	5.528	-21.494	0.666

Table 18b: 2003 – 2004 Cohen's d Test for Practical Significance

2003 – 2004 – Second Matched Pairs Group – Cohen's d		Test for Practical Significance		
Standardized Test	**Mean of Smaller Classes**	**Mean of Larger Classes**	**Pooled Standard Deviation**	**Cohen's d (Effect Size)**
TerraNova1 Reading	80.00	81.41	19.52	(0.07)
TerraNova1 LangArts	83.00	83.00	17.13	0.00
TerraNova1Math	70.97	82.07	22.09	(0.50)
TerraNova1Total	80.93	85.90	17.64	(0.28)
TerraNova2Reading	69.66	69.62	20.67	0.00
TerraNova2 LangArts	74.45	77.10	20.16	(0.13)
TerraNova2 Math	71.48	74.48	21.41	(0.14)
TerraNova2 Total	74.28	77.55	19.31	(0.17)
NJASK3 LangArts	221.55	222.48	13.69	(0.07)
NJASK3 Math	236.10	246.52	21.52	(0.48)