A COMPARISON OF STUDENT LITERACY ACHIEVEMENT BEFORE AND AFTER A CLASS SIZE REDUCTION IMPLEMENTATION IN SELECT PRIMARY CLASSROOMS

MARIA FARRELL

A COMPARISON OF STUDENT LITERACY ACHIEVEMENT BEFORE AND AFTER A CLASS SIZE REDUCTION IMPLEMENTATION IN SELECT PRIMARY CLASSROOMS

A Field Study Proposal

Presented to

The College of Graduate Studies

Austin Peay State University

In Partial Fulfillment

of the Requirements for the Degree
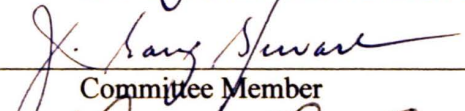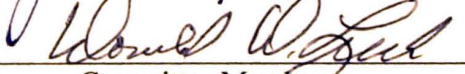
Education Specialist

Prepared by

Maria Farrell

December 2009

December, 2009

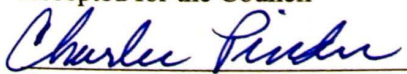To the College of Graduate Studies:

We are submitting a Field Study written by Maria Farrell entitled "A Comparison of Student Literacy Achievement Before and After a Class Size Reduction Implementation in Select Primary Classrooms." We have examined the final copy of this Field Study for form and content. We recommend that it be accepted in partial fulfillment of the requirements for the degree of Educational Specialist.

_____
Research/Committee Advisor/Chair

_____
Committee Member

_____
Committee Member

Accepted for the Council

_____
Dean, College of Graduate Studies

STATEMENT OF PERMISSION TO USE

In presenting this field study in partial fulfillment of the requirements for the Education Specialist degree at Austin Peay State University, I agree that the Library shall make it available to borrowers under rules of the Library. Brief quotations from this field study are allowable without special permission, provided that accurate acknowledgement of the source is made.

Permission for extensive quotation from or reproduction of this field study may be granted by my major professor, or in her absence, by the Head of Interlibrary Services when in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this field study for financial gain shall not be allowed without my written permission.

Signature _Maria Farrell_

Date _November 12, 2009_

DEDICATION

This field study is dedicated to my husband Brian Farrell and my family who have supported, loved and believed in me. Without their encouragement this would not have been possible. Thank you for your understanding in my long journey to further my education and learning.

ACKNOWLEDGEMENTS

ABSTRACT

This study analyzed DIBELS^{TM} Word Use Fluency (WUF) scores and DIBELS^{TM} Oral Reading Fluency (ORF) scores from 749 primary students in kindergarten, first, and second grade in two selected Title I schools in Clarksville Montgomery County Schools. Participants involved in the study were assigned to a regular size classroom with average class size ranging from 18-23 students for the 2006-2007 school year. Following a classroom size reduction (CSR) program, the participants were assigned to smaller classrooms with an average class size ranging from 16-18 students for the 2007-2008 school year.

STATView Statistical Software was used to conduct un-paired t-tests for each of the participating grade levels and DIBELS^{TM} subtests (WUF and ORF), to compare scores prior and post to implementation of CSR. Analysis of Variance, (ANOVA) was utilized using STATView to test for statistical significance with regard to gender and ethnicity. Three null hypotheses were tested and analyzed at the .05 level of significance.

Results of the study indicated that there was a statistically significant difference in literacy achievement on DIBELS^{TM} WUF and ORF scores between students assigned to regular sized classroom and students assigned to smaller classes. Further, statistical significance was shown based on ethnicity (majority and minority groups) and gender.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

Introduction

Statement of the Problem

Creating smaller class sizes within our schools is a common strategy used to help raise student achievement. The bar is set high by the No Child Left Behind (NCLB) Act and schools are striving to help all students achieve. In the past decade, class size reduction (CSR) has been a popular school reform strategy implemented to improve achievement and narrow the achievement gap (Konstantopoulos, 2008; Krueger, 2003; Nye, Hedges, & Konstantopoulos, 2004). The American Educational Research Association (2003) listed lowering class size as a top priority for education and stated, "While small classes benefit all kinds of students, much research has shown that the benefits may be greatest for minority students or students attending inner-city schools. For these students, smaller class sizes can shrink the achievement gap" (2003, p. 2).

The National Education Association (2006) has lobbied for small class sizes of no more than fifteen students and is reflected in the following statement:

> We know that this school reform strategy can play a key role in closing
> achievement gaps, by giving students more individual attention- teachers with
> small classes can spend more time and energy helping each child succeed. Small
> classes are a common sense strategy and essential to NEA's vision of a great
> public school for every child. (p.1).

CSR implementations across the nation have become controversial as questions are raised about the cost-effectiveness of the practice and interpretations of related

research (Hanushek, 1999; Johnson, 2002; Krueger, 2003; Reichardt, 2001). Critics argue that class size reductions in a time of teacher shortages can lower teacher quality and effectiveness, countering any benefit of smaller classes (Hanushek, 1999; Johnson, 2002; Krueger, 2003). They contend that CSR forces administrators to dip low into the applicant pool and hiring any warm bodies to fill the new positions (Hanushek, 1999; Johnson, 2002; Krueger, 2003; Reichardt, 2001). Title I schools with large numbers of disadvantaged students often have a great need for high quality, experienced teachers. It is these types of schools where CSR programs are often implemented as a strategy to shrink the achievement gap (Hanushek, 1999; Johnson, 2002; Krueger, 2003; Reichardt, 2001).

There is a general agreement among many researchers that smaller classes in primary grades increases student achievement on average for all students (Achilles, 2003; Biddle & Berliner, 2002; Smith, Molnar, & Zahorik, 2003). Smaller classes can give teachers more opportunity to meet the individual needs of students. With fewer students, teachers are left with more time to spend on direct instruction and less time on student discipline (Achilles, 2003; Biddle & Berliner, 2002; Finn & Achilles, 1990; Smith, Molnar, & Zahorik, 2003).

Several state-wide longitudinal CSR studies have helped compile data supporting the many benefits of smaller class sizes. Indiana's Project Prime Time (Gilman & Antes, 1985), Tennessee's Student-Teacher Achievement Ratio (STAR) Project (Word et al., 1990), Wisconsin's Student Achievement Guarantee in Education (SAGE) Program (Molnar et al., 1999), and California's Class Size Reduction Program (Bohrnstedt &

Stecher, 2002) are some of the more notable studies that have been reviewed and analyzed by many educational researchers.

Some research surrounding CSR programs suggest that not only does smaller class size increase achievement, but smaller classes have an even more significant impact on Black students and students who are economically disadvantaged (Molnar et al., 1999; Smith, Molnar, & Zahorik, 2003). Researchers reviewing the STAR data and Wisconsin's SAGE data found evidence that Black students and disadvantaged students benefit even more from smaller classes than White students. They found that Black students have greater gains when compared to White students in smaller classes (Biddle & Berliner, 2002; Finn, Gerber, & Boyd-Zaharias, 2005; Molnar, et al., 1999; Smith, Molnar, & Zahorik, 2003).

The findings that Black students and disadvantaged students benefit more from smaller classes than White students have been challenged by some researchers (Nye, Hedges, & Konstanopoulous, 2002; Konstanopoulous, 2008). Nye, Hedges & Konstanopoulous (2002, 2004) investigated the differential effects of small classes on low achieving, minority, and disadvantaged students and found weak evidence that smaller classes help these types of students. A more recent study released by Konstantopoulous (2008) found that higher achievers benefit more from smaller classes than lower achievers. Konstantopoulous further suggests that CSR will not reduce the achievement gap that exists between two groups of students in a classroom.

Purpose of the Study

The purpose of this study was to compare student achievement in several primary classrooms before and after a CSR implementation. This study measured achievement by

utilizing data from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) assessment. DIBELS™ is administered three times per year in all kindergarten, first, and second grade students in the participating school district. This study assessed any significant difference in DIBELS™ scores between ethnicity and gender.

Significance of the Study

With No Child Left Behind legislative requirements getting more rigorous, school districts are looking for reform strategies to help schools meet adequate yearly progress, increase academic achievement for all students, and close the achievement gap. In the 2005-2006 school year, Clarksville Montgomery County Schools (CMCSS) spent over 1.2 million dollars on CSR with the goal of improving equity and reducing the achievement gap that exists in the district. CMCSS hired 20 new teachers and 19 teacher aids for schools with at-risk students (Clarksville Montgomery County Schools, 2006, p. 21). As part of the strategic plan to improve student achievement the county has continued to allot the same 1.2 million dollars for school years 2007-2008 and 2008-2009 in order to pay the salaries of the newly created positions (Clarksville Montgomery County Schools, 2006, 2008). The 2008 Clarksville Montgomery County School budget states, "Elementary schools which were identified as high needs schools are targeted to receive additional teachers to lower class size, a strategy that has a proven record of improving student achievement" (p. 15). This field study was conducted to determine if the CSR implementation within two Title I schools impacted literacy achievement. By understanding the impact of class size on student achievement, the school system can make better policy decisions in the future and meet the needs of all learners.

Research Questions

The following research questions were generated to guide the study.

1. Is there a significant difference in DIBELS scores before and after lowering class size?

2. Is there a significant difference in DIBELS scores before and after lowering class size based on gender?

3. Is there a significant difference in DIBELS scores before and after lowering class size based on ethnicity?

Hypotheses

1. There is no statistically significant difference in DIBELS scores before and after lowering class size.

2. There is no statistically significant difference in DIBELS scores before and after lowering class size based on gender.

3. There is no statistically significant difference in DIBELS scores before and after lowering class size based on ethnicity.

Limitations

1. This study was conducted in two elementary schools in Clarksville, Tennessee. The small sample size used in this study limits the potential of the results of the study to be generalized to other populations.

2. Student achievement was based on and limited to two types of DIBELS scores only. A full range of student achievement is difficult to assess in primary grades as the district does not use standardized assessments in math, science, or social studies until students reach the third grade.

Assumptions

    1. All teachers involved in the study were highly qualified.

    2. All DIBELS™ administrators were highly qualified and follow the set assessment procedures.

    3. DIBELS™ scores were reported accurately using the Edusoft software.

Definition of Terms

    1. Class Size: The number of students for whom a teacher is primarily responsible during a school year.

    2. Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A set of standardized individually administered assessments of early literacy development. The subtest assessments that were used in this study were Word Use Fluency (WUF) and Oral Reading Fluency (ORF).

    3. Oral Reading Fluency (ORF): an individually administered standardized test of accuracy and fluency in oral reading.

    4. Word Use Fluency (WUF): an individually administered standardized test of vocabulary and oral language.

    5. Ethnicity: Ethnicity is noted by parents upon a student's enrollment to CMCSS schools. Ethnic groups are described as African-American, American Native/Alaskan, Asian/Pacific Islander, Caucasian, and Hispanic as determined by CMCSS enrollment data.

    6. Minority: Any student belonging to the ethnic groups of African-American, Hispanic, American Native/Alaskan or Asian/Pacific Islander.

    7. Non-Minority: Any student belonging to the ethnic group Caucasian.

8. Highly Qualified: To be deemed highly qualified by state and district standards, teachers must have a bachelor's degree, full state certification or licensure, and prove that they know each subject they teach.

9. Economically Disadvantaged: Any student that qualifies for free or reduced lunch based on family income.

CHAPTER II

Review of Related Literature

Researchers have reached a consensus in recent years that lowering class size has positive benefits and positive effects on student achievement (Biddle & Berliner, 2002; Bohrnstedt & Stecher, 1999; Nye, Hedges, & Konstantopoulous 2004; Smith, Molnar, & Zahorik; 2003; U.S. Dept. of Education, 2004). Class size reduction (CSR) policies have been a popular reform strategy used across the nation (Achilles, 1996, 2003; Biddle & Berliner, 2002; Bohrnstedt & Stecher, 1999; U.S. Dept. of Education, 2004). CSR has been controversial because it is typically one of the most costly educational interventions when compared to other educational reforms (Kruegar, 2003; Odden, 1990; Reichardt, 2001; Tomlinson 1988;).

The class size debate in America is not new. The debate is more than a century old. Small class size experimental studies have been documented in America since the 1900's (Biddle & Berliner, 2002; Glass & Smith, 1979).

The Glass & Smith Meta-Analysis

The Glass and Smith study of 1979 compiling years of class size research was titled, *Meta-Analysis of Research on Class Size and Achievement*. Glass and Smith's findings and questionable methodology heated the class size debate, and sparked even more interest nation-wide (Achilles, Finn & Pate-Bain, 2002; Berger, 1982; Biddle & Berliner, 2002; Education Research Service, 1980). Glass and Smith analyzed 77 studies from twelve different countries spanning over 70 years from 1900 to 1979. They broke down the studies into 725 separate class size comparisons and quantified each

comparison by using codes for the characteristics. For example, Glass and Smith reported on the year, publication type, age of participants, duration of instruction, number of instructional groups, demographics, pupil/instructor ratio, and other properties of the 77 studies. The Glass and Smith (1979) study quantified the outcomes of 725 comparisons through meta-analysis and reported using regression analysis relating to achievement on class size.

Glass and Smith (1979) found that there was no correlation between class size and achievement in studies conducted before 1940, but that there was a strong correlation between class size and achievement in studies conducted after 1960. They found a strong relationship between class size and student achievement, especially at class sizes with 15 students or less. The Glass and Smith study documented well-controlled and poorly controlled studies as a variable with their class size comparisons. They documented that over 100 comparisons within their meta-analysis came from controlled experiments with randomly assigned pupils, but that over 300 comparisons were from the uncontrolled category. This proved to be the main weakness in the Glass, Smith meta-analysis that drew much criticism on their findings (Berger, 1982; Educational Research Service, 1980; Odden, 1990; Tomlinson, 1988). The most important findings on class size were based on only 14 of the 77 studies, one of which was based on tennis instruction (Glass & Smith, 1979; Oden, 1990). Some of the class size comparisons in the uncontrolled category were from poorly designed studies that involved small samples, short-term exposure to small classes, and questionable measures of success (Berger, 1982; Educational Research Service, 1980; Oden, 1999; Tomlinson, 1988). Some researchers contend that the meta-analysis review by Glass and Smith (1979), set the stage for future

studies and debates on class size and achievement (Achilles, Finn, & Pate-Bain, 2002; Berger, 1982; Biddle & Berliner, 2002; Educational Research Service, 1980; Hedges & Stock, 1983). For example, Glass and Smith (1979) found that the optimal class size for student achievement should be around 15 or less. This continues to be the standard goal for many class size reduction policies seen today (Biddle & Berliner, 2002; NEA, 2006).

State legislatures and policy makers in the nineteen eighties allocated millions of dollars into CSR implementations (Gilman & Antes, 1985; Wisconsin Department of Public Instruction, 2008; Word, et al., 1990). Indiana's project Prime Time and Tennessee's STAR project, are two of the larger state-wide CSR implementations that began in the eighties (Biddle & Berliner, 2002). Some researchers began publishing studies and articles challenging the cost effectiveness of CSR and arguing that the so-called benefits of smaller classes do not outweigh the high costs to taxpayers (Grissmer, 1999; Hedges & Stock, 1983; Iacovou, 2002; Johnson, 2002; Kruegar 2003; Oden, 1990; Tomlinson, 1988).

Tomlinson (1988) reviewed a history of class size studies, their effects on student achievement, and focused on a cost-benefits analysis. Tomlinson concluded, "By itself, reducing class size, a very costly reform is unlikely to have tangible benefits for student achievement, at least not in a form or terms that are comprehensible and acceptable to the taxpaying public" (p. 2). He made a recommendation to the Department of Education and all policy makers:

Accordingly, rather than sink vast sums into an inefficient and unreliable method of school improvement, available resources should be instead be directed on

improving the quality of instruction and teachers' ability to manage the demands of classrooms as they are currently configured. (p. 2)

Tomlinson (1988) figured that considering the average class size in America was about 24 students, in order to reap the benefits of the statistically defensible "optimal" class size of 15 students as concluded by Glass and Smith (1979), an almost 40 percent class size reduction would be required to gain about a 10 percent improvement in learning.

Indiana's Project Prime Time

Indiana's Project Prime Time was an initiative financed by the Indiana Department of Education that aimed to reduce class size beginning with all first grade classes in 1984 to 18 students per class and gradually extended to all K-3 classes in the state by 1987 (Biddle & Berliner, 2002; Indiana State Department of Public Instruction, 1983; Sanogo & Gilman, 1994). Project Prime Time eventually reduced classrooms across the state from an average of 25 to an average of 18 students. The funding program was in response to the favorable results of the initial Project Prime Time pilot study that began in the 1981-1982 school year with a sample of 24 public schools and lasted two years. The pilot study analysis found strong evidence that student who attended small classes had substantially larger gains for reading and mathematics achievement (Indiana State Department of Public Instruction, 1983).

Several evaluations of Indiana's Prime Time Project found mixed results (Gilman & Antes, 1985; Malloy & Gilman, 1988; McGiverin, Gilman, & Tillitski, 1989). Gilman and Antes (1985) were critical of the Prime Time pilot study results. The evidence from the 1981 pilot study used to support the new statewide phase-in expansion of the project

in 1984 was based on results from classrooms with an average of 14 students. The funding initiative that followed only reduced class sizes to an average of 18 (Tomlinson, 1988, Sango & Gilman, 1994). Gilman and Antes (1985) also contended that the teachers in the pilot study were carefully chosen. The pilot study was considered a highly publicized program that may have caused a reactive Hawthorne Effect where teachers in smaller classrooms had more incentive to improve student achievement (Gilman & Antes 1985). Gilman and Antes (1985) also questioned the political undertones and reluctance of the Department of Education to allow independent evaluations of the study (Gilman & Antes, 1985; Tomlinson, 1988).

Gilman and Antes (1985) argued the validity of pilot study stating:

It should be noted that while many variables were measured in the pilot study, only those with significant results in favor of Prime Time were reported. The Indiana Department of Education as been reluctant to have Prime Time results evaluated by a statewide study. (p. 6)

Gilman and Antes (1985) conducted their own evaluation of the effectiveness of Prime Time by comparing achievement scores of first graders during the years of 1983-1984 when class sizes were 15-35 and 1984-1985 when class sizes were reduced to 15-22 due to the Prime Time initiative. They found that 40 of the statistical tests favored the smaller Prime Time classes from 1984-1985 and only four statistical tests favored larger classes. Gilman and Antes (1985) found, "It can be concluded that Indiana's Prime Time Program is largely successful during it's initial year. These results are overwhelmingly in favor of the smaller Prime Time classes" (p. 9). Malloy and Gilman (1988) studied over

113,000 third graders who had completed three years of smaller Prime Time classes and compared them to third graders who had no experience with Prime Time classes.

Malloy and Gilman (1988) found that there was a slight difference favoring the smaller Prime Time classes with third graders. Malloy and Gilman (1988) still concluded that the cost of the Prime Time program was not justified because the difference was marginal stating, "This difference has no practical utility since any difference would have been significant where there were over 133,000 subjects" (p. 3.). McGiverin, Gilman and Tillitski (1989) studied second graders who had completed two years in smaller Prime Time classes. They found that Prime Time students had higher achievement in basic skills when compared to cohorts in larger classes suggesting that smaller classes can benefit primary students.

The Tennessee STAR Project

The Tennessee STAR (Student/Teacher Achievement Ratio) Project was a longitudinal class size study that began in the mid-1980s and continued through1998 (Word et al., 1990). Many consider the STAR Project a landmark study in class size research (Biddle & Berliner, 2002; Achilles, 2003; Bohrnstedt & Stecher, 2002; U.S. Dept. of Education, 2004). Biddle and Berliner (2002) describe the STAR Project as "arguably the largest, best-designed field experiment that has ever appeared for education" (p. 6). This class size study was the first of its magnitude to implement a well-controlled experimental design that randomly assigned students to classrooms (Biddle & Berliner, 2002; Achilles, 2003; Bohrnstedt & Stecher, 2002; U.S. Dept. of Education, 2004). The successful findings of the Tennessee STAR Project jump-started policy changes and statewide CSR studies across the nation (Biddle & Berliner, 2002; Achilles,

2003; Bohrnstedt & Stecher, 2002; U.S. Dept. of Education, 2004). CSR programs have since appeared at the federal level and in various states in response to the STAR Project and other extensive studies conducted in the past two decades (Achilles, 2003; Biddle & Berliner, 2002; Bohrnstedt & Stecher, 2002; U.S. Dept. of Education, 2004).

Funded by the Comprehensive Education Reform Act of 1984 passed by the Tennessee Legislature, Tennessee State University began the comprehensive four-year class size experiment known as the STAR project. The STAR class size reduction project was carried out in three different studies analyzing the data. The original experiment in K-3, STAR (1985-1989) assessed the benefits of smaller classes compared to larger or standard classes. The Lasting Benefits Study (1989) tracked student achievement in STAR students after they returned to larger standard classrooms in fourth, fifth, sixth grades, and beyond. Project Challenge (1989) implemented smaller classrooms for grades K-3 in low-income schools with disadvantaged students who were at high risk of dropping out of school (Achilles, 2003).

Every school district in the state was invited to participate in the STAR program, provided they agreed to remain in the program for four years. Schools participating received only a monetary grant to hire additional teachers and aides. No money was allocated to fund additional space, resources, materials, or training. Due to the constraints of space and no additional support, many troubled schools could not participate. The sample for the first year of the study involved 79 schools, 328 classrooms and about 6,300 students. This sample came from a widespread population within the state and represented rural, urban, inner city, and suburban districts (Word et al., 1990).

Beginning in 1985, kindergarten students in 79 Tennessee schools were randomly assigned to one of three experimental classrooms: a standard full-sized class with one certified teacher and 22-26 students; a supplemented full-sized class with 22-26 students, one certified teacher and one non-certificated full-time teacher's aide; and a small classroom with one certificated teacher and 13-17 students (Word et al., 1990).

The project began the first year of the study with a cohort of kindergartners and the design of the study involved gathering achievement data on each subject as they progressed through third grade. Class sizes were maintained throughout the school day and throughout the year. In 1985 the STAR cohort participants were in kindergarten and assigned to the same types of classrooms every year for first, second and third grades. Student achievement was measured using the Stanford Achievement Test battery including reading, word study skills and math tests in the spring of each school year (Word et al., 1990).

With only 79 schools initially participating, the researchers were challenged to control sample populations from year to year due to shifting. Variations in samples were caused for several reasons. Kindergarten was not mandated in Tennessee at the time, families moved and relocated, and some students were retained to repeat a grade level. Some students were also permitted to switch from experimental to control classrooms and vice-versa. These types of shifts were due to parental requests, some involving conflicts between students who needed to be assigned to separate classrooms (Word et al., 1990).

Families moving from one STAR school to another STAR school filled vacant seats in corresponding experimental classrooms as much as possible. Some seats in the second, third, and fourth years of the study, however, were filled with new students. At

the end of the four years in 1989, the number of participants had almost doubled to about 12,000 students. Some students were assigned to small classrooms for four consecutive years, some were in small classrooms for three consecutive years, some two consecutive years, and some just one year (Achilles, 2003; Biddle & Berliner, 2002; Word et al., 1990). This migration of the sample population caused some critical reviewers to conclude that the STAR's results were biased and that the project did not constitute a controlled experiment (Hanushek, 1999).

The initial STAR findings showed that smaller classes had advantages when compared to standard classes. In terms of reading achievement when compared to students in standard classes, students who were in STAR classes for four years were .5 months ahead in kindergarten, 1.9 months ahead by the end of first grade, 5.6 months ahead in second grade and 7.1 months ahead by the end of third grade (Word et al., 1999). Students exposed to three, two, or just one year of smaller classes still showed advantages in achievement, although smaller than students who had continuous exposure for four years (Word et al., 1999; Biddle & Berliner, 2002).

The Lasting Benefits Study conducted by a team of researchers lead by Charles Achilles (Achilles, Nye, Zaharias, & Fulton; 1993) found that the benefits of smaller classes in primary grades continued even two years after students were returned to standard classes with up to 25 students. The Lasting Benefits Study results showed that in the fourth and fifth grades, students who had been in smaller classes in the early grades scored higher on the Comprehensive Tests of Basic Skills when compared to students who were in regular classes with or without a teacher aid in the early grades (Achilles et al., 1993). Eighth grade students who had attended one or more years of small classes in

K-3 were 4.1 months ahead in reading, 3.4 months ahead in math, 4.3 months ahead in science, and 4.8 months ahead in social studies when compared to students who did not attend smaller STAR classes in the early grades (Achilles et al., 1993; Biddle & Berliner, 2002).

In the Project Challenge Study (1989-1995) seventeen county school districts were selected to participate in a class size reduction project geared at improving academic achievement for at risk students in Tennessee. The poorest county districts were chosen based on the criteria that 75% or more students enrolled qualified for the free and reduced lunch program. Federal Chapter I funds helped to supplement the cost of reducing classrooms with a ratio of 1:15. The Project Challenge Study used the Tennessee Comprehensive Assessment Program (TCAP) to measure and evaluate the academic success of students. Project Challenge was different from STAR, as it was not an experiment. Project Challenge had no random selection or placement and no special testing. Staff at Tennessee State University worked with the Tennessee Department of Education to track and evaluate academic project of students in Challenge schools. Reports were provided each year using TCAP results by various researchers. Participating Challenge schools were reduced to sixteen after the third year of implementation (Achilles, Zaharias, Nye, & Fulton, 1995; Mosteller, 1995; Nye, Achilles, Boyd-Zaharias, Fulton & Wallenhorst, 1992).

Nye et al. and her colleagues,(1992) ranked schools from 138 Tennessee school systems based on their second grade TCAP scaled scores. Preliminary results of the first year, 1990-1991, showed that of the 17 Project Challenge systems, nine improved their rankings in reading ten improved in mathematics when compared to the 1989-1990

TCAP scores. Nye et al. (1992) and her colleagues also converted 138 Tennessee school systems average TCAP scores to z scores to determine Project Challenge school systems deviation from the state average scores in reading and math. While Project Challenge average z scores for reading and math were below the state average, there was a quarter standard deviation gain between 1990 and 1991, moving all 17 Project Challenge systems closer to the State TCAP mean test scores. Nye and her colleagues (1992) stated: "The gains in ranking and z-score comparisons suggest that on average, the second-grade results on TCAP indicated achievement benefits that might be attributed to participation in smaller classes" (p. 12).

On average, Project Challenge school systems continued to improve in rank over the next two years. Mostellar (1995) in his review of all the STAR related studies noted that the school districts in Project Challenge were performing well below the state average in mathematics before the project was implemented and those districts participating in Project Challenge moved above the state average after three years of implementation.

Achilles et al. (1995) found that Project Challenge schools collectively were below average in TCAP scores for 1990 and by 1993 were collectively above average in reading (Achilles, et al., 1995). Achilles, et al. (1995) also used the Tennessee Value-Added Assessment System (TVAAS) to help aggregate data for students taking the TCAP assessment. TVAAS is a statistical process where Achilles et al. (1995) could compare gains in Tennessee students to a large number of variables and rank average scores by system, school, or classes and the equivalent mean gain compared to the national norm gain. Achilles et al. found that Project Challenge schools produced

achievement results that paralleled those predicted from earlier STAR experiments showing support for positive effects of smaller class sizes. It was noted however, that a 1:15 ratio throughout Project Challenge schools was an assumed condition and not verified.

Hanushek's (1999) critical analysis, review, and independent investigation of the Tennessee STAR experiment challenged its validity and reliability due to design and implementation issues. In his independent study of the STAR data, Hanushek used econometric evidence to calculate the estimated effects of teacher-pupil ratios on student outcomes. Hanushek contended that the commonly cited STAR results were upwardly biased and that the benefits of class size reduction implementations are not worth the cost. Hanushek concluded in his 1999 analysis:

> Class size reduction represents one of the most costly reform policies actively discussed. Even if there are positive effects, they must be sufficiently large to justify the expenditure. Because of the very small (if any) effects of general class size reduction policies that have been found a thorough analysis of costs and effectiveness relative to other policies does not seem to be required here.
>
> (p. 159)

The Wisconsin SAGE Program

Under the supervision of STAR investigators from the Tennessee program, Wisconsin began targeting neighborhood schools with large enrollments of minorities and disadvantaged students for class size reduction (U.S. Dept. of Education, 2004). The five-year pilot study of the Wisconsin Student Achievement Guarantee in Education (SAGE) began in the 1996-97 school year (Molnar et al., 1999). In the 1996-97 school

year, 30 elementary schools representing 21 school districts became SAGE schools. The program required four interventions for participating schools. The first requirement was that participating schools reduce the pupil-teacher ratio within a classroom to at least 15 students per teacher beginning with kindergarten and first grade in the 1996-97 school year, second grade added by the 1997-98 school year, and third grade in the 1998-99 school year. The second requirement was to establish lighted school houses open in early morning until late at night offering more support for disadvantaged students. The third requirement was to develop rigorous curricula and the fourth requirement was for schools to establish a system of staff development and professional accountability (Molnar et al., 1999).

Annual evaluations of the effects of the program were mandated by SAGE legislation. Immediate implementation of all four required interventions were not possible by all participating schools, but the pupil-teacher ratio was met by all participating schools (Molnar et al., 1999). Most schools reduced class sizes down to about 15 students per one teacher in a single classroom. The SAGE legislation defined pupil-teacher ratio as the number of assigned students to one teacher. They did not specify that students had to be in one room with four walls and a teacher. Thus, some SAGE schools creatively found solutions to conquer the space problem of adding additional classrooms (Molnar et al., 1999). Out of the 176 SAGE classrooms, 10 in the 1996-97 school year and 11 in the 1997-98 school year consisted of a shared space classroom where a temporary wall separated classrooms with 15 students on one side with a teacher and 15 on the other with one teacher. Forty-two SAGE classrooms in the 1996-97 school year and 45 in the 1997-98 school year consisted of 30 students with two

teachers in one single classroom. The other SAGE classroom type was a floating teacher classroom and consisted of 30 students with one teacher except for during reading, language arts, and math when an additional teacher joined the class to reduce the ratio to 15:1. Ten classrooms in the 1996-97 school year implemented the floating teacher model, while only four classrooms in 1997-98 used the floating teacher model. One participating SAGE kindergarten classroom in 1996-97 and one first grade the 1997-98 school years consisted of 45 students in one single classroom space with three teachers (Molnar et al., 1999).

The comprehensive Test of Basic Skills (CTBS) was administered to first grade SAGE students and comparison schools (non-SAGE) and included pre and post assessments in reading, language arts, and mathematics (Molnar et al., 1999). Comparison schools were among the same district and resembled each SAGE school in terms of average family income, prior records of achievement in reading, racial composition, and K-3 enrollment (Molnar et al., 1999). A team of researchers headed by Alex Molnar, Philip Smith and John Zahorik were the lead researchers from the University of Wisconsin that analyzed the data and evaluated the effects of the first two years of implementation. Molnar and his colleagues used regression analysis to control for differences in individual student characteristics, and a hierarchical linear modeling to control for classroom characteristics when analyzing student achievement outcomes. Principals and teachers were surveyed, some were interviewed, and on-site observations were completed by researchers (Molnar et al.,1999).

The results from evaluations of first grade CTBS scores for the first two years of the SAGE study yielded similar findings as the Tennessee STAR program. The 1997-98

SAGE first grade students scored significantly higher on posttests in all subjects tested when compared to non-SAGE students from the comparison schools (Molnar et al., 1999; Smith et al., 2003). Molnar and his colleagues did report a ceiling effect in the results of the 1996-97 test score results showing a downward bias (Molnar et al., 1999). When analyzing the 15:1 classrooms and the 30:2 pupil-teacher ratio effects, first grade achievement results did not show a statistically significant difference on any subtests or total scores, suggesting that the benefits of smaller classes can be achieved without adding additional classroom space. The 45:3 pupil-teacher ratio classroom was not included in the analysis (Molnar et al., 1999).

The SAGE study offered a valid argument that Black students benefit more from smaller classes than White students (Molnar et al., 1999; Smith et al., 2003). Black SAGE students achieved greater gains in total CTBS scores than White SAGE students (Molnar, et al., 1999). For example, White SAGE student made average pre-posttest gains of 44.93 points (1996-97) and 48.56 points (1997-98), while Black average gains were 53.79 points (1996-97) and 53.93 points (1997-98) (Molnar et al., 1999).

The SAGE program is often cited as evidence that CSR can succeed as a vehicle to reduce the achievement gap that exists between Black and White students (Biddle & Berliner, 2002; Molnar, et. al., 1999; Nye, Hedges, & Konstantopoulos, 1999, 2000; Smith et al., 2003; U.S. Dept. of Education, 2004). Molnar and his colleagues (1999) noted that the gap between Black and White students was larger in SAGE schools than it was in comparison schools. In terms of the achievement gap, Molnar et al., (1999) stated "African American SAGE students achieved greater gains in terms of CTBS total scale scores than White SAGE students from pretest to posttest, closing the achievement gap

(although the gap remained statistically significant)" (p.170). The researchers in the SAGE study found that:

A strong case can be made for reducing class size in the early grades as an effective means of addressing the achievement gap between African American and White students, one of the most persistent and troubling problems in American public education (Molnar et al., 1999 p.177).

SAGE study surveys revealed that smaller classes helped teachers as they reported more knowledge about students, more instructional time due to fewer discipline problems, more time for individualized instruction, and the ability to cover more content. Smaller SAGE classes were linked to increased job satisfaction for teachers, reduced stress, and more time to collaborate with other teachers (Molnar et al., 1999).

The SAGE program has continued steadily since it was first implemented in 1996. Smith et al., (2003) tracked the academic performance of students in 30 schools from 21 school districts that began the SAGE program in 1996. They compared achievement of SAGE students to similar students in non-SAGE schools in the same districts. They tracked students for three full school years using the Terra Nova Comprehensive Test of Basic Skills with pre and posttests each fall and spring from 1996 to 2001. Smith et al., (2003) found that the greatest separation between achievement of SAGE and non SAGE students occurred in first grade with a 25-30 percent higher level of achievement than comparison students in larger classes. They also found that the advantages of smaller classes continued as SAGE students were achieving one-third to one-half a year ahead of comparison students in larger classes. Their analysis of the longitudinal SAGE data also found continued positive benefits of narrowing the achievement gap. Smith et al., stated:

The SAGE program narrows the achievement gap between African American and white students in 1st grade and prevents it from widening in 2nd and 3rd grade. In the larger comparison classrooms, the achievement gap between African American and white students widens each year (2003, p. 73.)

The Wisconsin Department of Public Instruction continues it's SAGE program today in it's schools across the state. The Wisconsin Department of Public Instruction website reports that Wisconsin now has approximately 475 SAGE schools and that SAGE schools receive $2, 250 for each low-income K-3 child (Wisconsin Department of Public Instruction, 2008).

While the Indiana, Tennessee, and Wisconsin programs helped to show the positive benefits of smaller classes in earlier grades, the state sponsored class size reduction program in California did not have the same successes (Biddle & Berliner, 2002; Achilles, 2003; Bohrnstedt & Stecher, 1999; Smith et al., 2003; U.S. Dept. of Education, 2004). Biddle and Berliner (2002) describe California's CSR program as " a near –textbook case of how not to reduce class size within a specific state" (p.13).

The California CSR Program

The release of the 1994 NAEP scores put California tied for last place in reading among fourth graders for the 39 states that participated in the national assessment (Bohrnstedt & Stecher, 2002). In 1994 California had the largest average class size in the country averaging 29 students per classroom. With published Tennessee's STAR Project successes, Governor Pete Wilson helped the California legislature passed a new law providing school districts $650 per student for each K-3 classroom that had 20 or fewer students. The new legislative funding also required that schools first reduce all first grade

classes, followed by second grades and then kindergarten or third grades (Bohrnstedt & Stecher, 2002). California spent almost 1 billion dollars in the first year of implementation of CSR (Bohrnstedt & Stecher, 2002). Evaluators of the California CSR program note that California's population is much more racially and ethnically diverse when compared to Tennessee. The number of students with limited English proficiency in California far outnumbers those in Tennessee as well (Biddle & Beliner, 2003; Bohrnstedt & Stecher, 2002). The first year of CSR in California involved almost 1 million students in 52,000 classrooms, over 150 times as many students as in the Tennessee STAR program (Bohrnstedt & Stecher, 2002).

The CSR Research Consortium, a partnership researching California's Class size Reduction Reform lead by Bohrnstedt & Stecher (2002) released a comprehensive evaluation and summary of the state-wide program. Class size reductions in California were associated with declines in teacher qualifications and inequitable distribution of credentialed teachers (Bohrnstedt & Stecher, 2002). The ultimate size of the CSR required the immediate hiring of thousands of new teachers and the addition of 18,000 classrooms. Schools serving inner city students and disadvantaged students found themselves in a jam as they were forced to choose from the bottom of the applicant pool to fill vacant positions. More qualified applicants took advantage of the many vacancies available in the suburbs due to the new legislation (Biddle & Berliner, 2002; Bohrnstedt & Stecher, 2002). Borhnstedt and Stecher (2002) painted a picture of the magnitude of the problem in their report:

> The proportion of K–3 teachers who were not fully credentialed (e.g., teachers with intern or emergency credentials) increased from 1.8 percent before the

program started to 12.5 percent in the second year of the program. Most of the uncredentialed teachers were hired by schools serving the most disadvantaged

students, in part because these schools were slower to implement CSR, and more certificated teachers had already been hired elsewhere. In 2000–01, more than one in five K–3 teachers were not fully credentialed in schools with high percentages of low-income, EL, minority, or Hispanic students (primarily large and urban (p. 10).

Urban schools also found it much more difficult to aquire the needed space to support the new number of additional classrooms (Bohrnstedt & Stecher, 2002). Operating costs for the new classrooms exceeded state payments and many schools took classroom space and funds from other programs to support CSR. Schools reported turning special education quarters, music rooms, art rooms, computer rooms, even libraries and teachers' lounges into makeshift classrooms to meet state requirements. Some schools had to tap into operating budgets to pay the rent on portables that were purchased to house new classrooms and delayed building repairs or purchasing needed curriculum materials (Bohrnstedt & Stecher, 2002).

While surveys showed that parents and students in California liked smaller classes and teachers reported giving students more individualized attention, evaluations of the CSR program showed no relationship of smaller class sizes to student achievement (Bohrnstedt & Stecher, 2002). Borhrnstedt and Stecher tracked achievement gains between cohorts of students with incrementally different patterns of exposure to CSR from kindergarten to third grade. While statewide average test scores increased on

average across cohorts, the gains in achievement could not be linked to CSR (Bohrnstedt & Stecher, 2002). Biddle and Berliner (2002) stated their view on the on the California

CSR program:

> This example should serve to remind us that smaller classes are not an educational panacea- that in order to be effective, programs for reducing class size should be planned with care and with thought given to the other needs and strengths of existing school systems. (p.13)

The Federal Class Size Reduction Program

In 1999 under the Clinton Administration, the Department of Education appropriated $1.2 billion for the Federal Class-Size Reduction Program (U.S. Department of Education, 2004). The Federal CSR program, although altered under the Bush Administration, still exists today and has since been folded into Title II, Part A of the No Child Left Behind (NCLB) Act of 2001 (U.S. Department of Education, 2004). The goal of the Federal CSR program is to improve student achievement, particularly in reading, by reducing class size in primary grades. Funds are distributed by the states to the most needy schools based on enrollment and poverty data. Funds allocated are allowed to be spent on recruitment, teacher salaries, and professional development and training of new teachers. In Fiscal Year 2001 the U.S. Department of Education allocated over $1.6 billion dollars to schools to be used under the guidelines of the Federal CSR Program (U.S. Department of Education, 2004).

An evaluation by the U.S. Department of Education (2004) of the Federal CSR Program was conducted by using 500 district staff surveys and 500 principal surveys.

Site visits to 24 schools and 48 classrooms were conducted in the spring of 2001. The evaluation was conducted to provide data on the distribution of federal funds for CSR, to provide data on the implementation of CSR, and effects of the program on class size. It did not provide data on the effects of CSR on classroom practices or student achievement (U.S. Department of Education, 2004).

The surveys of the Federal CSR program revealed facts about how Title II funds were spent (U.S. Department of Education, 2004). It was reported that while districts were permitted to use up to 15 percent of allocated funds for professional development, only 39 percent of districts in 2001 chose to spend it on that activity. Instead, districts used the funds to hire teachers to reduce class size (U.S. Department of Education, 2004). The Federal CSR program has had a modest impact on reducing class size in targeted schools (U.S. Department of Education, 2004). According to U.S. Department of Education surveys from 2001, in schools and grades where federally funded CSR teachers were placed, the average class size decreased by only one or two students. The overall average class size, after the Federal CSR implementation ranged from 18 in kindergarten, 20 in first grade and 21 in grades 2 and 3. Evaluators believed this to be due to increases in enrollment that mitigated the CSR and also due to the fact that 40 percent of the schools assigned CSR teachers to a special subject or team teaching instead of creating a new classroom (U.S. Department of Education, 2004).

Evaluators reported on two of the limitations of the Federal CSR program: available teachers and classroom space. Department of Education evaluators noted, "Variations in program implementation across districts demonstrates the limited ability of

small amounts of federal funds to create more effective learning environments for students throughout the nation" (U.S. Department of Education, 2004, p. 49).

The Long Reach of Tennessee's Project STAR

While Project STAR was conducted over two decades ago in the 1980's, due to the success of the study, it's data are still being reviewed and analyzed by researchers to help find more answers about CSR and its effects (Finn et al., 2005; Konstantopoulous, 2008; Krueger & Whitmore, 2001). Researchers are still studying the STAR data to find out just how far the effects of smaller classes can be related in life. Finn, Gerber & Boyd-Zaharias (2005) used STAR data to analyze graduation rates and concluded that graduating was related to K-3 achievement. Finn, Gerber, and Boyd-Zaharias (2005) found that attending small classes for 3 or more years increased the likelihood of graduating from high school, especially among disadvantaged students. Alan Krueger and Diane Whitmore at Princeton University reviewed Project STAR data in hopes of answering the questions about the relationship between smaller class sizes and reducing achievement gaps. Krueger and Whitmore (2001) found that low income and Black students tend to benefit more from attending smaller classes that White students. Krueger and Whitmore (2001) also found evidence that criminal conviction rates were 20 percent lower for adult Black males who were assigned to a small class in early years than for adult Black males who were assigned to a regular-class in early years. Krueger and Whitmore's (2001) finding for incarceration rates was noted, but was not statistically significant.

Nye, Hedges & Konstantopoulous' Class Size Studies

Nye, Hedges, and Konstantopoulous (1999, 2000a, 2002, 2004) used Tennessee Project STAR data and collaborated to further study the issue and effects of class size. Nye et al. (1999) conducted a five-year follow up of the students in the Tennessee STAR experiment and analyzed the long term effects of smaller classes in primary grades. They found results very similar to the Lasting Benefits Study and concluded that small class effects in primary grades do not disappear over time, but provide achievement benefits that last until high school. Nye et al. (1999) also found that students who experience more years of smaller classes in Kindergarten through grade 3 have higher levels of achievement five years later than students who have had fewer years of smaller classes (Nye et al., 1999). In 2000 (a), Nye et al. conducted a study using STAR data to determine if disadvantaged students benefit more from smaller classes. They found that the results of the study support the conclusion that small classes in the early grades lead to higher academic achievement. Nye et al. (2000a) also found what they called as "weak evidence" supporting that minorities benefit more from smaller classes (p. 1). Their results indicated that minority students experience small class effects on reading achievement that are 54-86 percent larger than the effects on White students and 30-46 percent as large as the reading achievement gap between minority and White students. Differential effects of small classes on minority students on math achievement were smaller and not statistically significant. None of the differential effect of small classes for low SES students were statistically significant. Minority students were only defined in the study as a student belonging to a minority group and were not given a specific group such as Black or Hispanic (Nye et al., 2000a).

In 2004, Nye, Hedges and Konstanopoulous conducted a similar study to their 1999 study using Project STAR five-year follow up data but differentiated to determine if minorities experience larger lasting benefits from smaller classes when compared to White students. Nye et al. (2004) used a repeated measures analysis and found that there were positive lasting benefits for minorities who attend four years of smaller classes in primary grades. In the case of reading achievement, minorities experienced a much larger small class effect when compared to White students in all grades (Nye et al., 2004). In the case of math achievement, the small class effect for minorities was only slightly larger in grades four and seven when compared to White students, and was about 30% larger in grades six and eight. In grade five, the data suggested that the small class effect in math achievement for White students was somewhat larger than minorities (Nye et al., 2004). They also found evidence that the negative differential effect for girls lead to a smaller gender difference in math scores among students who had four years of smaller classes when compared to those in regular-sized classes. The analysis of the data lead Nye et al. (2004) to conclude: "Thus, evidence indicated that small classes could help reduce overall racial and ethnic inequality in reading achievement and reduce gender inequality in mathematics achievement" (p. 99).

A study published in March 2008 conducted by Spyros Konstantopoulos at Northwestern University used Project STAR research data to investigate whether small class sizes had an effect on reducing the achievement gap among low and high achievers. Konstantopoulos reviewed data from Project STAR and found evidence to suggest that while decreasing class size increases achievement on average for all students, it does not appear to reduce the achievement gap within a class.

Konstantopoulous (2008) used meta-analysis and quantile regression methods to investigate the effects of small classes in primary grades on the achievement gap in mathematics and reading SAT scores. Previous studies in class size focused on estimating mean differences in student achievement between small and regular sized classes and found that class size reduction had larger positive effects for minority and disadvantaged students (Finn & Achilles, 1990; Nye et al., 2004; Kruegar & Whitmore, 2001). Konstantopoulous (2008) hoped to extend his investigation with a more in depth assessment and strived to "shed light on the mechanism through which small classes affect the achievement of low-and high-achieving students" (p. 274). Konstantopoulous (2008) examined the variability of student achievement between small and regular classes as well as the varied effects of small classes in the upper and lower tails of the achievement distribution. He does note the limitation of using SAT scores as they are norm referenced, and are not necessarily aligned with the school curriculum taught at each grade level. In his analysis, Konstantopoulous found that all students made significant gains in smaller classes when compared to regular sized classes. When examining variability he found evidence that smaller classes may not reduce the achievement gap (Konstantopoulous, 2008). When comparing very high achievers from the upper tail of the achievement distribution to the very low achievers at the lower tail of the achievement distribution, the very high achievers made significantly more gains from being in small classes then did very low achievers. Konstantopoulous (2008) concludes that while all types of students benefit from being in smaller classrooms, CSR will most likely not work as a strategy in closing the achievement gaps that exist within classrooms.

Conclusion

Student achievement for all is American education's top priority and the nation has a vested interest in closing achievement gaps that exist between groups of students. When taxpayer money is involved, there will always be a debate about what is the best use of funds in terms of education (Achilles, 2003; Achilles et al., 2002; American Educational Research Association, Fall 2003; Biddle, & Berliner, 2002; Borland et al., 2005; Finn et al., 2005; Gilman, & Kiger, 2003; Hattie, 2005; Iacovou, 2002; Konstantopoulos, 2008; Krueger, 2003; Nye et al., 2004). There is much research to support the claim that smaller classes in primary grades help students to achieve more academically and that the effects are lasting (Achilles, 2003; Achilles et al., 1998; Biddle & Berliner, 2002; Finn & Achilles, 1990; Glass & Smith, 1979; Grissmer, 1999; Nye et al., 2000b, 2004; Smith, et al., 2003). Whether different types of students benefit more from smaller classes than others is unclear, even after decades of research and debates. There is some evidence to support the claim that Black and disadvantaged students benefit more from smaller classes and that smaller classes in primary grades can contribute to closing the achievement gap(Achilles et al., 1998; Molnar et al., 1999; Nye et al., 2004; Smith, et al., 2008). Konstantopoulous' (2008) more recent study challenges those claims and helps to put the issue of CSR back in the limelight.

CHAPTER III

METHODOLOGY

Overview

This chapter explains the methods and procedures that were employed in the research process. Study design, participant information, data collection process, data analysis plan, and specific hypotheses will be included.

Research Design

The research design used was a quantitative descriptive study. This field study examined any significant relationships between smaller class sizes and increased academic achievement. The relationship between DIBELS™ scores prior and post to CSR implementation were examined for significance based on the variables of gender and ethnicity. The statistical design applied in this study was the *t* test and an *F* Test (ANOVA). The statistical package used was StatView.

Participants

This study accessed DIBELS scores from 749 primary students in kindergarten, first, and second grade in two selected Title I schools in Clarksville Montgomery County Schools. Participants involved in the study were assigned to a regular size classroom with average class size ranging from 18-23 students for the 2006-2007 school year. Post CSR implementation, the participants were assigned to smaller classrooms with an average class size ranging from 16-18 students for the 2007-2008 school year.

Instrument

This study measured achievement by utilizing data from the Dynamic Indicators of Basic Early Literacy Skills™ (DIBELS) assessment. DIBELS™ are a set of standardized individually administered measures of early literacy development. The DIBELS™ assessment tools were developed at the University of Oregon by Roland H. Good and Ruth A. Kaminski (2002). They are designed to be short (one minute) fluency measures, or subtests, used to regularly monitor the development of pre-reading and early reading skills. Clarksville Montgomery County administers DIBELS™ assessments three times a year: the beginning, middle and end of the school year, although subtests included vary depending on grade level and school.

This study compared end of year DIBELS™ scores from 2007 and 2008 using two separate measures or subtests of literacy development: Word Use Fluency (WUF) and Oral Reading Fluency (ORF). Word Use Fluency assessments were administered to all kindergarten, first, and second grade participants. Oral Reading Fluency assessments were administered to first and second grade participants. The ORF assessment is not recommended for kindergarten and is not administered to first graders until the middle of the year. Benchmarks and national norms indicate that most students do not begin reading fluently until the middle of first grade (Good & Kaminski, 2002).

The DIBELS™ Oral Reading Fluency (ORF) measure has been shown to be a good indicator of a student's overall literacy skills. (Fuchs, Fuchs, Hosp, & Jenkins, 2001). For this subtest, students are presented with a grade level reading passage and asked to read it aloud for one minute. Any words read aloud incorrectly, omitted, or hesitated upon for three seconds are considered incorrect. The ORF unit of measure

includes all words read correctly in one minute. The DIBELS™ Administration and Scoring Guide 6th Edition offers evidence of adequate criterion-related validity, concurrent validity, predictive validity, and alternate form reliability for the ORF subtest of the DIBELS™ battery (Good & Kaminski, 2002). Good and Kaminski (2002) correlate DIBELS™ subtests with other measures including the Woodcock-Johnson Psycho-Educational Battery-Revised Reading Cluster, the Woodcock-Johnson Psycho-Educational Battery-Revised Readiness Cluster, and Curriculum Based Measurement (CBM). The ORF measure had the strongest reliability and validity with an alternate form reliability score of .92, a test-retest score of .97, an average concurrent validity coefficient of .80, and a predictive validity coefficient of .66 (Shanahan, 2005).

The DIBELS™ Word Use Fluency (WUF) subtests measures a student's expressive vocabulary skills. Expressive vocabulary is the ability to use words to convey a specific meaning for a word. The WUF measures a students skills in correctly using vocabulary words in verbal utterances. (Good & Kaminski, 2002). For this subtest, students are orally provided with a vocabulary word and asked to use it in a sentence, phrase, expression, or utterance. The administrator continues to present words until the one-minute time limit is reached. Students hesitating or pausing for 5 seconds are provided with the next vocabulary word. The WUF measure is the total number of words spoken by the student in correct verbal utterances. Correct usage is defined as an utterance that conveys the correct meaning of the target vocabulary word or a definition or synonym of the target vocabulary word. An utterance may be a phrase, expression, or sentence (Good & Kaminski, 2002). The DIBELS™ WUF measure had an alternate form single probe reliability score of .71 in kindergarten, .65 in first grade, and .66 in second

grade (Kaminski et al., 2004). The DIBELS™ WUF measure had much more success however with multi-probe reliability using the aggregate of five probes. The DIBELS™ WUF measure had an alternate form multi-probe reliability score of .92 in kindergarten, .90 in first grade, and .91 in second grade (Kaminski et al., 2004). The WUF also had a concurrent criterion related validity of .44 to .55 with the Test of Language Development, .44 to .47 with a comprehensive language sample, .34 to .42 with the ORF, and .35 with the Oregon State Assessment at the end of third grade (Kaminski et al., 2004).

Data Collection Procedure

Prior to collection of data, permission was sought to conduct research from the Institutional Review Board at Austin Peay State University (see Appendix A) and the Clarksville Montgomery County Research Committee (See Appendix B). Data was gathered on WUF and ORF DIBELS scores for kindergarten, first, and second grade students in the two participating Title I schools. Demographic data and DIBELS scores were retrieved and coded by authorized school personnel so that no student or teacher identifiers were revealed. Demographic data were aggregated by grade level, gender, and ethnicity.

Data Analysis Plan

STATView Statistical Software was used to conduct un-paired *t* tests for each of the participating grade levels and DIBELS™ subtests (WUF and ORF) to compare scores prior and post to implementation of CSR. Analysis of Variance (ANOVA) was utilized using STATView to test for statistical significance with regard to gender and ethnicity. Using both the *t* test and the ANOVA, a detailed comparison and analysis was made to determine if there was a significant difference in the DIBELS™ subset scores before and

after the CSR was implemented in the two schools. Hypotheses were tested and analyzed

at the .05 level of significance.

CHAPTER IV
Results and Analysis of Data

Introduction

      This study examined class size reduction (CSR), a strategy implemented for two Title I schools in Clarksville Montgomery County beginning in the school year 2006. This study was undertaken to determine if there was a significant relationship between class size and literacy achievement. This study used DIBELS™ scores, namely WUF and ORF as measures of literacy achievement in 749 students during the school years 2006-2007 and 2007-2008. Both schools include a partial transient population due to the proximity to a military installation. The student body demographics of both schools are similar as they are both inner city Title I schools. Approximately 86% of the population in both schools is economically disadvantaged. Student demographics for the two schools include 52% White, 38% Black, 8% Hispanic, and 2% other. Table 1 details the average class sizes for participants during the study.

Table 1

Class Size Data

Grade Level	Average Class Size 2006-2007	Average Class Size 2007-2008
Kindergarten	18	16
First Grade	20	18
Second Grade	23	19

Using STATView, a statistical software program, the end of year DIBELS™ scores of students prior to the implementation of CSR in 2006-2007 were compared to the end of year DIBELS™ scores of students in 2007-2008 after four teacher positions were added and classes were reduced in size. Two subtests of DIBELS™ scores, Word Use Fluency (WUF) and Oral Reading Fluency (ORF) were utilized and sorted by grade level. Kindergarten does not administer ORF, thus, only WUF was used when comparing kindergarten end of year scores. Independent factors of gender and ethnicity were also researched. Three different hypotheses were analyzed using descriptive statistics. Hypotheses with two variables were analyzed using the unpaired *t* test. Hypotheses with more than two variables were analyzed using an Analysis of Variance (ANOVA). The relationship of each analysis was compared at the .05 level of significance. If statistical differences were found using the ANOVA, a post-hoc Tukey-Kramer test was used to differentiate which groups were statistically different.

Presentation and Analysis of Data

Research Question One

Is there a significant difference in DIBELS™ scores before and after lowering class size? This question was answered by the use of descriptive statistics. Five separate *t* tests were utilized to compare each grade level DIBELS™ subset scores prior and post implementation of class size reduction. Table 2 outlines the comparison of Word Use Fluency (WUF) scores and the Oral Reading Fluency (ORF) scores between the Non Class Size Reduction (NCSR) Year 2006-2007 and the Class Size Reduction (CSR) Year 2007-2008 for the respective grade levels in the two participating Title I schools.

Table 2   41

Unpaired t-Tests Evaluating End of Year DIBELS™ scores (WUF/ORF) 2006-2007 and 2007-2008 in Schools with Class Size Reduction

Grade Level /Variable	n	M	df	t	p
WUF					
K NCSR 2006-07	186	30.170			
K CSR 2007-08	197	35.910	382	-5.221	.0001**
1st NCSR 2006-07	101	61.270			
1st CSR 2007-08	104	73.180	204	-5.723	.0001**
2nd NCSR 2006-07	90	70.850			
2nd CSR 2007-08	71	78.960	160	-6.761	.0001**
ORF					
1st NCSR 2006-07	101	33.730			
1st CSR 2007-08	104	43.160	204	-5.872	.0001**
2nd NCSR 2006-07	90	79.660			
2nd CSR 2007-08	71	90.910	160	-15.288	.0001**

Note: **$p < .05$; NCSR = Non Class Size Reduction; CSR= Class Size Reduction;

Hypothesis One

The null hypothesis stated there was no statistically significant difference in DIBELS™ scores before and after lowering class size. The p-value of .0001 for all five t tests indicates there is a statistically significant difference in DIBELS™ scores. The null hypothesis is therefore rejected, and for the purpose of this study, it can be stated that

student literacy achievement in Word Use Fluency and Oral Reading Fluency significantly increased after the implementation of CSR.

Research Question Two

Is there a significant difference in DIBELS^{TM} scores before and after lowering class size based on gender? The second research question examined the variables of class size and gender for WUF and ORF. An ANOVA was utilized for each grade level to compare the differences in mean DIBELS^{TM} scores (WUF and ORF) of males and females before and after implementation of CSR. A post-hoc Tukey-Kramer test was utilized to differentiate which means were statistically significant and is indicated with subscripts next to each mean.

Table 3

ANOVAs evaluating End of Year WUF DIBELS^{TM} scores 2006-2007and 2007-2008 in schools with Class Size Reduction in terms of gender

Grade Level /Variables	n	M	df	F	p
WUF					
K Males NCSR 2006-07	89	31.123$_a$			
K Females NCSR 2006-07	97	29.217$_b$	1	89.599	.0001**
K Males CSR 2007-08	96	36.910$_c$			
K Females CSR 2007-08	101	34.910$_d$			

Note: **p < .05; NCSR = Non Class Size Reduction; CSR= Class Size Reduction; Means with different subscripts differ significantly at the p < .05 Tukey-Kramer significant difference comparison.

Table 3 continued 43

ANOVAs Evaluating End of Year WUF DIBELS^{TM} Scores 2006-2007and 2007-2008 in Schools with Class Size Reduction in Terms of Gender

Grade Level /Variables	n	M	df	F	p
WUF					
1st Males NCSR 2006-07	48	61.240$_a$			
1st Females NCSR 2006-07	53	61.300$_a$			
1st Males CSR 2007-08	49	72.680$_b$	1	9.275	.0030**
1st Females CSR 2007-08	55	73.680$_c$			
WUF					
2nd Males NCSR 2006-07	44	70.840$_a$			
2nd Females NCSR 2006-07	46	70.860$_a$			
2nd Males CSR 2007-08	35	77.960$_b$	1	7.694	.0067**
2nd Females CSR 2007-08	36	79.960$_c$			

Note: **$p < .05$; NCSR = Non Class Size Reduction; CSR= Class Size Reduction; Means with different subscripts differ significantly at the $p < .05$ Tukey-Kramer significant difference comparison.

Table 4

ANOVAs Evaluating End of Year ORF DIBELS^{TM} Scores 2006-2007 and 2007-2008 in Schools with Class Size Reduction in terms of Gender

Grade Level /Variables	n	M	df	F	p
ORF					
1st Males NCSR 2006-07	48	70.650$_a$			
1st Females NCSR 2006-07	53	76.150$_b$			
1st Males CSR 2007-08	49	77.100$_c$	1	7.966	.0058**
1st Females CSR 2007-08	55	83.650$_c$			
ORF					
2nd Males NCSR 2006-07	44	33.180$_a$			
2nd Females NCSR 2006-07	46	34.280$_b$			
2nd Males CSR 2007-08	35	42.800$_c$	1	39.026	.0001**
2nd Females CSR 2007-08	36	43.520$_d$			

Note: **$p < .05$; NCSR = Non Class Size Reduction; CSR= Class Size Reduction; Means with different subscripts differ significantly at the $p < .05$ Tukey-Kramer significant difference comparison.

Hypothesis Two

The null hypothesis stated there was no statistically significant difference in DIBELS^{TM} scores before and after lowering class size based on gender. Based on the statistical analysis using the ANOVAs which compared end of year DIBELS^{TM} scores for males and females post and prior CSR, the null hypothesis must be rejected. Each p-value for the five ANOVAs indicates a statistical difference with relation to gender at the .05 level of significance. The Tukey-Kramer post-hoc tests indicate a statistically

significant difference between the means for kindergarten males and females in WUF at the .05 level. In first grade and second grade however, the Tukey-Kramer post-hoc tests indicate that there was no statistically significant difference between the means of males and females in WUF in the Non-CSR year, 2006-07. Tukey-Kramer post-hoc tests also revealed that the means did not differ significantly between first grade males and first grade females in ORF for the CSR year, 2007-08.

Research Question Three

Was there a significant difference in DIBELS™ scores before and after lowering class size based on ethnicity? The third research question examined the variables of class size and ethnicity for WUF and ORF. An ANOVA was utilized for each grade level to compare the differences in mean DIBELS™ scores (WUF and ORF) of students belonging to the majority (Caucasian) and students belonging to the minority (African American, Hispanic, Asia/Pacific Islander, Other) before and after implementation of CSR. A post-hoc Tukey-Kramer test was utilized to differentiate which means were statistically significant and is indicated with subscripts next to each mean.

Table 5

ANOVAs Evaluating End of Year WUF DIBELS™ Scores 2006-2007 and 2007-2008 in Schools with Class Size Reduction in Terms of Ethnicity

Grade Level /Variables	n	M	df	F	p
WUF					
K Majority NCSR 2006-07	96	31.123_a			
K Minority NCSR 2006-07	90	29.217_b			
K Majority CSR 2007-08	102	36.910_c	1	89.599	.0001**
K Minority CSR 2007-08	95	34.910_d			
WUF					
1st Majority NCSR 2006-07	54	62.078_a			
1st Minority NCSR 2006-07	47	60.078_b	1	4.479	.0369*
1st Majority CSR 2007-08	56	73.204_c			
1st Minority CSR 2007-08	48	73.157_c			
WUF					
2nd Majority NCSR 2006-07	47	71.647_a			
2nd Minority NCSR 2006-07	43	70.020_b			
2nd Majority CSR 2007-08	37	79.588_c	1	13.336	.0004**
2nd Minority CSR 2007-08	34	78.306_d			

Note. NCSR = Non Class Size Reduction; CSR= Class Size Reduction
**$p < .05$. Means with different subscripts differ significantly at the $p < .05$ Tukey-Kramer significant difference comparison.

Table 6

ANOVAs evaluating End of Year ORF DIBELS[TM] scores 2006-2007 and 2007-2008 in schools with Class Size Reduction in terms of ethnicity

Grade Level /Variables	n	M	df	F	p
ORF					
1st Majority NCSR 2006-07	54	34.471_a			
1st Minority NCSR 2006-07	47	32.471_b			
1st Majority CSR 2007-08	56	43.961_c	1	3.004	.0863
1st Minority CSR 2007-08	48	42.327_c			
ORF					
2nd Majority NCSR 2006-07	47	79.745_a			
2nd Minority NCSR 2006-07	43	79.745_a			
2nd Majority CSR 2007-08	37	90.910_b	1	75.284	.0001**
2nd Minority CSR 2007-08	34	90.592_b			

Note. NCSR = Non Class Size Reduction; CSR= Class Size Reduction
**$p < .05$. Means with different subscripts differ significantly at the $p < .05$ Tukey-Kramer significant difference comparison.

Hypothesis Three

The null hypothesis stated there was no statistically significant difference in DIBELS[TM] scores before and after lowering class size based on ethnicity. Four of the five ANOVAs utilized to test this hypothesis indicated a statistical difference at the .05 level with relation to ethnicity: kindergarten WUF, first grade WUF, second grade WUF and second grade ORF. Based on the statistical analysis using four ANOVAs comparing end of year DIBELS[TM] scores for majority and minority students post and prior CSR, the null

hypothesis must be rejected. However, the null hypothesis must be retained when considering the statistical analysis using the ANOVA for first grade ORF. When analyzing ANOVAs comparing end of year DIBELS™ scores for majority and minority students post and prior CSR it was found that although there was a difference in mean scores and students performed better in smaller classes for the CSR year, it was not statistically different at the .05 level. The difference revealed a p value of .0863.

The Tukey-Kramer post-hoc tests indicate that all of the means differ for kindergarten minority and majority students in WUF at the .05 level of significance. In first grade however, the Tukey-Kramer tests indicate that the means did not differ significantly between majority and minority students in WUF in the CSR year, 2007-08. In terms of ORF, only one Tukey-Kramer post-hoc test indicated a significant difference between the means in respect to majority and minority first grade students in the Non-CSR year, 2006-07.

CHAPTER V

Summary, Findings, Conclusions, Recommendations

Summary

The purpose of this study was to investigate the relationship between class size and literacy achievement in Clarksville Montgomery County Schools by comparing student achievement in several primary classrooms before and after a Class Size Reduction (CSR) implementation. This study measured achievement by utilizing end of year data from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS™) assessment for the school years 2006-07 and 2007-08. This study also assessed any significant difference in DIBELS scores between ethnicity and gender.

The relationship between class size and student achievement was the focus of many earlier studies and state-wide longitudinal studies mandated by state legislators in several states. Educational researchers examining studies such as the Tennessee STAR program and the Wisconsin SAGE program have suggested that not only does smaller class size increase achievement, but smaller classes have an even more significant impact on Black students and students who are economically disadvantaged (Molnar et al., 1999; Smith et al., 2003). These findings have been challenged recently. Nye, Hedges & Konstanopoulous (2002, 2004) found weak evidence that smaller classes help minority and economically disadvantaged students. Konstantopoulous (2008) found that higher achievers benefit more from smaller classes than lower achievers.

Clarksville Montgomery County Schools (CMCSS) has spent millions of dollars in past years on CSR within select Title I schools with the goal of improving equity and reducing the achievement gap that exists in the district (Clarksville Montgomery County

Schools, 2006, 2008). The goal of this field study was to determine if the CSR implementation within two Title I schools for the 2007-08 school year impacted literacy achievement. Understanding the impact of class size on student achievement can assist the district in making better policy decisions in the future as the district strives to meet the challenges of the requirements of NCLB.

The sample for this study was 749 primary students in kindergarten, first, and second grade in two selected Title I schools in Clarksville Montgomery County Schools. Participants involved in the study were assigned to a regular size classroom with average class size ranging from 18-23 students for the 2006-2007 school year. Post CSR implementation, the participants were assigned to smaller classrooms with an average class size ranging from 16-18 students for the 2007-2008 school year. Literacy achievement and class size were examined using the following variables: gender and ethnicity (majority or minority status). The study was conducted to test three hypotheses stated in the null. Unpaired t-tests and Analysis of Variance, (ANOVA) were utilized using STATView statistical software to test for statistical significance at the .05 level. Tukey-Kramer honestly significant difference post hoc analyses were conducted for any tests indicating statistically significant differences to determine which means differed.

Findings

The major purpose of this study was to determine if class size had a significant effect on literacy achievement as measured by DIBELS™ WUF and ORF.

Hypothesis one compared the end of year DIBELS™ WUF and ORF scores of students before and after the CSR implementation. This hypothesis was tested for all students in the study sample. Kindergarten is not administered ORF, thus only end of year WUF scores were used to compare kindergarten literacy achievement before and after the CSR implementation. First and second grade end of year scores were compared in both WUF and ORF to measure literacy achievement before and after the CSR implementation. All unpaired *t* Tests evaluating end of year DIBELS™ scores (WUF/ORF) 2006-2007 and 2007-2008 in schools with class size reduction were found to be statistically significant at the .001 level. The rejection of the null hypothesis indicated the relationship between literacy achievement as measured by DIBELS™ WUF and ORF was statistically significant when considering kindergarten, first and second grade students. Students enrolled in smaller classes in 2008-09 performed better on DIBELS assessments than students enrolled in larger classes the previous year.

Hypothesis two compared the end of year DIBELS™ WUF and ORF scores of males and females before and after the CSR implementation. This hypothesis was tested for all students in the study sample. Kindergarten is not administered ORF, thus only end of year WUF scores were used to compare kindergarten literacy achievement before and after the CSR implementation based on gender. First and second grade end of year scores were compared in both WUF and ORF to measure literacy achievement before and after the CSR implementation based on gender. All ANOVAs evaluating end of year DIBELS™ scores (WUF/ORF) 2006-2007 and 2007-2008 in schools with class size

reduction based on gender were found to be statistically significant at the .05 level. Therefore, the null hypothesis was rejected. Males and Females performed better in smaller class sizes in 2007-2008.

Further analysis of the relationship between class size and literacy achievement using Tukey-Kramer post-hoc tests indicate that while some statistical differences were found between the means of males and females, neither sex benefited more from smaller classes than the other. For example, when looking at ORF, female first graders in 2006-07 scored an average of 34.280 words per minute. Female first graders in 2007-08 performed better with smaller classes averaging 43.520 words per minute, a difference of 9.24 words. When looking at first grade males in ORF between the two years, males in 2007-08 performed better in smaller classes with 9.62 more words than the males in 2006-07. While the male gain is slightly higher, it is certainly comparable and almost equal to the first grade female performance gains in ORF for the same year. When comparing DIBELS^{TM} scores between 2006-07 and 2007-08 first grade females had slightly higher gains in WUF than males with a 12.38 point increase in average score. Males in first grade improved in WUF with smaller classes in 2007-08 by 11.44 points. Again, while the first grade female group appears to have made slightly greater gains in smaller classes with WUF when compared to the first grade males, it is very close in number and differs by less than 1 point. Differences between the means, while statistically significant as indicated by the Tukey-Kramer post-hoc tests, do not indicate that one gender group benefits more from smaller class sizes than the other. Rather, both genders benefit equally from smaller class sizes.

Hypothesis three compared the end of year DIBELS™ WUF and ORF scores of students before and after the CSR implementation based on ethnicity (majority/minority status). This hypothesis was tested for all students in the study sample. Kindergarten is not administered ORF, thus only end of year WUF scores were used to compare kindergarten literacy achievement before and after the CSR implementation. First and second grade end of year scores were compared in both WUF and ORF to measure literacy achievement before and after the CSR implementation. ANOVAs evaluating end of year DIBELS™ scores (WUF/ORF) 2006-2007 and 2007-2008 in schools with class size reduction in terms of ethnicity were found to be statistically significant at the .05 level for all tests except first grade DIBELS™ ORF. Thus, the null hypothesis was rejected when considering ethnicity with kindergarten WUF, first grade WUF, second grade WUF and second grade ORF. When using these literacy measurements, it appears that smaller class sizes can make a difference and benefit both majority and minority students. The null hypothesis must be retained however when considering ethnicity and first grade DIBELS ™ ORF. The ANOVA statistical analysis for first grade DIBELS™ ORF based on ethnicity indicated a difference that was not statistically significant and only had a p-value of .0863.

Further analysis of the relationship between class size and literacy achievement using Tukey –Kramer post-hoc tests indicate that while some statistical differences were found between the means of majority and minority students, neither group benefited more from smaller classes than the other. When analyzing points gained in mean averages in the CSR year for the majority and minority groups, both groups were similar in improvement when compared to the students in the previous year. For example, in

second grade ORF, majority students performed better in the CSR year with 11.165 more points than majority students in 2006-07. The second grade minority group showed similar gains with 10.847 points more than minority students in ORF in 2006-07. When looking at kindergarten WUF gains, the majority group gained 5.787 points in the CSR year and the minority group gained 5.693 points in the CSR year. The data suggest that smaller classes can benefit both majority and minority students about the same.

Conclusions

The purpose of this study was to determine if there was a statistically significant relationship between smaller classes and literacy achievement as measured by DIBELS™ Word Use Fluency Measures (WUF) and Oral Reading Fluency Measures (ORF). The study examined the DIBELS™ scores of 749 primary students in kindergarten, first, and second grade in two selected Title I schools in Clarksville Montgomery County Schools. Participants involved in the study were assigned to a regular size classroom with average class size ranging from 18-23 students for the 2006-2007 school year. Post CSR implementation, the participants were assigned to smaller classrooms with an average class size ranging from 16-18 students for the 2007-2008 school year. Based on the findings of this study, the following conclusions were presented:

1. There was a statistically significant difference in mean scores on DIBELS™ WUF and ORF assessments between students who were assigned to smaller classes and students assigned to regular sized classrooms. Smaller classes were found to have an impact on literacy achievement.

2. There was a statistically significant difference in mean scores on DIBELS™ WUF and ORF assessments between students who were assigned to smaller classes and students assigned to regular sized classrooms based on gender. Tukey-Kramer post-hoc test revealed differences between the means, but neither gender was found to benefit more than another.

3. There was a statistically significant difference in mean scores on DIBELS™ WUF and ORF assessments between students who were assigned to smaller classes and students assigned to regular sized classrooms based on ethnicity. Tukey-Kramer post-hoc test revealed differences between the means, but neither majority nor minority were found to benefit more than one another.

Recommendations

Based on the findings of this study, the following recommendations are made:

1. The analyses of data in this study appear to support the Clarksville Montgomery County Schools Class Size Reduction implementation. CMCSS should continue funding the Class Size Reduction program and providing extra teachers for Title I schools.

2. The data have revealed positive benefits for minority primary students in literacy outcomes after participating in smaller classes. Former research and the findings of this study suggest that all primary students could benefit from smaller classes. Research should be conducted to assess the financial impact of adding extra primary teachers to schools with high minority or at-risk populations.

Future Research

1. It would be beneficial to broaden this study to include more participants and evaluate how smaller class sizes might effect literacy achievement in students using different forms of literacy measurements. This study was limited to only two Title I schools and two forms of literacy assessments, DIBELS™ WUF and ORF. This study could be replicated to include more CMCSS elementary primary classrooms if funds were available to add extra teachers to other schools.

2. This study could be expanded to compare and evaluate achievement gaps between groups and if smaller classes have an impact on the achievement gaps. This field study was limited to a small sample size in two Title I schools with high minority populations, thus examining achievement gaps between groups was not feasible. If CSR was expanded to more schools, the impact of class size on achievement gaps could be examined.

References

Achilles, C. M. (2003). How class size makes a difference: What the research says. The impact of class-size reduction (CSR). *Paper presented at the SERVE Research and Policy Symposium on Class Size Reduction and Beyond.* Retrieved from ERIC database (ERIC Document Reproduction Service No. ED475012).

Achilles, C. M., Finn, J. D., & Pate-Bain, H. (2002). Measuring class size: Let me count the ways. *Educational Leadership, 59(5),* 24-26.

Achilles, C.M., Nye, B.A., Zaharias, J.B., & Fulton, B.D. (1993). *The lasting benefits study (LBS) in grades 4 and 5 (1990-1991): A legacy from Tennessee's four year (K-3) class-size study (1985-1989), project STAR.* A paper presented at the annual meeting of the North Carolina Association for Research in Education, Greensboro, NC. Retrieved from ERIC database (ERIC Document Reproduction Service No.ED356559).

Achilles, C.M., Zaharias, J.B., Nye, B.A. & Fulton, D. (1995). *Analysis of Policy Application of Experimental Results: Project Challenge.* Paper presented at the Annual Convention of the Mid-South Educational Research Association (MSERA), Biloxi, MS. Retrieved from ERIC database (ERIC Document Reproduction Service No. ED393151).

American Educational Research Association (Fall 2003). Class Size: Counting students can count. Research Points. *Essential Information for Education Policy.* *1(2)*, 1-4. Retrieved on March 16th, 2007 from

http://www.aera.net/uploadedFiles/Journals_and_Publications/Research_Points/RP_Fall03.pdf

Berger, M. A.(1982). *Class size is not the issue.* Paper presented at the annual meeting of the National School Boards Association Atlanta, GA, April 1982. Retrieved from ERIC database (ERIC Document Reproduction Service No. ED231061).

Biddle, B. J., & Berliner, D. C. (2002). Small class size and its effects. *Educational Leadership, 59(5),* 12-23.

Bohrnstedt, G.W. & Stecher, B. M. (2002). *What we have learned about class size reduction in California.* Sacramento, CA: California Department of Education.

Clarksville Montgomery County Schools (2006). *CMCSS 2006-2007 School System Budget.* Retrieved on March 10th, 2008 from

http://www.cmcss.net/archives/2007/Budget.pdf

Clarksville Montgomery County Schools (2008). *CMCSS 2008-2009 School System Budget.* Retrieved on October 14th, 2008 from

http://www.cmcss.net/archives/2008/Budget.pdf

Educational Research Service, (1980). Class size research: A critique of meta-analyses. *Phi Delta Kappan,* 62(4), 239-241.

Finn, J.D. & Achilles, C.M. (1990). Tennessee's class size study: Findings, implications, misconceptions. Educational Evaluation and Policy Analysis, *21(2),* 97-109.

Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology, 97* (2), 214.

Gilman, D.A. & Antes, R.L., (Fall, 1985). *The educational effects of the introduction of a state supported program of smaller classes. A study of the first year of Indiana's project PRIME TIME and its effects on test results (1984-85). A comprehensive analysis.* Retrieved from the ERIC database (ERIC Document Reproduction Service No. ED286842).

Glass, G.V. & Smith, M.L. (1979). Meta-analysis of reaserach on class size and achievement. *Educational Evaluation and Policy Analysis, 1*(1), 2-16.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). Dynamic Indicators of Basic Early Literacy Skills (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. http://dibels.uoregon.edu/

Good, R.H., Kaminski, R.A., Shinn, M., Bratten, J., Shinn, M., Laimon, L., Smith, S., & Flindt, N. (2004). *Technical Adequacy and Decision Making Utility of DIBELS* (Technical Report No. 7). Eugene, OR: University of Oregon.

Hanushek, E. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21*(2), 143-163.

Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research, 43*(6), 387-425.

Hedges, L.V., & Stock, W. (1983). The Effects of Class Size: An Examination of Rival Hypotheses. *American Educational Research Journal, 20(1)* 63-65.

Iacovou, M. (2002). Class size in the early years: Is smaller really better? *Education Economics, 10(3)*, 261-290.

Indiana State Department of Public Instruction (1983). *Project Prime Time: 1982-1983 Report.* Retrieved from ERIC database (ERIC Document Reproduction Service No. 239765).

Johnson, K. A. (2002). The downside to small class policies. *Educational Leadership, 59(5)*, 27-29.

Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from Project STAR. *The Elementary School Journal, 4(108)*, 275-291.

Krueger, A.B. (2003). Economic considerations and class size. *Economic Journal, 113 (485)*, F34-F63. Retrieved from ERIC database (ERIC Document Reproduction Service No. EJ667303).

Malloy, L. & Gilman, D. (1988). *The cumulative effects of Indiana's Prime Time: A state sponsored reduced class size program, on basic skills achievement.* Indiana State University. Retrieved from ERIC database (ERIC Document Reproduction Service No. ED300139).

McGiverin, J., Gilaman, D. & Tullitski C. (1989). A meta-analysis of the relation between class size and achievement. *Elementary School Journal, 90(1)*. 47-56.

Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis, 21*(2), 165-177. Retrieved from ERIC database (ERIC Document Reproduction Service No.EJ592471)

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5*(2), 113-127.

National Education Association (2006). *Closing Achievement Gaps: An Association Guide.* Retrieved March 16, 2008 from

http://www.nea.org/achievement/policy.html

Nye, B.A., Hedges, L. V., & Konstantopoulos, S. (2002). Do low achieving students benefit more from smaller classes? Evidence from the Tennessee class size experiment.*Educational Evluation & Policy Analyisis. 24(3)*, 94-100.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2004). Do minorities experience larger lasting benefits from small classes? *Journal of Educational Research, 98*(2)

Odden, A. (1990). Class size and student achievement: Research-based policy alternatives. *Educational Evaluation and Policy Analysis, 12*(2), 213.

Oregon State Department of Education (2005). *Closing the achievement gap: Oregon's plan for success for all students.* Oregon Department of Education. Retrieved from ERIC database (ERIC Document Reproduction Service No. ED486625).

Reichardt, R. (2001). *Reducing class size: Choices and consequences. policy brief.* Mid-Continent Research for Education: CO. Retrieved from ERIC database (ERIC Document Reproduction Service No. ED456532).

Sanogo, Y. & Gilman, D. (1994). *Class size and student achievement: Tennessee's Starla Quiroz and Indiana's Prime Time Projects.* Retrieved from ERIC database (ERIC Document Reproduction Service No. ED370680).

Smith, A., Molnar, A. & Zahorik, J. (2003). Class-size reduction: a fresh look at the data. *Educational Leadership. 61(1),* 72-74.

Tomlinson, T. M., & Office of Educational Research and Improvement (ED), Washington, D.C. (1988). *Class size and public policy: Politics and panaceas* Retrieved from ERIC database (ERIC Document Reproduction Service No. ED292216).

U.S. Department of Education (2004). *A Descriptive Evaluation of the Federal Class-size Reduction Program: Executive Summary*, Washington, D.C.

Viadero, D. (2006). Race report's influence felt 40 years later: Legacy of Coleman study with new view of equity. *Education Week, 25*(41) Retrieved from ERIC database (ERIC Document Reproduction Service No. EJ739368).

Wisconsin Department of Public Instruction (2008). *Student achievement guarantee in education (SAGE) program. Website.* Retrieved on March 21st, 2007 from http://dpi.wi.gov/sage/

Word, E., Achilles, C.M., Bain, H., Folger, J. Johnston, J. & Lintz, N. (1990). *Student/Teacher Achievement Ratio (STAR) Tennessee's K-3 Class Size Study. Final Summary Report 1985-1990.* Nashville, TN: Tennessee Department of Education. Retrieved on March 20, 2007 from http://www.heros-inc.org/star.htm

Appendices

Appendix A

Clarksville Montgomery County Approval

Clarksville Montgomery County SCHOOL SYSTEM

Sallie Armstrong, Ed.D.
Curriculum & Instruction Director

Board of Education   621 Gracey Avenue   Clarksville, Tennessee 37040
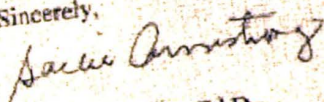931-920-7819   Fax: 931-920-9819   email: sallie.armstrong@cmcss.net

April 16, 2008

Maria Farrell
East Montgomery Elementary
920 McAdoo Creek Rd.
Clarksville, TN 37043

Dear Ms. Farrell,

Your request to conduct research in Clarksville Montgomery Co. School System has been approved. Note that one of the requirements of the district is that you provide us with a copy of the completed study.

Sincerely,

Sallie Armstrong, Ed.D.

SA:bd

Appendix B

Austin Peay State University

Institutional Review Board Approval

Austin Peay
State University

College of Graduate Studies

April 17, 2008

Maria Farrell
3977 Lakewood Dr.
Clarksville TN 37043

RE: Your application regarding study number 08-014: A comparison of Student Literacy Achievement Before and After a Class Size Reduction Implementation in Select Primary Classrooms

Dear Maria Farrell:

Thank you for your recent submission. We appreciate your cooperation with the human research review process. I have reviewed your request for expedited approval of the new study listed above. This type of study qualifies for expedited review under FDA and NIH (Office for Protection from Research Risks) regulations.

Congratulations! This is to confirm that I have approved your application through one calendar year. This approval is subject to APSU Policies and Procedures governing human subject research. The full IRB will still review this protocol and reserves the right to withdraw expedited approval if unresolved issues are raised during their review.

You are granted permission to conduct your study as described in your application effective immediately. The study is subject to continuing review on or before April 17, 2009, unless closed before that date. Enclosed please find the forms to report when your study has been completed and the form to request an annual review of a continuing study. Please submit the appropriate form prior to April 17, 2009.

Please note that any changes to the study as approved must be promptly reported and approved. Some changes may be approved by expedited review; others require full board review. If you have any questions or require further information, contact me at (221-7415; fax 221-7641; email pinderc@apsu.edu).
Again, thank you for your cooperation with the APSU IRB and the human research review process. Best wishes for a successful study!

Sincerely,

Charles A. Pinder, Ph.D.
Chair, Austin Peay Institutional Review Board
Cc: Dr. Tammy Shutt

www.apsu.edu

P.O. Box 4458 • Clarksville, TN 37044 • P: (931) 221-7414 • F: (931) 221-7641