2006

Student achievement as a function of class size and pupil-teacher ratio

Robert T. Becker

STUDENT ACHIEVEMENT AS A FUNCTION OF CLASS SIZE AND

PUPIL-TEACHER RATIO

by

Robert T. Becker, Jr.

Submitted to the Department of Leadership and Counseling

Eastern Michigan University

in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

Educational Leadership

Dissertation Committee Members

Charles M. Achilles, Ed. D.

Helen Ditzhazy, Ph. D.

Dr. Ella Burton, Ed. D.

Michael Paciorek. Ph. D.

November 13, 2006

Ypsilanti, Michigan

Acknowledgments

I would like to thank the instructors and staff in the Department of Leadership and Counseling at Eastern Michigan University (EMU) for all their work and encouragement. I have been touched by how ready to lend a hand the instructional staff was. Their encouragement and assistance made this accomplishment possible.

I want to give special thanks to Dr. Charles Achilles, my dissertation chair. His support and guidance brought this dissertation to completion. His commitment to his students is singular. I also want to recognize Dr. Helen Ditzhazy, my graduate advisor, for her support throughout my pursuit of a doctoral degree; her counsel kept me focused. I want to also thank the rest of my committee members (Dr. Ella Burton and Dr. Michael Paciorek) for their support and encouragement.

In addition to the staff at Eastern Michigan University, I want to thank my wife, Joan, and children, Brigette Dumont, Robert T. Becker, III, Brittany Jagoda, and Blake C. Becker. They are the most important presence in my life, and I appreciate the sacrifices they made as I pursued my doctoral degree.

I also want to thank my parents, Robert T. and Valerie A. Becker, for all their love and support. Your confidence sustained me.

To Eric, Nick, Barb, Linda, Kathy, and Kori – the GVSU/EMU 2000 Cohort – you're the BEST!

Finally to Dr. M – couldn't have done this without you.

Abstract

Background: The degree to which class size is able to produce positive, enduring effects on student achievement has been and continues to be vigorously debated. Comparative studies have clouded the issue with imprecise use of the terms class size (CS) and pupil-teacher ratio (PTR), making it difficult to draw clear conclusions regarding the effects of class size on student achievement.

Purpose: To assess differences, if any, in achievement between students attending classes where the class-size is approximately n = 17 and students attending classes where the class-size is approximately n = 25 and the pupil-teacher ratio is approximately 15:1.

Setting: Six public schools from suburban and rural locations in Michigan. A total of 117 third-grade students and 125 fourth-grade students participated.

Research Design: Nonexperimental, ex-post facto, cross-sectional study.

Data Collection and Analysis: Student achievement data were gathered from standardized assessment tools normally collected at each school. Student achievement scores were analyzed by using a two-tailed *t* test to determine differences in the scores of students in class-size and pupil-teacher-ratio settings. Information was gathered to determine the extent to which the schools providing data for class-size settings had implemented class-size reform as defined by current research.

Observation data were gathered to determine differences in the amount of square footage per student and possible behavioral differences between students in the two settings.

Findings: Significant differences in student-achievement-outcome scores between students in class-size and pupil-teacher-ratio settings were identified in third-grade mathematics and fourth-grade reading but not in third-grade reading or fourth-grade mathematics. None of the sites supplying class-size data for the study had fully implemented class-size settings according to current research. Students in small classes exhibited significantly fewer instances of disruptive behavior and had significantly more square feet per student than did the students in pupil-teacher-ratio class settings.

Conclusions: Analyses showed that in one half of the categories there were significant differences in the achievement of students in CS and PTR settings, while in the other half of the categories there were no significant achievement differences. Class-size initiatives not implemented in accordance with current research do not produce positive effects on student achievement; however, CS did improve the classroom environment.

Table of Contents

List of Tables

Chapter 1

Background

Introduction

An interest in improving the quality of education for students is the reason I chose to become an educator. My independent research is designed to advance my knowledge of how class size influences student achievement.

With the advent of high-stakes testing in many states, school district leaders search for how to meet the test requirements has intensified. Educating students in small classes from kindergarten through third-grade (K-3) has demonstrated greater and more sustained improvement in student achievement than has educating those students in traditional-size classes (25-28 students). The gains achieved in the first four years by being in a small class are sustained when students return to regular-size classes, and they continue to grow, in a long-term trajectory of success.

The Problem for this Study

The decision to study the differences in achievement between students in small classes, or class-sizes (CS), and those in classes with a pupil-teacher ratio (PTR) approximately equal to that of a small class surfaced after reading the misinformation generated by opponents of class-size reductions (CSR). Opponents of CSR as a means of improving student achievement so commingle the terms class-size and PTR that many school leaders are confused when PTR data and outcomes are used to represent CS data and outcomes. Examining student achievement and learning in small classes and comparing those to student learning in classes with a PTR within the range of small classes may help address the issue of clarity. Comparing student achievement in these two settings should help educators and policy personal determine which method improves student performance more advantageously (PTR or CS) and then use this information to improve student achievement.

Background for the Study

In this independent study, the researcher analyzed student achievement as it relates to CS and PTR in an attempt to shed some light on differences in outcomes, if any, between students taught in small classes (about n = 17 students) and students taught in larger classes (about n = 25 students) in schools and grades that have a PTR about equal to that of a small class (about 15:1). This study was undertaken to try to obtain some clarity in the use of two concepts that have been used imprecisely for many years.

Statement of the Problem

Individuals using PTR data to report on CS may be improperly dismissing the positive impact of CS research on student achievement. This PTR-CS comparison sows the seeds of confusion among practitioners. The practice of using PTR data to describe CS effects on student achievement seems to be one way to show that there is no positive effect of CS on student achievement. The use of imprecise comparisons of PTR and CS data on student achievement could be hindering a demonstrated method of helping students learn and achieve higher outcomes. This becomes especially critical when considering the school experiences of minority students, students of low socioeconomic status (SES), and limited-English-proficiency (LEP) students. Indeed, research shows that these students have continuously demonstrated highly positive achievement gains from being in small classes in early grades (e.g., Word et al., 1990) when compared to gains of other students.

This issue becomes critical in terms of student achievement, when taken in context with the research of Bloom (1984), who found that through tutoring and a mastery learning approach, “the average student is two-sigma above the average control student” (p. 4). Bloom stated that “the tutoring process demonstrates that most of the students have the potential to reach this high level of learning” (p. 4). Researchers are using CS research to demonstrate positive effects on student achievement and to seek a reasonable group size for student learning between tutoring (one-to-one) and the large classes of most schools today. Other researchers are reporting that CS has little or no positive effect on student achievement. A difficulty arises when some researchers use PTR data to *support* their explanation that CS does not work, whereas in fact the two terms (PTR and CS) are not synonymous.

This would seem like an easy issue to resolve if researchers and others discussed data only from each classification (i.e., PTR and CS as separate concepts) and did not mix the two different concepts. CS data are derived from studies of students actually in each class, achieved by counting the students in attendance. PTR data are derived from studies of students in a particular grade or school divided by the number of education professionals at the site, often including administrators, librarians, specials teachers, etc. Clarification of these issues needs addressing for the sake of precision, clarity, and improved data use in order to assist all children. Thus, the definitions used in the present study need to be explicated clearly.

Definition of Key Terms Used in This Study

The following is a list of terms and definitions that are important to this study.

Achievement – student scores on standardized tests, behavior within the classroom, and participation in class discussions. In the present study, however, the primary focus was on test-score outcomes, as these data are more readily available and standard than are other outcome data.

Average Class-size – "Average class-size is derived by dividing a unit's (i.e., grade level, building, etc.) total student enrollment by the number of general education or 'regular' classroom teachers" serving the site (Sharp, 2002, p. 12).

Class-size (CS) – "the number of students for whom a teacher is primarily responsible during a school year" (Lewit & Baker, 1997, p. 113). Arrived at by adding the number of students present in the classroom.

Circulation Square Footage – the area, free from student or teacher furniture, that is available for student movement within the classroom.

Classroom Square Footage – the length, in feet, of a classroom multiplied by the width, in feet, of a classroom.

Duration of Observations – the number of minutes a researcher spent in CS and PTR class settings observing student behavior.

Early Intervention – starting students in small classes when they enter school in kindergarten (K) or pre-kindergarten (PK).

Intense Treatment – "the pupil spends all day, every day in the small class. Avoid PTR events such as 'pull-out' projects or team teaching." (Achilles 2005, p. 15)

Interruptions of Single Student – any interruption of another student, group of students, or class caused by a single student with no apparent purpose except to cause a disturbance.

Interruptions of Student to Student – any interruption between two or more students wherein they are trying to communicate to another student or group of students.

Number of Students – the actual number of students in attendance in a classroom, determined by counting each student in a classroom or derived from the class list.

Number of Teachers – the teacher or teachers responsible for instruction in a single classroom.

Number of Teacher Aides – the number of individuals other than the teacher of record who assist in instructional delivery in a classroom, calculated using Full Time Equivalents (FTE).

Pupil-Teacher Ratio (PTR) – "the number of students in a school or district compared to the number of teaching professionals" (McRobbie et al., 1998, p. 4). The PTR is arrived at by dividing the number of students attending a school by the number of teachers, instructional aides, pull-out teachers, and other educators as determined by the district. "The difference between PTR and CS in USA elementary schools is about n = 10" (Achilles & Sharp, 1998). In PTR calculations, the divisor is important but not a standardized variable.

Questions of Content – questions asked by students of their teacher regarding the content of the lesson.

Questions of Instruction – questions asked by students of their teacher regarding instructions on how to complete an assignment.

Regular Classes – classes of 24-28 students per teacher.

Small Classes – classes of 13-17 students per teacher (Word, et al., 1990), up to classes of about 15-18 students per teacher.

Square Feet Per Student – the classroom square footage divided by the number of students in a classroom.

Sufficient Duration – "Maintain the small-class environment for at least three, preferably four years for enduring effects"(Achilles 2005, p. 15).

Teacher Aides – Individuals other than the teacher of record, who assist in instructional delivery in a classroom, calculated using Full Time Equivalents (FTE).

Delimitations of the Study

In this study, the researcher

- Examined CS and PTR in public schools in Michigan. This did include charter schools, but did not include private schools.
- Studied only achievement outcomes of students in grades three and four
- Relied upon student-achievement-outcomes from the 2004–2005 school year as obtained from the Michigan Education Assessment Program (MEAP), a statewide criterion-referenced test given in grade four. Third-grade student-achievement outcomes were obtained from California Achievement Test (CAT), California Test of Basic Skills (CTBS), Iowa Test of Basic Skills (ITBS), Northwest Evaluation Association (NWEA), and Terra Nova (TN).
- Observed students in CS and PTR classrooms as a nonparticipant observer

Limitations of the Study

The study had several limitations, identified here:

- The random sampling of CS and PTR in third- and fourth-grade classrooms in Michigan public elementary schools limits the ability of valid generalizations made from these data to these grade levels in other Michigan public schools.
- Valid generalizations can not be made regarding CS and PTR conditions in other states, as the researcher used Michigan CS and PTR data exclusively.
- Actual CS data are difficult to obtain. States do not regularly collect class-size or average class-size data.
- Because the study was voluntary, the only data available were limited to data from those districts willing to participate.
- Third-grade, standardized student achievement data came from a variety of testing instruments because there was no state standardized testing instrument for third-grade.
- Random observation data of three CS and three PTR classrooms are the subjective assessment of the researcher, but data were collected in a common format.
- Random observation data of CS and PTR classrooms limit the ability to valid generalizations made from these data at similar levels in other Michigan public schools.

In this nonexperimental, ex-post facto, cross-sectional explanatory study, the researcher used data received from the specific standardized testing instruments (mentioned above) in third-grade and MEAP results for fourth-grade students to estimate the effect or noneffect that CS or PTR has on student achievement. To the degree possible, the researcher selected comparability in CS and PTR settings to assure that treatments for both groups were the same, such as assignments to classes, assignment of instructional staff, and assignment of state foundation grant. To the degree attainable in a nonexperimental study, the only difference in the groups was CS and PTR.

Importance of Findings to Research and Practice

The impact of CS on student achievement has been one of the longest lines of research in education and also one of the most hotly debated topics in education research. Notable research journals and professional magazines such as Educational Evaluation and Policy Analysis and Educational Leadership have dedicated entire issues to the class-size debate. Yet confusion remains surrounding class-size effects. This confusion seems to settle around those researchers who take the position that there is no consistent positive CS effect on student achievement. Invariably, however, these researchers cite PTR data to support their CS position. This confusion leaves educators looking outside of class-size for answers on improving student achievement.

Study results should provide clarity to the educational community as to which treatment, CS or PTR, influences student achievement more positively. This clarification allows school district personnel to move in the appropriate direction and improve the schooling condition for students by choosing class-size organization or pupil-teacher ratio organization.

Organization of the Study

Chapter 1 provided an introduction, the problem for this study, the purpose of the study, the background for the problem, and the rationale for the study. Included as well were the delimitations and limitations of the study, a brief review of the design and methodology, definitions of key terms used in the study, and the potential importance of the findings to research and practice. Chapter 2 consists of a review of the literature related to both CS and PTR. Chapter 3 details the research design and methodology used in the gathering and analysis of data. Chapter 4 presents the results and analysis of the data. Chapter 5 includes the summary of findings, conclusions, and recommendations for practice, for policy, and for further research.

Chapter 2

Review of Related Research and Literature

Because every day educators assess programs to improve student achievement, there must be clarity surrounding the data used to make educational decisions. It is common for the large pupil-teacher ratio (PTR) database to be referenced by researchers commenting on class-size (CS) issues. Researchers using large databases often blur the lines between these two concepts and settings, confusing educators. As administrators review the voluminous research on CS and PTR and their effects on student achievement, the administrators must have precise definitions of terms in order to improve their understanding of actual conditions and effects.

The review of research and literature consists of studies on CS and/or PTR. Several large CS studies supply data for analysis and comparisons. The Tennessee Student-Teacher Achievement Ratio (STAR) experiment, Project Prime Time in Indiana, and the Student Achievement Guarantee in Education (SAGE) in Wisconsin are but a few of the CS studies available. The review of research and literature includes many studies done to evaluate the effects of CS and PTR on student achievement.

Clarity of Data

Clarity in definition of terms related to CS and PTR conditions is evident in Texas H.B. 72 of 1984 (Appendix A), which initiated CS reform in Texas. H.B. 72 mandated that each school district employ a sufficient number of certified teachers to maintain an average of not less than one teacher for each 20 students in average daily attendance. The focus of this bill was to limit the number of students enrolled in a classroom, true class-size reduction (CSR). H.B. 72 focused CSR efforts on students in grades K-2, where the CS benefit would improve student achievement immediately, and the lasting effect would continue when students began to attend classes with larger enrollments.

In the report Smaller Classes, Not Vouchers, Increase Student Achievement, Molnar (1998) identified the issue at the heart of the CS and PTR confusion: “low pupil-teacher ratios do not always mean small class-sizes” (p. 28). “The terms pupil-teacher ratio and class-size are often used interchangeably in everyday conversation. Most people understand both terms to mean the average number of students in a typical classroom with one teacher; this is a false assumption” (p. 28). Molnar then provided one definition of the difference between the two concepts.

> One calculates pupil-teacher ratio by dividing the number of students by the number of instructors holding teaching certificates whose primary responsibility it is to teach. These instructors include teaching specialists in areas such as physical education, art, reading, and special education, as well as Chapter I “pull out” teachers (pull-out teachers remove students from the regular classroom who qualify for means-tested specialized instruction.) One calculates average class-size by surveying classroom teachers and asking how many students are in their classes. (p. 28)

There is no standard or uniform formula for districts to use in determining PTR consistently. Because the claim of a favorable PTR seems a good public relations technique, some districts’ divisors for PTR may include counselors, administrators, and teacher aides. A favorable PTR would certainly be attractive to parents, confused by imprecise reporting of CS/PTR research, who seek small class settings for their children, only to find that the PTR was achieved by large classes with many pull-out programs.

The Confusion of Terms

If research and analysis are to be understood and used by practitioners, the data presented must be clear and concise. Boozer and Rouse (1995) addressed the CS and PTR confusion directly and found some important differences in CS and PTR. Figures 1a and 1b of their study relate to school size and demonstrated that "the larger the school, the more variance and thus, the larger was the difference between CS and PTR" (p. 5). They noted that "the correlation between the pupil-teacher ratio and the average class-size is relatively low at 0.13 in the New Jersey Survey and 0.26 in the NELS" (Boozer & Rouse, 1995, p. 5). Boozer and Rouse found that "the pupil-teacher ratio does not (statistically) increase in schools with a larger proportion of black students, but that the average class-size does" (p. 8). Boozer and Rouse also stated,

> The fact that school average class-size matters, but pupil-teacher ratio does not suggest a second reason that researchers may not have detected racial differences in class-size: the pupil-teacher ratio (more than the average class-size). The way to accomplish keeping the PTR the same while the average class-size increases is to add additional staff for "pull-out" interventions, projects like Title I, our nation's largest remediation effort. (p. 9)

Researchers and analysts need to be clear in use of the terms CS and PTR. Proponents of class-size, (e.g., researchers such as Achilles, Boyd-Zaharias, Finn, Glass, Krueger, Molnar, Pate-Bain, etc.) have generally presented their data within the framework of clearly defined terms. Economists, such as Hanushek (1998, 1999), a leading commentator and sometime critic of CS research, may add to policy makers' confusion between CS and PTR by reporting PTR outcomes labeled as CS data. For example, Achilles (2005) noted the confusion caused by Hanushek (2002), who, while discussing class-size, offered a Table 2-1 titled “Pupil-Teacher Ratio and Real Spending, 1960-1995” (p. 39). Achilles (2005) showed that Hanushek’s (2002) data were PTR data from Table 65 in Digest of Education Statistics1999 (National Center for Education Statistics [NCES], 2000), labeled “Pupil-Teacher Ratio.” Table 65 data called PTR (NCES, 2000) are the same as the data Hanushek (2002) called CS (see Table 1).

Table 1

Comparison of Pupil–Teacher Ratio Data: Digest of Education Statistics (1999) Table 65 and Hanushek (2002) CS Data

Term/Concept	1960	1970	1980	1990	1995	Source
Pupil-teacher ratio	25.8	22.3	18.7	17.2	17.3	(Hanushek, 2002)
Pupil-teacher ratio	25.8	22.3	18.7	17.2	17.3	(NCES, 2000))

Source: Achilles, 2005, p. 13

Hanushek follows his table with the statement “Perhaps the most astounding part of the current debates on class-size reduction is the almost complete disregard for the history of such policies. Pupil-teacher ratios fell dramatically throughout the 20th Century” (Hanushek, 2002, p. 39). Statements like these that mix PTR and CS as synonyms may lead to incorrect conclusions regarding the effectiveness of CS, necessitating diligence on behalf of the reader to untangle the web of confusing rhetoric.

In their article “What Research Says About Small Classes and Their Effects,” Biddle and Berliner (2002) stated that “most of the studies Hanushek has reviewed do not provide evidence on class-size, but some seemed to focus on the class-size issue, and after reviewing the latter as well, Hanushek has announced that class-size also appears to have little impact" (p. 4). Thus, readers and policy people must understand the research/analyses they are reading because, as Biddle and Berliner pointed (2002) out, "although Hanushek is clearly aware that student-teacher ratio is not the same thing as class-size (see Hanushek, 1999, p. 145) he has continued to argue that his reviews of literature based on the former imply findings to the latter" (p. 18). Characteristics such as the research background of the author, the accuracy of the information cited, the political bent of the institution publishing the study, and the sources used to corroborate the findings help inform the reader of the quality of the study and of the actual phenomenon being studied. Practitioners cannot rely upon editors of secondary or tertiary reports to cull out all inaccuracies or inconsistencies that authors present in their papers. For example, the February 2002 issue of Educational Leadership, the magazine of the Association for Supervision and Curriculum Development (ASCD), was dedicated to research on class-size and school size. In the article "The Downside of Small Class Policies" (Johnson, 2002), the statistics used to demonstrate a drop in class-size were, in reality, data for PTR over time; the exact data from Table 65 of Digest of Education Statistics 1999 (NCES, 2000), cited precisely. This misrepresentation was not discovered by ASCD editors but by one of the principal investigators (PI) of the Tennessee STAR experiment. The correction was not published until May 2002, and then an interested reader could only find the correction buried in an inconspicuous place. Achilles (2002), stated,

> In one article, however, two very different ways to think about class numbers—*class-size* and *pupil-teacher ratio* were treated as synonyms. The numbers used to support the argument about class-size in "The Downside of Small Class Policies"

> come directly from Table 65 of the Digest of Education Statistics (NCES, 2000), which is about pupil-teacher ratio, not class-size, Table 69 reports class-size. (p. 88)

The importance of ensuring that research/analysis in education is accurate is poignantly made by Achilles and Finn (2002), who stated,

> Educators often make decisions about other people's children and money, so their decisions need to be good; their sources valid; their applications in line with the research; and their evaluations able to be interpreted based both on the research (quality and fidelity) and on the theory (predictability). (p. 3)

Conflict generated by confusing, and then comparing, CS and PTR data as the same, is detrimental to seeking a solution to Bloom's two-sigma challenge (1984). This conflict is evident in the analysis of data contained within the National Assessment of Educational Progress (NAEP) information. Hanushek (1998) reported how the NAEP results for 17-year-olds stood out in three aspects: "first, overall performance is approximately the same since 1970; second, there has been some convergence of scores between whites and blacks and Hispanics; third, the convergence of scores by race and ethnic groups may have stopped during the 1990's" (pp. 5-6). Researchers such as Grissmer, Flanagan, Kawata, and Williams (1998, 2000), Hauser (1998), and Hedges and Nowell (1998) also reviewed the NAEP data. They found gains of 0.6-0.8 standard deviation units for Black students and 0.2-0.3 for Hispanic students. Additionally, when NAEP data, (mathematics summary trend data for nine-year-olds) were reviewed, they showed that in 1990, the year following the conclusion of the STAR experiment, there was a large improvement in scores. This increase was across the board, including males, females, Whites, Black, Hispanics, and others. Improvements ranged from an increase of 6 points for Black students to one of 13 points for other student groups.

In a section on "Econometric Evidence," Hanushek (1998) blurred the lines between CS and PTR as follows: "These econometric estimates relate class-size (CS) or teacher intensity (ed. whatever that is?) to measures of student performance" (p. 20). This statement followed an earlier point, in the same article, that stated, "First, pupil-teacher ratios are not the same as class-size" (Hanushek 1998, p. 12). Hanushek confused CS/PTR differences when summarizing the results of "377 econometric studies of the determination of student performance, 277 consider pupil-teacher ratios" (p. 21). Hanushek then concluded that the top line of the table showed only a "15% statistically significant relationship between teacher intensity and student performance—the expected result if class-size systematically matters" (p. 21). (Emphasis added)

Hanushek continued the confusion regarding data analysis on student achievement with conflicting statements about what data were available. He stated, "The only data that are available over time reflect pupil-teacher ratio data" (Hanushek, 1999, p. 12). In the same source, he then recounted "several hundred separate estimates of the effects of reduced class-size" but used the PTR data to make this claim (p. 33). It would appear there were sufficient data available to evaluate the impact of class-size on student achievement. By using PTR data instead of CS data when reporting on the effect of reduced class-size, , Hanushek commingled parts of each definition to achieve misleading results because of unclear data naming and use.

Practitioners under the No Child Left Behind (NCLB) Act are challenged to use Scientific-Based Research (SBR) as the basis for school and student-achievement-outcomes. To do this, they need to be able to believe the research they read. "In an ERIC Clearinghouse book and after warning people that PTR and CS are not the same, the author then used PTR data to project cost for small classes. [Chapter 4, Picus (2001) in The Search of More Productive Schools]" (Achilles, 2005, p. 5). Another example of federally supported publication source's not reporting data accurately can be found in a chapter by Harris and Plank titled "Making Policy Choices: Is Class-size Reduction the Best Alternative?" This chapter in the Federal Education publication Using What We Know: A Review of the Research on Implementing Class-Size Reduction Initiatives for State and Local Policy Makers, by Laine and Ward (2000) "had no class-size data; instead, the authors used PTR data from Table 65 of The Digest of Education Statistics 1999 (Achilles, 2005, p. 6). The American Education Research Association's (AERA) American Educational Research Journal published an article of which B. Nye was a co-author. The paper was titled "The Effects of Small Classes on Academic Achievement: The Results of the Tennessee Class-size Experiment." The way the article is titled could lead the reader to assume that Dr. Nye was a part of the STAR study. Neither Nye nor either other author was an investigator of STAR.

Administrators and policy makers rely upon research published by research organizations such as the AERA as well as the federal government for data reported clearly and precisely. Achilles (2005) stated that when these "scientific outlets get it wrong, and policy leaders read and believe the rationale put forth through these outlets, they feel no need to inquire further to find out the validity of the research" (p. 12).

This review of literature and research thus far has investigated some confusion caused by reporting PTR research findings as if they were CS findings and the need for clarity in definition of terms when reporting data (see Figure 1). A clear definition of terms may encourage the use of CS to obtain longlasting positive, enduring effects on student-achievement-outcomes.

The Class-Size Findings

Authors of research emanating from major CS studies, such as Indiana's Prime Time, Tennessee's Project STAR, Wisconsin's SAGE initiative, and others, report their data and findings as CS results. This is exemplified in Word et al. (1990, p. 110, Table V-6 and p. 111, Table V-7), in which the authors presented longitudinal analyses of average annual reading and math scores. The tables are clear and concise, displaying scores in reading and math from each of the three treatments, that is, small classes (S), regular classes (R), and regular classes with a full-time teacher's aide (RA). The results clearly showed that outcomes of students in small classes were higher in each geographic school setting (inner city, suburban, rural, and urban) than were the scores of randomly assigned peers in large classes (n = 22-25 students) or large classes with a full-time teacher's aide. Never in the narrative, table title, or body of the table is there reference to anything other than CS data. A clear definition of terms aids the clarity of CS research. Word et al. (1990) stated exactly the parameters of the two treatment and one control groups: "small classes have an enrollment of 13-17, regular classes of 22-25, and regular classes with a full-time aide of 22-25. Each participating school must accommodate at least one of each treatment group, from kindergarten through third-grade" (p. 10). There is no ambiguity of terms, not one mention of PTR, just straightforward representation of intent and discussion of CS.

The Indiana legislature instituted class-size reductions (CSR) in 1981 with its Project Prime Time. The state provided funds in 1984 for school corporations to “reduce first-grade classes to an average of 18 students (or 24 with an instructional assistant)” (Chase, Mueller, & Walden, 1986, p. 1). The fall of 1985 saw second-grade added to Prime Time, and the final addition came in the fall of 1986 with Indiana school corporations given the choice of adding third-grade or kindergarten with an aide. The result from Project Prime Time demonstrated modest gains at best. In second-grade mathematics, only 10% of corporations sampled had significantly higher post-Prime Time than pre-Prime Time scores. Only 20% of corporations reported higher second-grade reading scores; 30% of corporations reported higher scores in first-grade mathematics scores. The largest gains in post-Prime Time versus pre-Prime Time scores came in first-grade reading, with 50% of Indiana School Corporations’ reporting higher student-achievement-outcomes. When third- and fourth-grade teachers were asked to identify what they considered the major advantages of Prime Time, “53% responded that the opportunity for individualized instruction was the most important advantage. The third item on the teachers list of advantages was the ability to more quickly diagnose student needs” (Chase, Mueller, & Walden, 1986, p. 152).

Analysis of the Wisconsin’s SAGE program showed achievement gains similar to those found in the STAR experiment in the early years. SAGE differed in design from STAR in two notable ways: first, SAGE used a quasi-experimental design with naturally occurring classes as the control group; second, SAGE measured small-class achievement against that of naturally occurring control groups. The STAR researchers, however, employed a randomized experimental design with two control groups (R, RA) for each small class. SAGE was not only a class-size study. Some of the comparison treatments included the use of "2-Teacher Teams and Floating Teachers" (Molner et al., 2001, p. 16); this PTR distinction was clearly spelled out. If there were not enough rooms available in a school to lower CS, another teacher was put into the classroom with 28-30 students to bring down the number of students per teacher (a PTR treatment). The reporting of achievement data is clearly shown as either CS or PTR. The positive results of SAGE were similar to those of STAR, as shown in SAGE that "first-grade students scored 4 scale points higher in language arts, 4.3 scale points in reading, 4.6 scale points higher in mathematics and 4.6 scale points higher in total test scores than Comparison school students" (Molnar et al., 2001, p. 9). Because all data were reported, researchers could separate the true CS sections from the PTR sections for analysis.

Burke County, North Carolina, initiated CS reductions in the 1991-1992 school year by instituting a pilot program of small classes in the first-grade classrooms of four elementary schools. Burke County schools had been designated as low wealth by the North Carolina Department of Public Instruction; approximately 34% of Burke County students received free or reduced-price lunch. Small classes were systematically implemented from pilot schools the first year to all first-grades in the second year. That is, in the 1992-1993, school year administrators implemented small classes in all first-grade classrooms in Burke County, and the original four schools then piloted small classes in the second-grade through grades 1-3. Egelson and Harmon (2000) reported that "currently, all first-, second-, and third-grade classrooms at the 17 elementary schools in Burke County use small classes" (p. 281). The administration at Burke County realized that the transition to small classes might require staff development to alter teaching methods to complement the CS change. “Staff development focused on training staff to take advantage of one-on-one interactions with students, as well as how to develop effective student-centered activities” (Egelson & Harmon, 2000, p. 282). Additional staff development opportunities helped teachers “identify and respond to student needs, building on the strengths of students to mediate weaknesses, and develop positive classroom management” (p. 282). The systemic implementation of CS by Burke County educators resulted in significant gains in student-achievement-outcomes, even beyond the K-3 implementation: “Between 1992 and 2000 the district students improved their math scores in third-grade from 61.6% to 80.5%, in fourth-grade from 66.4% to 89.4%, and in fifth-grade from 61.9% to 88.6%” (Sharp, 2002, p. 31).

The state of California initiated class-size reduction (CSR) in 1996, following the success of Tennessee’s STAR study. California’s CSR targeted reductions in class-size from 28-30 students per class to 20 students or fewer in kindergarten through third-grade (K-3) over time. The California CSR effort was implemented beginning with first and second-grades state-wide during the 1996-1997 school year, involving more than 1.6 million students. The results associated with the CSR reform in California were not large: “the percent of students whose SAT-9 scores were above the 50th national percentile rank was 2-4 percentage points greater for third-grade students in reduced sized classes compared to those in larger classes” (Bohrnstedt, Stecher, & Wiley, 2000, p. 205). These results should not have been a surprise. Students in second-grade would have had only one year in a CS class, which is not, as pointed out by STAR research, a long enough duration to have a positive impact on student-achievement outcomes.

There were other issues with the California CSR implementation, as Bohrnstedt, Stecher, and Wiley (2000) stated,

> A number of differences exist between the two CSR research contexts. The two reforms differ in term of size and scope, available facilities, availability of trained teachers, actual class-size, student diversity, existing curriculum, and existing state assessments; each of these conditions could have mitigated the impact of CSR in California. Policy makers in other states who are considering the adoption of CSR should look for ways to address these important differences. (p. 207)

The CS/PTR debate has enticed numerous researchers to examine data on both sides of the issue, drawing their own conclusions. Nye, Hedges, and Konstantonpoulos (1999) used Hierarchical Linear Modeling (HLM) on STAR data and "estimated between 0.11 and 0.20 standard deviations," finding a stronger effect than did the original analyses of STAR data. Krueger (1999), Molnar (1998), and Nye et al., (1999) investigated the assertion of flaws in STAR research, made by economist Hanushek (1999), such as claims regarding the presence of the Hawthorne effect, differing school curriculum, and types of students and teachers, that may make the results not generalizable."(p. 153) Results of the various investigations supported the robustness of STAR. Krueger (1998) tested STAR data to determine the impact of the Hawthorne effect. Krueger (1998) determined that "there is little support for the view that the main experimental results are contaminated by the Hawthorne effects" (p. 36). Nye et al. (1999) found that one group of students and staff was randomized within each school and that it is unlikely that school characteristics could have biased the effects.

Grissmer, Flanagan, Kawata, and Williams (1998), Hauser (1998), and Hedges and Nowell (1998) have analyzed NAEP data. Each researcher or team of researchers demonstrated achievement gains related to small classes; only Hanushek either did not find or report them. Bracey (1999) stated with regard to NAEP data,

> The increasing proportion of minority students in the NAEP sample has attenuated the overall average and hidden gains made by blacks and Hispanics at all grades and performance levels. Low-scoring white students also showed gains. If one were to say that NAEP scores are "stagnant" as Hanushek does, one could only be referring to the scores of high-performing whites. (p. 2)

In their review of the effects of CS, Biddle and Berliner (2002) responded to Hanushek's work by stating that "Hanushek has not responded well to criticism; rather he has found reasons to quarrel with the details and continue publishing reviews claiming small classes have few to no effects." (p. 15). Additionally, Biddle and Berliner (2002) asserted, "Because of these responses and activities, it is no longer possible to give credence to Hanushek's judgment of class-size" (p. 15).

Conditions/Principals for Enduring CS Effects

Ramey and Ramey (1998) identified six probable principles of program efficacy for early interventions. Relative to CS effects, the most important of the six are (a) developmental timing, (b) program intensity, and (c) direct provision of learning experiences. They found that intervention programs must start early in a child's education, in kindergarten or pre-kindergarten; that the numbers of hours and days of the week are key in children's learning; and that the program must be delivered by someone other than parents if there is to be an enduring impact on later school performance. Biddle and Berliner (2002) reported, "When it (CS) is planned thoughtfully and funded adequately, long-term exposure to small classes in the early grades generates substantial advantages for students in American schools, and those extra gains are greater the longer students are exposed to those classes" (p. 14) .Subsequent work done by Achilles (2005), citing studies done by Finn and Achilles (1999), Finn, Gerber, Achilles, and Boyd-Zaharias (2001), and Nye, Hedges, and Konstantonpoulos (1999) on the STAR experiment data have shown that three conditions must be met for early class-size gains to have an enduring effect: "Those conditions are (a) Early Intervention (when the child starts school), (b) Intensity (all day, every day), (c) Duration (pupils must remain in small classes (13-17 students) three and preferably four years to demonstrate lasting effects" (Achilles, 2005, p. 1). Finn, Gerber, and Boyd-Zaharias (2005) noted that the most robust, enduring effects "were strongest for students who entered small classes in kindergarten or Grade 1 and who remained in small classes for three or more years" (p. 216).

In an update of the STAR research following students into high school and beyond, Finn (2005) stated that the "direct provision of learning experiences, rather than relying on intermediary sources" (p. 1) would contribute the lasting impact of CS on student-achievement-outcomes. Ladson-Billings and Gomez (2001) reported findings of their qualitative study regarding students' receiving assistance from professionals, such as in Title I and Reading Recovery programs, throughout the school day. They found that "students who received services from a variety of professionals were more likely to be confused about to whom they were responsible" (p. 677).

Implementation Issues

Early intervention programs have been successful in achieving enduring effects. The Perry Preschool Project and most Head Start programs have seen student early-achievement benefits decline, and even disappear, soon, up to three years after the participants have left the program. Key to this evaporation of enduring effects may be the limited intensity of the above-mentioned programs. According to Finn, Gerber, and Boyd-Zaharias (2005), students in the Perry Preschool Program "received only 2.5 hours of school time daily and the typical Head Start program involved 3.5 hours of class time 4 or 5 days a week" (p. 216). Contrast this to STAR, in which students were in small classes all day throughout the entire school year and students continued in small classes for up to three to four years with teachers randomly assigned to classes each year. Achilles (2005) noted that Blatchford, Bassett, Brown, Martin, and Russell (2004) found "an important 'disruption' effect on children's educational progress ... moving to a class of a different size, especially a larger class was disruptive ... it is advisable to maintain smaller classes and to seek … stability in class-size across years" (p. 2)

In their study "Small Classes in the Early Grades, Academic Achievement, and Graduating from High School," Finn et al. (2005) found that "attending small classes for three or four years in the early grades had a positive effect on high school graduation, above and beyond the effect on early academic performance" (p. 219). "The odds of graduating were 67.0% greater for students attending small classes for three years and almost 2.5 times greater for students attending small classes for four years" (Finn et al.

2005, p. 219). Benefits of small classes extend beyond academic achievement to reduced incidents of inappropriate student behavior and an improved sense of classroom community.

Following “Project Success” class-size reduction, Principal Jean Owens, of Oak Hill Elementary, High Point, North Carolina, interviewed the staff for differences in the classroom experience from the previous years. Among the information learned through Owens’s interviews was that “classroom management and discipline are better… students develop better human relations and have greater regard for others … students learn how to function more effectively as members and leaders of groups” (Achilles, 1999, p. 47). Olson (1971) researched the impact of different class-size groupings and developed “nine defensible generalizations about class-size when teachers teach fewer rather than more students” (p. 65). The results of the interviews conducted by Owens reflected the same found by Olson almost 30 years earlier: “Teacher and student attitudes improved; discipline and classroom management is better; and students display good human relations” (p. 65).

Similar student behavioral results were documented by researchers of the Project SAGE study, who noted that “little time is required to manage the class or to deal with discipline problems” (Achilles, 1999, p. 48). Summarizing the “Immediate Observable Outcomes of Class-size of 15 and Class-size of 24,” Achilles (1999) noted a pair of opposites regarding students; in the “Class-size of 15” there was “more personal space for each student and a sense of peacefulness in the class” (p. 50). In the “Class of 24” there were “more student conflicts in the classroom and less space for each child” (p. 51).

Community, Engagement, and PSOC

During the 1993-1994 school year, Achilles, et al. (1994) conducted a “year-long observation study of teaching behaviors of teachers in small classes and teaching behaviors of teachers in classes of about 24 students” (p. 52). The small classes had about 14 students. One of the noteworthy observations was that students in small classes “reduced their discipline problems by one-half as they experienced small classes” (Achilles, 1999, p. 53). Mosteller stated that “reducing [the size of classes in the early grades] reduces the distractions in the room and gives the teacher more time to devote to each child.” (1995, p. 125) The knowledge that discipline and classroom management issues have interfered with content instruction is not new: “such problems are less prominent in small classes” (Biddle & Berliner, 2002, p. 15). Biddle & Berliner reported that “small groups can create supportive contexts in which learning is less competitive and students are encouraged to form supportive relationships with one another” (p. 15). The findings by Biddle and Berliner correspond to those found by Johnson (1990), who reported an increased psychological sense of community (PSOC) in small classes. These findings regarding how small classes can create a supportive context for learning and an increased sense of community are important when taking into consideration the Carnegie Council on Adolescent Development 1995 report, which showed “changes in the structure and cohesiveness of families and communities have left many children with fewer positive social supports, less adult guidance and fewer sources of positive role models” (Carnegie Council on Adolescent Development, 1995). Sarason (1974) coined the term *psychological sense of community* (PSOC) to describe the fundamental psychological need all humans have for being part of a community. Goodenow (1993) examined the psychological sense of membership in the classroom and found it to be correlated with student's academic self-efficacy and expectations of success. Participation and engagement in the school and classroom at all levels was identified by Finn and Rock (1997) as the single most important antecedent of at-risk behavior and academic failure even when controlling for SES, race, and ethnicity. Batemen (2002) noted that "the key to feelings of belonging in a community is the level of personal investment that individuals make in the community process" (p. 70). Small classes have been associated with increased opportunities for student collaboration, which fosters the individual investment necessary to create community. "Learning community classrooms are such that, students' learning needs are facilitated and enriched by their teachers and peers" (p. 72). Research previously cited in this review shows that students' individual needs are better met in small classes. Further, students in small classes have a greater opportunity to contribute knowledge to the common goals of the class, helping them become valued members of the classroom community. Bateman found that "research indicates students in small classrooms [*sic*] report lower levels of antisocial behavior and higher affective evaluations of their peers" (p. 73).

STAR results show clearly that there are specific steps for successfully implementing class-size:

1. <u>Early Intervention.</u> Start when the pupil enters "schooling" in K or even pre-K.
2. <u>Sufficient Duration.</u> Maintain the small-class environment for at least 3, preferably 4, years for enduring effects. Encourage parent involvement in schooling.

3. Intense Treatment. The pupil spends all day, every day in the small class. Avoid Pupil-Teacher Ratio (PTR) events, such as "pull-out" projects or team teaching. Develop a sense of "community," close student-teacher relations and coherence. Teacher aides may be used in the site but not for teaching.
4. Use Random Assignment in early grades to facilitate peer tutoring, problem-solving groups, student-to-student cooperation, and active participation and engagement. (STAR).
5. Employ a Cohort Model for several years so students develop a sense of family or community. STAR results show the power of both random assignment and a cohort model. "Looping" adds teacher continuity to the cohort, and may be a useful strategy for added benefits. (Research is needed here).
6. Evaluate process and outcomes carefully, and share results. (Achilles, 2005, p. 15)

Classroom Size and Student Space

Few researchers have looked into the density of occupants in classrooms. The size of the classroom and the amount of furniture within it directly affect one's perception of being crowded. Weinstein (1979) stated, "Nowhere else are large groups of individuals packed so closely together for so many hours, yet expected to perform at peak efficiency on difficult learning tasks and to interact harmoniously" (p. 585). Tanner (2000a) researched the problem of how many students should be in a given space, using social-distance research findings that each student needs 49 square feet of space. He stated that given “the recommended size of an elementary classroom in the United States is 900 square feet; a classroom of this size should house only 17 students” (p. 1). Research done by Tanner (2000b) indicated that “if there are too many people in a given space, we usually react negatively. Children react both by withdrawing, physically and socially, and by acting aggressively” (p. 5). Because space within the classroom is limited, only small amounts of learning materials can be used at any one time; thus, space, or the lack of it, restricts the learning opportunities of students. Duncanson (2003) pointed out that “the lack of large spaces that students can self-select to work in forces the teacher to schedule all events in a one-size fits all modality, focus on the delivery of general instruction to all students, and deal with one activity at a time” (p. 4). Crowded classrooms provide students with few opportunities to engage the teacher one on one in meaningful conversation. Conversely, where there are broad areas for student movement and work, students “direct their own learning activities and become independent-minded investigators while working on several inquiry activities at once” (Duncanson, 2003, p. 4).

In a recent study, Finn, Pannozzo, and Achilles, (2003) reviewed research from ten studies on student behaviors in large- and small-class settings. They found that in the Success Starts Small (SSS) program in High Point, North Carolina, “discipline referrals decreased consistently in the two years after small classes were implemented. There was a 26% drop from the first year to the second year and a 50% drop from the second to the third year” (p. 337). A review of teacher responses regarding the most important differences between large and small classes in the California CSR initiative showed that "easier class discipline emerged as the fourth most important difference, with 20% of all teachers listing this in their responses" (Finn, Pannozzo, & Achilles, p. 339).

A brief review of research literature demonstrated the confusion generated by reporting PTR data as if they were CS data. The review also showed that CS researchers did not mix PTR and CS data in reporting their studies and that independent researchers corroborate the findings of CS researchers. Additionally this review showed that students experiencing small classes have fewer disciplinary and classroom management problems because teachers are better able to engage students one on one. The literature identified that successful implementation of small classes begins with early intervention, K or pre-K, kept students in small classes for at least three, and preferably four, years, and ensured that students stayed in the class all day, every day.

This review of literature and research on CS and PTR has led to the theoretic framework presented in Figure 1 (p. 30). The theoretic framework included the several components of CS necessary to produce a positive, enduring effect on student-achievement-outcomes. The review of literature associated with CS research indicated that data should be accurately reported, that CS data and PTR data must have clarity in the definition of terms (clarity of definition). Also, interventions must be implemented as stated in the research. The theoretic framework delineated several steps in the implementation of CS (e.g., random assignment of students to classes, early intervention, sufficient duration, and intensity of treatment), which, if not present, seem to nullify positive enduring student-achievement-outcomes (intervention implementation). Benefits to students attending CS classes were shown to go beyond achievement outcomes to include a more community-oriented learning environment that meets students' psychological sense of community needs, fewer disruptive student behaviors, and more one-on-one student–teacher opportunities.

Summary

Students have no control over the class environment, CS or PTR, in which they spend the day; building administrators and policy makers are responsible for considering the research and data presented in order to provide the most effective learning environment for children. Clarity in the definitions of CS and PTR when used in reported research is critical to this effort. There must be an understanding of what constitutes a properly implemented CS program and that there are characteristics (e.g., early intervention, random assignment of students to classes, sufficient duration, and intensity of treatment) and the positive effects for staff and students that go beyond achievement outcomes.

The literature review set the stage for the study to compare classes of the same or very similar CS and PTR parameters to try to get a clear determination of their actual or differential effects on student achievement. Chapter 3 details the research design and methodology used in the gathering and analysis of data. Chapter 4 presents the analysis and results of these data. Chapter 5 includes the summary of findings, conclusions, and recommendations for practice, for policy, and for further research.

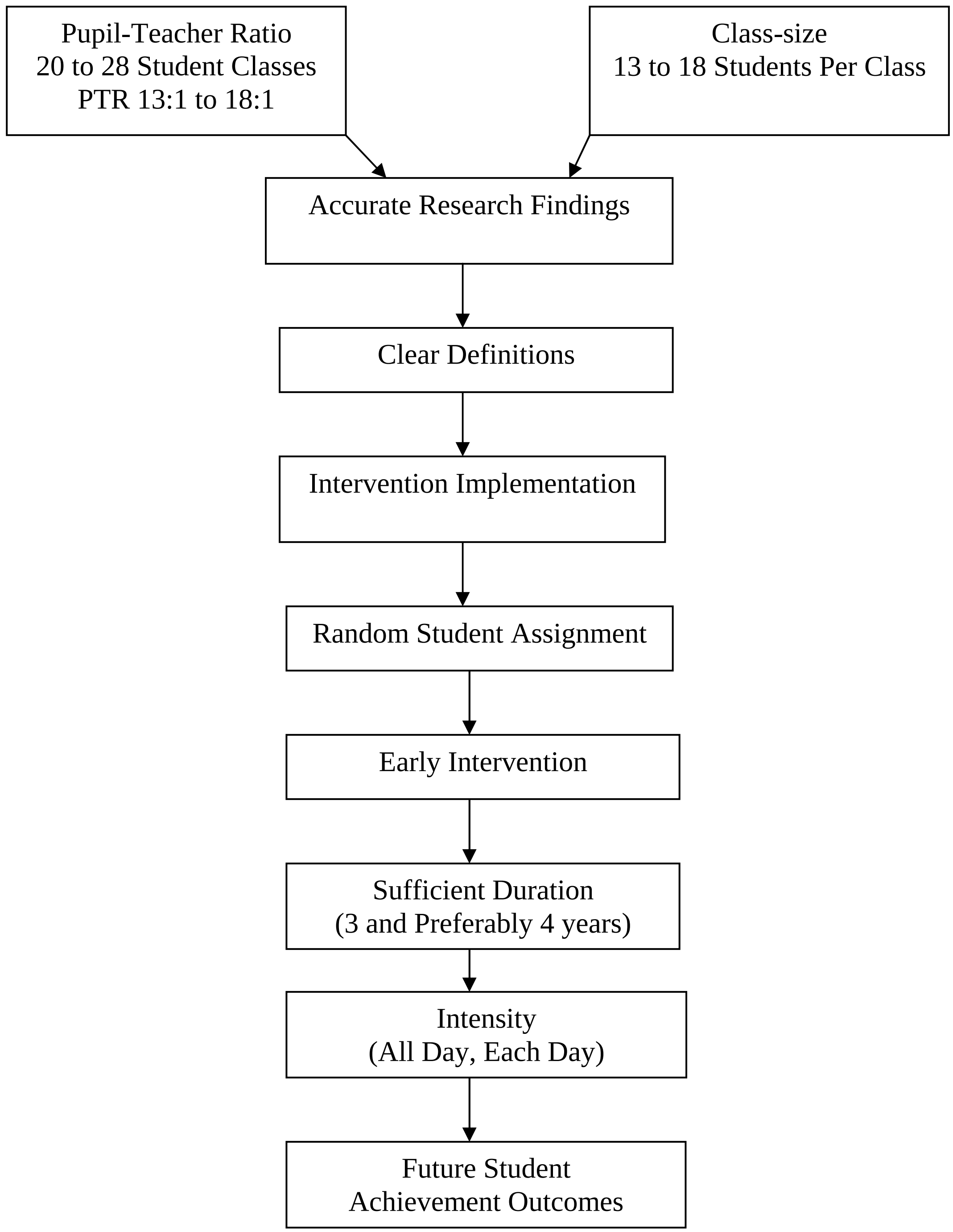

Figure 1. Theoretic model for understanding the influence of class size and pupil-teacher ratio on student-achievement outcomes.

Chapter 3

Research Design and Methodology

School administrators need to investigate opportunities to improve the delivery of education to students. Researchers assist educators in this practice by conducting and analyzing research to ascertain those approaches that improve student engagement and achievement. The study of class-size (CS) and pupil-teacher ratio (PTR) is the approach reviewed in this study.

A review of research and literature related to CS and PTR class settings and their effects demonstrated confusion in definitions of terms. In reporting PTR data as if they were CS data, educators and economists alike use the terms CS and PTR as synonyms, contributing to the confusion and not adding to the clarity surrounding the settings (e.g., Hanushek, 1998; Johnson, 2001).

Design

The present study was a nonexperimental ex-post-facto, "cross-sectional" study (Johnson, 2001, p. 9; Kerlinger, 1986) commonly used when studying human performance in real-world situations. The data analyzed were from student-achievement-outcomes on standardized test instruments for students in third-grade and from standards-referenced tests for fourth-grade students in the two (CS & PTR) conditions. The data supplied are normally collected data available at each school. In this case, participants were third- and fourth-grade students in two different class settings. Those settings were classes of between 13 and 20 students with one teacher, designated as CS, and larger classes of between 24 and 28 students located in schools with a PTR of between 13:1 and 20:1. Third-grade student outcomes from standardized tests, such as the California

Achievement Test (CAT), California Test of Basic Skills (CTBS), Iowa Test of Basic Skills (ITBS), and Terra Nova (TN) were analyzed. Fourth-grade student outcomes from the Michigan Education Assessment Program (MEAP), a criterion-referenced test (SRT), were analyzed. The targeted populations were third- and fourth-grade students attending public schools in Michigan during the 2002-2003 school year. There was no control group in the traditional sense, as the focus was on comparisons of effects of CS and PTR on student achievement. The student populations (n) in the comparisons were generally analogous, that is, data from CS conditions of 13-20 were compared to data from students in larger classes (24-28 students) in a school with a PTR of from 13:1 to 20:1. Factors including school setting, urban, suburban, rural, state-foundation-grant amount, and percentage of free and reduced lunches were as similar as possible. Approval for the study was received from the Eastern Michigan University Human Subjects Review committee on January 30, 2004 (Appendix B), and an approval to extend the protocol for one year was received on January 28, 2005 (Appendix C).

Methodology

To the degree possible, the researcher sought comparability in the two settings, attempting to assure that treatments for both groups were the same, such as assignments to classes and assignment of instructional staff. To the degree attainable in a non-experimental study, the only differences in the groups were the CS and PTR settings.

The student-achievement-outcome data gathered, to the degree possible, were disaggregated by gender, ethnicity, and free- and reduced-lunch classifications as a proxy for socio-economic status (SES). The comparison of like classes in similar schools within comparable districts was essential to the validity of this study. The standardized test scores for each class were entered into SPSS statistics software for analysis. Student-achievement-outcome data were the normally collected data drawn from student testing and available in school archives. No specific testing was done for the study, in which the researcher used available data on achievement. Data were analyzed and reported only as group data (CS or PTR), and to ensure confidentiality, no student, teacher, or school was named. Test scores for each third-grade class setting were converted into z-scores. Mean scores (MS) and standard deviation (SD) information (Appendix D) were received for the Iowa Test of Basic Skills and Terra Nova tests; similar information was not received for the Gates–MacGinitie reading test, making those scores unusable. Z-scores were created for third-grade student-achievement-outcome scores by subtracting the mean score (MS) for each test from the actual student score (AS) and dividing the difference by the standard deviation for the specific test [(AS-MS)/SD]. Z-scores were used for cross-test comparison because “these scores are comparable since they are standardized in units of standard deviations” (Salkind, 2000, p. 155).

The preponderance of literature demonstrating the positive impact of CS on student achievement suggested that a one-tailed *t* test could be used to analyze the gathered data. However, because the study focused on data and on theory, not on literature, a two-tailed *t* test was used in data analysis of the null hypothesis, that there was no difference in student-achievement-outcomes related to classroom settings. This analysis approach should determine whether either treatment, CS or PTR, has a significant impact on student achievement. The researcher used two-tailed *t* tests to compare the effects of CS and PTR on student achievement by gender, ethnicity, and poverty levels.

In addition to statistical analyses, the researcher investigated like environmental factors in each CS and PTR class setting, specifically including such items as square feet per student (ft2) and circulation space. The ft2 of each classroom, CS and PTR, was calculated and divided by the number of students within each class to determine the square feet per student (ft2/student). If a greater effect were found for either CS or PTR, the impact of how much room each student has might also be a factor in positive student achievement. Similarly, through simple observation in selected classrooms, the numbers of times individual students questioned their teachers regarding the repeating or clarifying of instructions or content were recorded. The numbers of disturbances, student-to-student or individual student, were recorded, as such events may help explain differences and similarities between treatments and whether the size of a classroom, as well as CS, influences student behavior.

The observational research was conducted with the researcher as a non-participant, as defined by Lofland and Lofland (1995). The goal was to gather the best possible data by observation to aid the researcher's understanding of why CS or PTR has a positive effect on student achievement. Data were recorded by the non-participant researcher; student behavior was observed in both CS and PTR settings, ensuring the quality of the data.

Site Selection

Utilizing data detailing the per-pupil state foundation grant for each school district for fiscal years 2002, 2003, and 2004 provided by the Michigan House Fiscal Agency, 102 school districts that received a per-pupil state foundation grant of $6,500.00 were purposefully selected for possible participation in the study. School districts located in

Wayne County, Michigan, were excluded because this urban district has the highest population of minority students and was not comparable to any other school districts across the State. The randomization of selection was aided by using the House Fiscal Agency data because no demographical information was contained within the report. Following initial selection of the 102 school districts, a letter asking for participation in the study was sent to the superintendent of each district selected, allowing the researcher to use standardized test scores from one third- and one fourth-grade classroom (Appendix E). Because there was a slow rate of return, a follow-up letter was sent to superintendents on April 16 (Appendix F). By the second week in July 2004, 20 school districts had given initial permission to participate in the study; those districts are represented in Table 2. The responding districts ranged in size from 11,800 students to 166 students, with an average enrollment (less the 11,800 and 166 student districts) of 44,447/18, or 2,339 students. The percentage of students in the identified schools that participated in this study who qualified, under the Federal Free and Reduced Lunch program, to receive free lunch, ranged from 70% to 4% percent, with an average enrollment (less the 70% and 4%) of 533/18, or 30%. Ten of the responding districts had reported enrollment between 4,900 and 1,593 students, and 17 of the 20 districts reported per-student foundation grant amounts between $6,700 and $6,626. The amount of per-student foundation grant amounts received from the responding districts ranged from $7,000 to $6,616. Individual school-building enrollment ranged from 1,850 students to 166 students, with an average enrollment (less the 1850 and 166) of 8,815/18, or 429 students. The data presented in Table 2 demonstrate the difficulty in collecting CS data, as of the 20 responding districts, only 4 indicated classes that met CS parameters. One of the four withdrew before the study was completed.

Data Gathering

Because participation was voluntary, the respondents dictated the selection of those schools whose data would be used as the CS representatives. Of the 20 school districts that responded, only 3 had class-sizes for third and fourth-grades that fell within the criteria set for CS in this study. Four respondents were excluded because they did not have standardized testing at the third-grade level. Of the remaining 16 school districts, three schools were identified to gather the PTR student-achievement-outcome scores from because their building data closely matched each other's. Once identified, a letter was sent to the principal of each elementary school selected by the district superintendent as a participating school (Appendix G). Each letter sent was followed by a phone call to the principals of the participating schools two weeks prior to the start of the 2004/2005 school year. At this time several schools withdrew their support for participation in the study; the most crucial was one of three CS schools, and the building principal stated reluctance to release student test data. The researcher called the superintendent of the particular district and expressed the importance of that district's staying in the study. The superintendent subsequently made arrangements for the necessary standardized test data to be shared. This ensured that the necessary (and minimum) standardized test data from three schools representing both CS and PTR settings in both the third and fourth grades would be available. At this point, the data for the study included descriptive data on the three school districts for each CS and PTR setting; these data are in Table 3.

Table 2

Initial Responding School District Descriptive Data Comparison

District enrollment	District certified staff	Foundation grant	Building enrollment	Classroom teachers	Special education teachers	Support teachers	Building administrators	Pupil-teacher ratio	% Free lunch students	% Reduced lunch students	Third-grade students	Fourth-grade students
N	N	N	N	N	N	N	N	N	N	N	N	N
2600	171	6700	443	20	8	3	2	14.3	28	6	19	20
3515	198	6700	798	28	6	8	3	19	31	9	0	26
792	54	6700	400	20	3	6	2	13.8	50	10	22	17
310	22	6700	310	13.5	3	1.5	1.5	17.2	20	5	19	20
11800	802	6700	450	21	4	5	4	15	70	15	14	19
605	45	6626	320	14	1	110	1.5	12.8	12	18	25	20
2007	123	6626	650	28	5	6	3	16.7	31	9	23	28
1321	86	6626	401	18	5	3.2	3.2	15.3	44	13	0	0
3320	0	6700	406	16	2.4	1.2	1	20.7	13.5	1	25	27
1573	103	6700	430	16	5	7	3	15.4	36	0	24	31
2304	122	6626	504	21	4	2	2	18.7	32	10	26	24
2032	115	6616	214	26	3	1	1.3	7.13	26	5	29	26
2300	130	6700	515	21	4	7	1	16.1	33	6	24	23
1079	97	6700	260	10	1	1	1	21.7	42	11	21	20
4977	277	6700	472	17.5	1.5	4	2	20.5	28.8	9.7	28	25
322	22	6700	322	15	5	2	2	14.6	58	17	23	24
3950	272	6762	1850	94	6	16	8.5	16	11	10	23	24
4900		6800	320	12	4	2	1.3	17.8	4	2	28	27
4201	232	6700	506	22	5	1	1.5	18.1	12.3	18.1	23	25
166	19	7000	166	18	0	1	1	8.73	24	10	12	13

The CS schools had an average enrollment of 353 students, and the PTR schools showed an average enrollment of 509 students. The schools representing CS in the study had an average CS of 15 students per class in third-grades and 17 in fourth grades, whereas the PTR schools averaged 25 third-grade students and 27 in fourth grades.

The standardized test data were identified as fourth-grade-student test scores from the Michigan Education Assessment Program (MEAP) test, a criterion-referenced test and the student test scores from any standardized test regularly given in third-grade. To make gathering the data as easy as possible for the person in each school, choices in the method used to provide the necessary information were offered: (a) a disk formatted with an Excel spreadsheet was provided, wherein a school person could enter the test data in the appropriate columns; (b) schools could print or copy the information and send it to the researcher who would enter the data; or (c) if the information were already in electronic media, the person could email the data. The complete third-grade CS data are in Appendix H; complete third-grade PTR data are in Appendix I; complete fourth-grade CS data are in Appendix J; and complete fourth-grade PTR data are in Appendix K.

Data on student behavior and classroom size from PTR and CS conditions were gathered from observations in six classroom settings, three each of CS and PTR. Each observation was 90 minutes long. Classrooms were randomly selected from third- and fourth-grade classes that met CS and PTR criteria. The observer had no interaction with the students or teacher. Arrangements were made with teachers prior to observation to coordinate the schedule of the observer to ensure that observations were made approximately within the same time frame during the school day and that there was a

Table 3

District Descriptive Data Comparison of Participating Class-size and Pupil-Teacher Ratio Sites

	District enrollment	District certified Staff	Foundation grant	Building enrollment	Classroom teachers	Special education teachers	Support teachers	Building administrators	Pupil-teacher ratio	% Free lunch students	% Reduced lunch students	Third-grade students	Fourth-grade students
	N	N	N	N	N	N	N	N	N	N	N	N	N
Class size	11800	802	6700	450	21	4	5	4	15	70	15	14	19
	2600	171	6700	443	20	8	3	2	14.3	28	6	19	20
	166	19	7000	166	18	0	1	1	8.73	24	10	12	13
PTR	4977	277	6700	472	17.5	1.5	4	2	20.5	28.8	9.7	28	25
	3320	0	6700	406	16	2.4	1.2	1	20.7	13.5	1	25	27
	2007	123	6626	650	28	5	6	3	16.7	31	9	23	28

location where the observer was out of the way of instructional needs and student sight lines. The observer entered the classrooms and proceeded directly to the designated observation location. Information about the classroom setting (e.g., size of the classroom, number of students, number of adults, and type of furniture) was recorded following the 90-minute observation period. Observations were recorded for Interruptions of Questions and Interruptions of Student Behaviors (see Table 9). The numbers of observed interruptions were totaled following the observation. Questions of content and clarification were recorded because they represent students' not concentrating on directions given or materials presented by the teacher. Student behavior was recorded as incidents of individual students and multiple-student disruptions. Distractions by individual or groups of students create distraction in the classroom environment, keeping students from concentrating fully.

Information was also gathered concerning the size of each classroom and the amount of furniture within each room (teacher, student, and ancillary) for each CS and PTR classroom to determine the amount of space available for student learning and movement. The CS-PTR observational tool is Appendix L.

Data were sought from each of the selected CS schools to determine the level of implementation regarding those characteristics identified by Achilles (2005) as key for small classes to have an enduring effect on student achievement. Data were sought on whether or not students began their education in small classes (early intervention), how many concurrent years students attended small classes (sufficient duration), whether students spent all day, every day, in a small class and/or if there additional pull-out interventions that removed students from their classroom for part of the day (intense treatment), and whether students were randomly assigned to classes (random assignment) (see Table 10).

Data on student achievement were gathered from six CS and PTR classroom settings within the selected schools, three CS and three PTR. These settings were as similar as possible in the areas of per-student foundation grant amount, percentage of students participating in the federal free and reduced lunch program, and school setting—urban, suburban, or rural. Data were extracted from standardized test scores used in each district and extracted from nonparticipant observation in CS and PTR classrooms.

Analysis of these data was accomplished by using SPSS 10.0 statistical analysis software. Third-grade student-achievement-outcome data were converted to z-scores because "these scores are comparable since they are standardized in units of standard deviations" (Salkind, 2000, p. 155). A two-tailed test was used to determine whether or not there were differences in student achievement as a result of either the CS or PTR classroom settings. CS and PTR classroom settings were observed, and data were recorded on the size of each classroom, amount of furniture, and circulation space. Information was solicited from participant schools to determine the implementation of small classes, the grade levels at which students were in small classes, the length of time spent in small classes each day, and the existence of pull-out programs.

Chapter 3 has included information on the study design and research methods. Chapter 4 presents the results and analysis of the data. Chapter 5 includes the summary of findings, conclusions, and recommendations for practice, for policy, and for further research.

Chapter 4

Presentation and Analysis of Data for PTR/CS Study

Introduction

The purpose of the study was to determine if there was a difference in student achievement based on students' being in a small class (CS) (13-17) or a large class (20-28) with a pupil-teacher ratio (PTR) between 13:1 and 17:1. It was not the purpose of the researcher to estimate or determine the impact of a particular setting. The study consisted of gathering and analyzing student-achievement-outcome data from standardized tests taken by students in their particular CS or PTR setting and gathering student-behavioral data from CS and PTR classrooms observed by the researcher.

Third-grade student-achievement-outcome data were received from a variety of standardized testing instruments: the Iowa Test of Basic Skills (ITBS), Northwest Evaluation Association (NWEA), Gates McGinitiy (GM) reading test, and Terra Nova Test (TN). Data displayed in Table 4 demonstrate the number of student-achievement-outcomes supplied in each standardized testing format, which exemplified one of the difficulties of the study, as one PTR site that supplied student-outcome data reported achievement data in only reading and not in mathematics and reading, as requested.

Fourth-grade student-achievement-outcome data were received from schools in two standardized tests, the ITBS and the Michigan Educational Assessment Program (MEAP). These data, presented in Table 5, show that one of the CS schools reported student-outcome-achievement data from the ITBS in reading only.

Descriptive/Demographic Data for Student Achievement

The demographic data, shown in Table 6, describe the third- and fourth-grade

Table 4

The Number of Third-Grade Student-Achievement-Outcome Data Sets Received by Standardized Testing Format, CS and PTR Study

CS Scores (n = 44)	Reading	Mathematics
ITBS	32	32
NWEA	12	12
PTR Scores (n = 73)		
GM	25	
TN	48	48

Table 5

The Number of Fourth-Grade Student-Achievement-Outcome Data Sets Received by Standardized Testing Format, CS and PTR Study

CS Scores (n = 49 and 32)	Reading	Mathematics
ITBS	17	0
MEAP	32	32
PTR Scores (n = 76)		
MEAP	76	76

students from whom the student-achievement-outcome data were received. The number of male and female students at each grade level was evenly distributed between CS and PTR classes. The ethnicity of students in each grade level from whom achievement-outcome data were received were predominantly White children, with only 24 students of 242, or 9.9%, classified as Native American, Asian, Black, or Hispanic. Table 5 also demonstrates a wide spread in the number of students participating in the National School Free and Reduced Lunch Program among those schools reporting PTR data.

The total student-achievement-outcome data received from participating CS and PTR schools is found in Appendices H, I, J, and K. Table 6 portrays the demographic information of the students whose scores were used in the study. Table 6 entries demonstrate one difficulty in analyzing these data between settings because student demographic data were not provided from some sites for all requested categories. The total scores for CS (n = 44) and PTR (n = 48) reflect the difference in mean z-scores between the two settings.

Table 6

Demographic Information of Students Who Participated in the CS/PTR Study

	Gender				Ethnicity				Lunch Status			
Grade level/demographics	Male		Female		White		Other		Regular		Free/Reduced	
	N	%	N	%	N	%	N	%	N	%	N	%
Third-grade (CS n = 44)	24	55	20	45	38	8	6	14	28	64	16	36
Third-grade (PTR n = 73)	37	51	36	49	25	34	0	0	60	82	13	18
Fourth-grade (CS n = 49)	25	51	24	49	37	75	12	24	27	55	22	45
Fourth-grade (PTR n = 76)	41	54	35	46	70	92	6	8	52	68	14	18

Test-Score Outcomes, Grades 3 and 4

Table 7 displays mean z-scores for total CS (n = 44) and PTR (n = 48) third-grade student-achievement-outcome data received for reading and mathematics and by the category (lunch status, ethnicity, or gender) and setting (CS & PTR). Z-scores were used for cross-test comparison because “these scores are comparable since they are standardized in units of standard deviations” (Salkind, 2000, p. 155). Z-scores were created for third-grade student-achievement-outcome scores by subtracting the mean score (MS) for each test from the actual student score (AS) and dividing the difference by the standard deviation (SD) for the specific test [(AS – MS)/SD]. Table 7 also demonstrates the difficulty in analyzing the data between settings because all student demographic data were not provided from sites for each requested category. The scores for CS (n = 44) and PTR (n = 48) reflect the difference in mean scores between settings.

Table 7

Third-Grade Mathematics and Reading Mean Z-Scores by Setting (CS & PTR) and Student Category (Lunch Status, Ethnicity, & Gender)

		Mean Z-scores		
CS	N	Reading	N	Mathematics
Total	44	-0.19402	44	-0.15058
Regular	28	-0.02444	28	.110436
Free/Reduced	16	-0.49078	16	-0.60734
White	38	-0.13089	38	-0.18473
Other	6	-0.17638	6	-0.38646
Male	24	-0.15225	24	-0.27277
Female	20	-0.24414	20	-0.00394
PTR				
Total	48	-0.1399	48	1.2642
Regular	37	-0.14401	37	1.2354
Free/Reduced	11	-0.12608	11	1.3609
White	0	0	0	0
Other	0	0	0	0
Male	26	-0.08235	26	1.1649
Female	22	-0.20792	22	1.3815

Table 8 shows mean scores for total CS (n = 49) and PTR (n = 76) fourth-grade student-achievement-outcome data received in reading and mathematics, as well as mean scores for the categories of lunch status, ethnicity, and gender.

Table 8

Fourth-Grade Reading and Mathematics Mean Scores by Setting (CS & PTR) and Category (Lunch Status, Ethnicity, & Gender)

		Mean scores		
CS	N	Mathematics	N	Reading
Total	49	540.34	49	547.50
Regular	19	544.05	27	550.42
Free/Reduced	13	534.92	22	543.23
White	27	535.80	37	542.20
Other	5	541.18	12	548.48
Male	16	531.44	25	539.56
Female	16	549.25	24	555.44
PTR				
Total	76	555.33	76	554.64
Regular	34	565.88	34	561.85
Free/Reduced	14	547.85	14	553.36
White	70	555.99	70	555.41
Other	6	547.66	6	545.66
Male	41	556.85	41	556.07
Female	35	553.54	35	552.97

Each individual student received a student number to differentiate him/her by grade level and class setting. The reading and math student-achievement-outcome scores represent different standardized tests. These data needed to be equalized for comparison. Following the translation of the raw third-grade data into z-scores, data were analyzed using a two-tailed *t* test because for this study the researcher utilized the nondirectional null hypothesis that there would be no difference between the achievement outcomes of students in small classes and the outcomes of students in larger classes housed in a facility with a PTR similar to the number of students in the small class. The two-tailed *t* test was used to determine whether or not there was a difference in the achievement level of students in CS and PTR classes but not the particular direction of the difference. A .95 confidence interval was used to compute the two-tailed *t* tests. Referencing "Table B.2 T values Needed for Rejection of the Null Hypothesis" (Salkind, p. 335), the *t* value needed to reject the null hypothesis had be greater than 1.96 ($p \leq .05$).

Data were gathered regarding student behaviors, the size of individual CS and PTR classrooms, and the available space to facilitate student learning as the researcher observed three CS and three PTR settings for 90 minutes each. During the observations, the researcher recorded data regarding how often students asked questions regarding content or instruction, as well as how often individual students or groups of students exhibited behaviors that disrupted the decorum of the class. The form used to record the observational data is shown in Appendix L. Observation data were then totaled for each observed CS or PTR classroom so that comparisons between the settings could be made.

Data were gathered from participating CS sites to determine the extent to which those characteristics of SC that had been demonstrated by research (e.g., Ramey &

Ramey, 1989; Achilles, 2002) to promote positive and enduring effects on student achievement (Table 10) were implemented. Data were gathered on early intervention by determining the availability of SC in pre-kindergarten (pre-K) or kindergarten (K). Data were gathered regarding the number of consecutive years students spent in small classes to determine the duration that students were able to spend in SC. Data were gathered to determine how CS sites assigned students to classes (randomly or assigned) and if pull-out interventions were used to improve student achievement.

Analysis of Student Outcome Data by Variables (Tabled Data are in Appendices)

Analysis of third-grade reading student-achievement-outcome z-scores, fully displayed in Appendix M, showed that the achieved t value of -.316 was less than the t value 1.987 necessary for rejection of the null hypothesis, based on 90 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). Additionally the achieved significance value of .753 was greater than $p \leq .05$, based on 90 degrees of freedom found in “Table B” (p. 335) which means it was “too large to reject the null hypothesis” (Norusis, 2000, p. 244). These results determined there was no significant difference in student achievement between students in CS or PTR classes, which determination led to the acceptance of the null hypothesis. The result indicated the difference in reading z-scores of students in third-grade did not occur by something other than chance.

Analysis of third-grade mathematics student-achievement-outcome data z-scores (fully displayed in Appendix N) showed that the achieved *t* value of -6.866 was greater than that of the 1.987 necessary for rejection of the null hypothesis, based on 90 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). The achieved significance value of 0.00 was less than the $p \leq .05$ necessary for the rejection of the null hypothesis based on 90 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). According to Norusis (2000), “when the significance level is small, you reject the null hypothesis” (p. 244). The analysis determined that there was a significant difference in student-achievement-outcome z-scores between students in CS and PTR classes, which supported the rejection of the null hypothesis. This indicates that the difference in the mathematics z-scores between third-grade students in CS and PTR classes did occur by something other than chance.

Analysis of third-grade reading student-achievement-outcome z-scores for female students (fully displayed in Appendix O) showed that the achieved *t* value of -.153 was less than the 2.021 necessary for rejection of the null hypothesis, based on 40 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). The obtained significance value of .879 was greater than $p \leq .05$, “too large to reject the null hypothesis” (Norusis, 2000, p. 244). These results demonstrate there was no significant difference in the student-achievement-outcome z-scores for reading between female students in third-grade CS and PTR classroom settings.

Analysis of third-grade mathematics student-achievement-outcome z-scores for female students (fully displayed in Appendix P) showed that the achieved *t* value of -4.489 was greater than the 1.987 necessary for rejection of the null hypothesis. The obtained significance value of .000 was less than the $p \leq .05$ necessary for the rejection of the null hypothesis. According to Norusis (2000), “when the significance level is small, you reject the null hypothesis” (p. 244). These results demonstrate there was a significant difference in the student-achievement-outcome z-scores for reading between female students in third-grade CS and PTR classroom settings and that the difference occurred by something other than chance.

Analysis of third-grade reading student-achievement-outcome z-scores for male students (fully displayed in Appendix Q) showed that the achieved *t* value of -.282 was less than the 2.014 necessary for rejection of the null hypothesis, based on 48 degrees of freedom found in "Table B" (Salkind, 2000, p. 335). The obtained significance value of .779 was greater than $p \leq .05$, "too large to reject the null hypothesis" (Norusis, 2000, p. 244). These results demonstrate there was not a significant difference in the student-achievement-outcome z-scores for reading between male students in third-grade CS and PTR classroom settings.

Analysis of third-grade mathematics student-achievement-outcome z-scores for male students (fully displayed in Appendix R) showed that the achieved *t* value of -3.152 was greater than the 2.014 necessary for rejection of the null hypothesis, based on 48 degrees of freedom found in "Table B" (Salkind, 2000, p. 335). The obtained significance value of .003 was less than the $p \leq .05$ necessary for the rejection of the null hypothesis. According to Norusis (2000), "when the significance level is small, you reject the null hypothesis" (p. 244). These results demonstrate there was a significant difference in the student-achievement-outcome scores for reading between male students in third-grade CS and PTR classroom settings and that the difference occurred by something other than chance.

Analysis of third-grade reading student-achievement-outcome z-scores for students who purchased regular lunch (fully displayed in Appendix S) showed that the achieved *t* value of .550 was less than the 2.001 necessary for rejection of the null hypothesis, based on 63 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). The obtained significance value of .584 was greater than $p \leq .05$, “too large to reject the null hypothesis” (Norusis, 2000, p. 244). These results demonstrate there was not a significant difference in the student-achievement-outcome z-scores for reading between students who purchased regular lunch in third-grade CS and PTR classroom settings.

Analysis of third-grade mathematics student-achievement-outcome z-scores for students who purchased regular lunch (fully displayed in Appendix T) showed that the achieved *t* value of -4.421 was greater than the 2.001 necessary for rejection of the null hypothesis, based on 63 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). The obtained significance value of .000 was less than the $p \leq .05$ necessary for the rejection of the null hypothesis. According to Norusis (2000) “when the significance level is small, you reject the null hypothesis” (p. 244). These results demonstrate there was a significant difference in the student-achievement-outcome z-scores for mathematics between students who purchased regular lunch in third-grade CS and PTR classroom settings and that the difference in z-scores occurred by something other than chance.

Analysis of third-grade reading student-achievement-outcome z-scores for students who participated in the National School Free and Reduced Lunch Program (fully displayed in Appendix U) showed that an achieved *t* value of .-1.405 was less than the 2.060 necessary for rejection of the null hypothesis, based on 25 degrees of freedom found in “Table B” (Salkind, 2000, p. 335) and an obtained significance value of .172 was greater than $p \leq .05$, “too large to reject the null hypothesis” (Norusis, 2000, p. 244). These results demonstrate there was not a significant difference in the student-achievement-outcome scores for mathematics between students who participated in the National School Free and Reduced Lunch Program in third-grade CS and PTR classroom settings.

Analysis of third-grade mathematics student-achievement-outcome data for students who participated in the National School Free and Reduced Lunch Program (fully displayed in Appendix V) showed that an achieved *t* value of -6.038 was greater than the 2.001 necessary for rejection of the null hypothesis, based on 25 degrees of freedom found in "Table B" (Salkind, 2000, p. 335). The obtained significance value of .000 was less than the $p \leq .05$ necessary for the rejection of the null hypothesis. According to Norusis (2000) "when the significance level is small, you reject the null hypothesis" (p. 244). These results demonstrate that there was a significant difference in the student-achievement-outcome scores for mathematics between students who participated in the National School Free and Reduced Lunch Program in third-grade CS and PTR classroom setting and that the difference in z-scores occurred by something other than chance.

Analysis of fourth-grade reading student-achievement-outcome data (fully displayed in Appendix W) showed that the achieved *t* value of -2.811 was greater than the 1.96 necessary for rejection of the null hypothesis, based on 106 degrees of freedom found in "Table B" (Salkind, 2000, p. 335). Additionally, the achieved significance value of 0.006 was less than $p \leq .05$, and according to Norusis (2000), "when the significance level is small, you reject the null hypothesis" (p. 244). These results determined that there was a significant difference in student-achievement-outcome data between students in CS and PTR classes, which supported the rejection of the null hypothesis. This indicated that the difference in scores of fourth-grade student reading tests in the two settings did occur by something other than chance.

Analysis of fourth-grade mathematics student-achievement-outcome data (fully displayed in Appendix X) showed an achieved *t* value of -1.099 that was less than the 1.96 necessary for rejection of the null hypothesis, based on 106 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). The achieved significance value of 0.274 was greater than $p \leq .05$, “too large to reject the null hypothesis” (Norusis, 2000, p. 244). These results showed that there was no significant difference in student achievement between students in CS and PTR classes, which determination led to acceptance of the null hypothesis. The difference in z-scores of students in fourth-grade mathematics in the two settings did not occur by anything other than chance.

Analysis of fourth-grade mathematics student-achievement-outcome data for female students (fully displayed in Appendix Y) showed that the achieved *t* value of -.263 was less than the 2.009 necessary for rejection of the null hypothesis, based on 49 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). The obtained significance value of .793 was greater than $p \leq .05$, “too large to reject the null hypothesis” (Norusis, 2000, p. 244). These results demonstrate that there was no significant difference in the student-achievement-outcome scores for reading between female students in fourth-grade CS and PTR classroom settings and that the difference in scores did not occur by anything other than chance.

Analysis of fourth-grade reading student-achievement-outcome data for female students (fully displayed in Appendix Z) showed an achieved *t* value of -.700 that was less than the 2.009 necessary for rejection of the null hypothesis, based on 49 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). The obtained significance value of

.487 was greater than $p \leq .05$, "too large to reject the null hypothesis" (Norusis, 2000, p. 244). These results demonstrate that there was no significant difference in the student-achievement-outcome scores for reading between female students in fourth-grade CS and PTR classroom settings and that the difference did not occur by anything other than chance.

Analysis of fourth-grade mathematics student-achievement-outcome data for male students (fully displayed in Appendix AA) showed an achieved *t* value of -1.310 that was less than the 2.004 necessary for rejection of the null hypothesis, based on 55 degrees of freedom found in "Table B" (Salkind, 2000, p. 335). The obtained significance value of .196 was greater than $p \leq .05$, "too large to reject the null hypothesis" (Norusis, 2000, p. 244). These results demonstrate that there was no significant difference in the student-achievement-outcome scores for reading between male students in fourth-grade CS and PTR classroom settings and that the difference did not occur by anything other than chance.

Analysis of fourth-grade reading student-achievement-outcome data for male students (fully displayed in Appendix AB) showed that the achieved *t* value of -3.152 was greater than the 2.004 necessary for rejection of the null hypothesis, based on 55 degrees of freedom found in "Table B" (Salkind, 2000, p. 335). The obtained significance value of .003 was less than $p \leq .05$, and according to Norusis (2000), "when the significance level is small, you reject the null hypothesis" p. 244). These results demonstrate that there was a significant difference in the student-achievement-outcome scores for reading between male students in fourth-grade CS and PTR classroom settings, and that the difference occurred by something other than chance.

Analysis of fourth-grade mathematics student-achievement-outcome data for students who purchased regular lunch (fully displayed in Appendix AC) showed that the achieved *t* value of .072 was less than the 2.009 necessary for rejection of the null hypothesis, based on 51 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). The obtained significance value of .943 was greater than $p \leq .05$, “too large to reject the null hypothesis” (Norusis, 2000, p. 244). These results demonstrate that there was no significant difference in the student-achievement-outcome scores for mathematics between students who purchased regular lunch in fourth-grade CS and PTR classroom settings and that the difference in scores did not occur by anything other than chance.

Analysis of fourth-grade reading student-achievement-outcome data for students who purchased regular lunch (fully displayed in Appendix AD) showed that the achieved *t* value of -2.162 was greater than the 2.009 necessary for rejection of the null hypothesis, based on 51 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). The obtained significance value of .035 was less than $p \leq .05$, and according to Norusis (2000) “when the significance level is small, you reject the null hypothesis” (p. 244). These results demonstrate there was a significant difference in the student-achievement-outcome scores for reading between students who purchased regular lunch in fourth-grade CS and PTR classroom settings and that the difference in scores occurred by something other than chance.

Analysis of fourth-grade mathematics student-achievement-outcome data for students who participated in the National School Free and Reduced Lunch Program (fully displayed in Appendix AE) showed that the achieved *t* value of -2.289 was greater than the 2.060 necessary for rejection of the null hypothesis, based on 25 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). The obtained significance value of .031 was less than $p \leq .05$, and according to Norusis (2000) “when the significance level is small, you reject the null hypothesis” (p. 244). These results demonstrate that there was a significant difference in the student-achievement-outcome scores for mathematics between students who participated in the National School Free and Reduced Lunch Program in fourth-grade CS and PTR classroom settings and that the difference in scores occurred by something other than chance.

Analysis of fourth-grade reading student-achievement-outcome data for students who participated in the National School Free and Reduced Lunch Program (fully displayed in Appendix AF) showed that the achieved *t* value of -2.787 was greater than the 2.060 necessary for rejection of the null hypothesis, based on 25 degrees of freedom found in “Table B” (Salkind, 2000, p. 335). The obtained significance value of .010 was less than $p \leq .05$, and according to Norusis (2000) “when the significance level is small, you reject the null hypothesis” (p. 244). These results demonstrate that there was a significant difference in the student-achievement-outcome scores for mathematics between students who participated in the National School Free and Reduced Lunch Program in fourth-grade CS and PTR classroom settings and that the difference in scores occurred by something other than chance.

Observational Data

The observational data, displayed in Table 9, showed that students attending PTR classes had from 2.5 to 4.33 times as many disruptions as did those attending CS classrooms. The disruptions were caused by students’ asking questions to clarify instructions or understand content or because of student-behavior issues. Students in CS

Table 9

Results Obtained from Observing CS and PTR Class Settings

Variables	CS	CS	CS	PTR	PTR	PTR
	Class setting					
Students (n)	14	18	17	26	25	23
Teachers (n)	1	1	1	1	1	1
Teacher aides (n)	0	0	0	1.5	1	1
Student desks (n)	22	20	24	28	28	29
Teacher desks (n)	2	1	1	1	2	1
Other desks/tables (n)	1	2	1	1	1	2
Classroom ft2	1008	952	896	720	900	960
Circulation ft2	364	252	284	152	212	238
ft2 per student	72.00	52.89	52.71	27.69	36.00	41.74
Duration of observation (min)	90	90	90	90	90	90
Instruction questions (n)	4	8	11	32	27	31
Content questions (n)	4	7	6	20	18	23
Interruption of single student (n)	7	19	12	28	21	24
Interruption student to student (n)	24	16	27	87	79	83
Total interruptions & questions (n)	39	50	56	167	145	161

classes had 1.25 to 2.6 times more square footage per student than those students in PTR settings; this provided the students in CS settings a greater amount of circulation area than their PTR counterparts experienced.

Implementation Data

Data gathered on the degree of implementation of those characteristics, identified by research (e.g., Ramey & Ramey, 1998; Achilles, 2002), which contributed to the enduring effects of CS on student achievement are displayed in Table 10. These data demonstrated that the characteristics necessary for the enduring effects of CS on student achievement were lacking in two of the three CS settings. Two CS classes continued to use pull-out interventions, removing students from their classrooms for remediation treatments, ensuring that students missed what was taught while they were out of the classroom.

Table 10

Implementation of the Characteristics of Small Class Size Necessary to Exert Positive, Enduring Effects on Student Achievement

				Total	
	CS-1	CS-2	CS-3	Yes	No
Early intervention					
Small classes available in pre-K	N	Y	Y	2	1
Small classes available in K	Y	Y	Y	3	0
Duration (Consecutive years in CS Setting)					
One year in small class	Y	Y	Y	3	0
Two years in small class	Y	N	Y	2	1
Three years in small class	N	N	Y	1	2
Four years in small class	N	N	Y	1	2
Intensity					
Pull-out interventions	Y	Y	N	2	1
Random class assignment	N	N	N	0	3

Note. Y = Yes, Characteristic Available; N = No, Characteristic Not Available.

Assignment of students to classes at all CS sites was not random. Rather, assignment of students to classes was done collaboratively by teachers and building administrators, each allowing some parental input. In only one school were students able to attend small classes for three to four years consecutively and that was because the school had only one class per grade level and enrollment was capped at 17 students per class. In none of the other CS settings were students able to attend small classes for more than two consecutive years.

Effect Size

Effect size is used to assist in the determination of whether the statistical significance discovered through data analysis reflected a trivial or a meaningful difference. Because sample size directly influences the significance of a *t* test, calculating the effect size provides clarity to how meaningful any differences were. Effect size expresses observed sample mean differences in standard deviation units. “The larger the effect size, the more likely it is that the observed difference is a meaningful difference” (McNamara, 1992, p. 195). Table 11 displays the effect sizes for each third- and fourth-grade category, which support the statistically significant findings of the data analyses. Students in those categories in which statistical significance was found scored, as shown by effect size, from .67 standard deviation units to 1.29 standard deviation units higher on student-achievement-outcome measures. The effect size found in the reading difference of the third-grade free or reduced lunch category may be educationally important, though the small “n” (n = 27) may have been too small for the difference to be considered statistically significant.

Summary

In Chapter 4 the researcher presented and analyzed data related to student-achievement-outcomes in CS and PTR settings. The researcher presented observational data of third- and fourth-grade student behavior and available space per student, in CS and PTR settings. Data were presented that demonstrated the degree to which those characteristics of (CS) that are necessary to produce positive, enduring effects on student-achievement-outcomes were implemented in the CS sites. The explanations of these data and results are presented in Chapter 5.

Table 11

Third- and Fourth-Grade Effect Sizes

	Students			Lunch status	
	Total	Female	Male	Regular	Free or reduced
Third-grade	CS v PTR	CS v PTR	CS v PTR	CS v PTR	CS v PTR
Math	1.18	1.11	1.14	1.29	.82
Reading	.07	.05	.15	.13	.53
Fourth-grade					
Math	.20	.07	.33	.02	.80
Reading	.67	.22	1.16	.75	1.27

Chapter 5

Summary, Conclusions and Recommendations for Practice, Policy and Future Research

Researchers in education and other fields (e.g., economics) have labored to assess the effect that class-size (CS) has on student achievement for decades. Complicating the assessment of CS effects is that much (even most) data reported on the size of classes in schools or school districts is computed and reported as pupil-teacher ratio (PTR) data, or average class-size, not actual class-size. Multiple studies and critiques, authored by education researchers and others, use the terms CS and PTR synonymously, leading to confusion over CS effects because of the imprecise use of the two terms. This confusion has hampered education leaders in their attempts to enact policy and program implementation to improve student achievement.

Purpose of the Study

The purpose of this study was to determine what differences in student achievement and behaviors seemed influenced by class-size (CS) and pupil-teacher ratio (PTR) in each of the two class settings. Student-achievement-outcome scores were gathered from standardized tests given in CS and PTR classrooms and analyzed to determine what, if any, differences occurred in student achievement in two areas: (a) in class-sizes of 13-20 students and (b) in classes of 24-28 students in buildings that have a PTR in the 13:1-20:1 range. Observational data were gathered in CS and PTR classes to determine if there were differences in the amount of space per student or any student behavioral differences based on individual classroom settings. Data were gathered from the CS settings to determine how many of the CS characteristics identified by research as producing positive, enduring effects on student achievement had been implemented.

Scope of the Study

The primary populations identified for this study were public school third- and fourth-grade students in CS and PTR class settings, during the 2002-2003 school year, who had outcome-test results. The student populations (n) in the comparisons were as analogous as possible, that is, data from CS classes of 13-20 were compared to data from students in larger classes (24-28 students) in a school with a PTR of 13:1 to 20:1, as the collected data provided allowed.

Methodology

After reviewing data detailing the per-pupil state foundation grant allotment that each school district received for fiscal years 2002, 2003, and 2004, provided by the Michigan House Fiscal Agency, the researcher sent sent invitations to participate in the study to 102 school districts receiving a per-pupil state foundation grant of $6,500.00. Twenty school districts indicated tentative willingness to participate in the study and share their standardized test data. Student outcomes from standardized tests, such as the California Achievement Test (CAT), the California Test of Basic Skills (CTBS), Iowa Test of Basic Skills (ITBS), Northwest Evaluation Association test (NWEA), and Terra Nova test (TN), and the criterion-referenced Michigan Education Assessment Program (MEAP), were collected and analyzed. There was no control group in the traditional sense, as the focus was on comparisons of influence of CS and PTR on student achievement.

The researcher investigated environmental or context factors in each CS and PTR class setting, specifically including such items as square feet, square feet per student, and circulation space. The number of times students asked questions of the teacher regarding the repeating or clarifying of instructions or content were recorded, as were the numbers of disturbances, student-to-student or individual-student. The researcher also compiled data gathered from CS settings to determine how many of the CS characteristics identified by research as producing positive enduring effects on student achievement (early intervention, sufficient duration, and intensity) were available or had been implemented in the target districts.

Summary of Findings

The analysis of student-achievement-outcome data in third-grade reading and fourth-grade mathematics in CS and PTR classes determined there was no difference in student achievement and that the difference in z scores of students in third-grade reading and the test scores of students in fourth-grade mathematics did not occur by something other than chance. This finding lead to the acceptance of the null hypothesis: There is no difference in student achievement between students in CS and PTR classroom settings for third-grade reading and fourth-grade mathematics.

Analysis of student-achievement-outcome data of third-grade mathematics z scores and fourth-grade reading scores, in CS and PTR class settings determined there was a significant difference in the scores of students in the two settings. The difference between the z-scores between third-grade mathematics students and the test scores of fourth-grade students reading students in CS and PTR classes occurred by something other than chance, which supported the rejection of the null hypothesis that there would be no difference in student achievement between students in CS and PTR classroom settings in third-grade mathematics and fourth-grade reading.

Analysis of third-grade student-achievement-outcome data as z-scores in mathematics and reading between male and female students in CS and PTR settings determined there was a significant difference in the achievement of both male and female students in mathematics. There was no difference in male or female third-grade student-achievement-outcomes in reading.

Analysis of third-grade student-achievement-outcome data as z scores in mathematics and reading between students in CS and PTR settings who purchased regular lunch showed there was a significant difference in achievement for mathematics but not for reading.

Analysis of third-grade student-achievement-outcome data in mathematics and reading between students who participated in the National School Free and Reduced Lunch Program in CS and PTR settings showed there was a significant difference in student achievement in mathematics but not in reading.

Analysis of fourth-grade student-achievement-outcome data in mathematics and reading between male and female students in CS and PTR settings determined there was a significant difference in the achievement of male students in reading. There was found to be no difference in female student achievement in mathematics or reading, as well as no difference in male student achievement in mathematics.

Analysis of fourth-grade student-achievement-outcome data in mathematics and reading between students who purchased regular lunch in CS and PTR settings showed there was a significant difference in student-achievement-outcomes in reading but not mathematics.

Analysis of fourth-grade student-achievement-outcome data in mathematics and reading between students who participated in the National School Free and Reduced Lunch Program in CS and PTR settings showed there was a significant difference in student achievement in both mathematics and reading.

Observation data (see Table 9) demonstrated that students in PTR classrooms had from 2.5 to 4.33 times as many disruptions as did students in CS classrooms. Students in CS classrooms had 1.25 to 2.6 times more square footage per student than did students in PTR settings. The observational data showed that all three of the observed CS settings provided greater square footage per student than that recommended Tanner by (2000a), 49 square feet per student.

Data gathered on the degree of implementation of those characteristics that research, (e.g., Nye, Hedges, & Konstantonpoulos, 1994; Ramey & Ramey, 1998; Finn et al., 2001, Finn, Fox, McClellan, Achilles, & Boyd-Zaharias 2006; Achilles, 2002) determined were factors necessary to achieve the positive, enduring effects of class-size (CS) on student achievement demonstrated that these factors were lacking in two of three CS settings (see Table 10). Two CS settings continued to use pull-out interventions, removing students from their classrooms for remediation, effectively ensuring that some students would miss what was being taught while they are out of the classroom. In all schools participating, the assignment of students to classes was done collaboratively by teachers and building administrators, each allowing some parent input. In only one CS setting did students attend small classes three to four years consecutively, and that was because the school had only one class per grade level and enrollment was capped at 17 students per class. In none of the other CS settings were students able to attend small classes for more than two consecutive years.

Conclusions

Results derived from analysis of the student-achievement-outcome data, as shown in Table 12, were inconclusive in the aggregate. Analyses support the acceptance of the null hypothesis, that there is no difference in student-achievement-outcome data between students in CS and PTR class settings in half the categories. Results also support the rejection of the null hypothesis in half the categories, indicating that there is a significant difference in student-achievement-outcome data between students in CS and PTR class settings. The results indicate that differences were consistent in subject area, that is, mathematics or reading, across the analyzed categories except for the fourth-grade female students and fourth-grade students participating in the National Free and Reduced Lunch Program, for whom the differences were significant in both mathematics and reading.

Student-achievement-outcomes, however, were only part of the data gathered to determine the differences between CS and PTR settings. Data on the implementation of the CS characteristics were gathered (see Table 10) to show at what grade level CS was available to students (early intervention), how many years a student could attend class in a CS setting (sufficient duration), if students were removed from the class for pull out interventions (intensity), and how students were assigned to classes (random assignment). These data describe the extent to which CS was correctly implemented in the targeted sites according to the theoretic framework generated from prior research. Without correct implementation, it is probable that CS cannot be optimally effective.

Importance of Size to a Study

Boozer and Rouse (1995) discussed the importance of school size in a study; they indicated “the larger the school, the larger the variance” (p. 5). The student population in

Table 12

Summarization of the Differences in Student-Achievement Outcomes Identified by Analysis

	Students			Lunch status	
	Total	Female	Male	Regular	Free or reduced
Third grade	CS vs. PTR	CS vs. PTR	CS vs. PTR	CS vs. PTR	CS vs. PTR
Math	Yes	Yes	Yes	Yes	Yes
Reading	No	No	No	No	No
Fourth grade					
Math	No	No	No	No	Yes
Reading	Yes	No	Yes	Yes	Yes

Note. Yes indicates a difference in student-achievement-outcome.
No indicates there was not a difference in student-achievement-outcome.

the participating schools ranged from 166 to 650 students; none of the school sites would be considered excessively large. Demographic information in Table 6 (p. 46) illustrates the homogeneousness between CS and PTR sites. The small and similar size of the schools led to little difference between CS and PTR sites in the numbers of male and female students (gender), numbers of White students and students of other races (ethnicity), and the numbers of students participating in the National Free and Reduced Lunch Program (Lunch Status). The total number (n) of participants in third-grade CS and PTR classes, as noted in Table 3 (p. 41), was n = 121; however, only n = 92 scores were provided for analysis in mathematics. The total number (n) of participants in fourth-grade CS and PTR classes, as noted in Table 3 (p. 38), was n = 132; however, only = 108 scores were provided for analysis in mathematics. It is difficult to expect much variance between student achievement scores in CS and PTR settings with this level of participation. The number (n) of students available for the study was directly correlated to the number of participating CS sites, as CS and PTR sites were used in equal numbers to make comparisons. Of the four sites where administrators indicated that they had small classes in their districts, one building principal withdrew a school, leaving only three CS sites. Thus, only three comparison PTR sites could be used, limiting the amount of data available for use in the study. A study limitation, or impediment, was that principals were reluctant to provide the requested standardized student-achievement outcome-data. Twice the researcher had to call the district superintendent who had agreed to participate in the study to ensure that building principals would supply the requested student-achievement-outcome data. Superintendent intervention resulted in two building principals' complying by command but providing only partial student-achievement-outcome data, as well as partial student demographic information. One building principal of a fourth-grade CS site provided standardized student-achievement-outcome data from a test other than the requested Michigan Education Assessment Program (MEAP). Follow-up calls to procure missing data elements were unsuccessful; the principals would not provide any further data or assistance than that already supplied.

Class-Size Implementation

Successful small classes are more than just adding teachers to schools, as reported by Biddle and Berliner (2002), who stated that

> when it (CS) is planned thoughtfully and funded adequately, long-term exposure to small classes in the early grades generates substantial advantages for students in American schools, and those extra gains are greater the longer students are exposed to those classes." (p. 14).

> Nye, Hedges, and Konstantonpoulos (1994), Ramey and Ramey (1998), Finn et al. (2001, 2006), and Achilles (2002) have identified implementation steps, noted in the theoretical framework, that need to be in place for small classes to have a positive enduring, impact on student-achievement-outcomes. Those conditions are "a) Random Assignment (individual class placement), b) Early Intervention (when the child starts school), c) Intensity (all day, every day), d) Duration (pupils must remaining small classes (13-17 students) three and preferably four years to demonstrate lasting effects" (Achilles, 2002, p. 2). Information provided in Table 10 demonstrates that two of the three CS settings provided small classes in pre-kindergarten (Pre-K), and all three CS settings provided small classes in kindergarten (K).

Class-Size Enduring Effects

Research has demonstrated that for small classes to have positive, enduring effects on student achievement that certain characteristics, identified in the theoretical framework, must be in place. When the actual implementation characteristics of the three CS sites providing student-achievement-outcome data were contrasted to the implementation characteristics portrayed in the theoretical framework, the results received were what should have been expected.

Class-size implementation must be based on accurate research findings that use clear definitions of CS and PTR so that the intervention will include all CS characteristics identified to ensure positive, enduring effects on student-achievement-outcomes. Molnar (1998) underscored this point: "The terms pupil-teacher ratio and class-size are often used interchangeably in everyday conversation. Most people understand both terms to represent the number of students in a typical class with one teacher: This is a false assumption" (p. 28).

Students need to be assigned randomly to their individual classes, as it has been demonstrated to promote increased student-to-student cooperation and to facilitate peer tutoring, and active participation. Data gathered from the CS sites providing student-achievement-outcome data show that at no site were students assigned to classes randomly.

Ramey and Ramey (1989) determined that CS intervention programs must start early in a child's education and that for there to be an enduring impact, the intervention must be delivered by someone other than the parent. Biddle and Berliner (2002) noted that "exposure to small classes in the early grades generates substantial advantages for students and those gains are greater the longer that students are exposed to those classes" (p. 14). Students in only one of the three CS sites had the opportunity to attend small classes for three or four years consecutively.

The positive, enduring effect that the numbers of years students spend in small classes have on student achievement has been well documented. Finn et al. (2006) noted that graduation rates increased with each additional year students spent in small classes. Krueger and Whitmore (2001) determined that students attending small classes K-3 were more likely to take college entrance exams than were students attending large classes. The advantage of students' being in small classes for three to four years is shown in the results obtained in Burke County, North Carolina, schools. The systemic implementation of CS by Burke County educators resulted in significant gains in student-achievement-outcomes, even beyond the K-3 implementation: "Between 1992 and 2000 the district students improved their math scores in third-grade from 61.6% to 80.5%, in fourth-grade from 66.4% to 89.4%, and in fifth-grade from 61.9% to 88.6%" (Sharp, 2002, p. 31). Data provided in Table 10 illustrate that in only one of the three CS sites were students in small classes K-3, and this particular school had only one class per grade level and caps on enrollment at 17 students per class. Students attending the two remaining CS sites spent up to two years in small classes, but not always consecutively. The CS in these two settings was determined by the number (n) of students enrolled in the grade levels, K-3, in any particular year. Because only one of three CS settings that provided student-achievement-outcome data for this study enabled students to attend small classes in all grades K-3, the positive, enduring effects of CS would be difficult to detect by analysis of the CS data provided for this study. This would be consistent with Blatchford et al. (2004), who found that moving students to classes of different sizes caused a disruption effect on children's educational process: "Moving to a class of a different size, especially a larger class was disruptive … it is advisable to maintain smaller classes and to seek … stability in class-size across years" (p. 2).

The amount of time each day a student attends small classes is vital to the positive impact and the enduring effects of CS on student achievement data in Table 10, reveal that two of the three CS sites used pull-out interventions with their students, reducing their time in the classroom, perhaps confusing students about which teacher they were responsible to, and ensuring that pull-out students missed the instruction going on in their class in their absence. Ladson-Billings and Gomez (2001) found that "students who received services from a variety of professionals (e.g., Title 1) were more likely to be confused about to whom they were responsible" (p. 677). Further, removing children from their class setting for pull-out treatments means that the student will miss some on-going instruction while they are out of the classroom (coherence, intensity), ensuring that those students continue to fall behind their classmates. Finn, Gerber, and Boyd-Zaharias (2005) researched the decline in student achievement benefits achieved through the Perry Preschool Project and many Head Start programs. They found limited intensity in each of the programs: Students in the Perry Preschool Program “received only 2.5 hours of school time daily and the typical Head Start program involved 3.5 hours of class time 4 or 5 days a week” (p. 216). In contrast to students in the Perry Preschool Program and Head Start, students who participated in the STAR study spent all day, every day, in their small classes and demonstrated significant positive, enduring achievement effects.

Because only one of three CS settings that provided student outcome achievement data for this study did not use pull-out interventions, the positive enduring effects of CS would be difficult to detect with the data provided. The characteristics of small classes identified in the theoretical framework were minimally present in the three CS sites supplying student-achievement-outcome data to the study. None of the CS sites randomly assigned students to classes; at each site, small classes were available as students started school; however, only one site provided small classes for three or four consecutive years; only one site did not utilize pull-out student interventions. Given the minimal implementation of the characteristics necessary to realize the positive, enduring effects of CS as noted in the theoretical framework and the small number of student participants, it is clear the results obtained by analyzing the available data only minimally reflected the research on CS implementation and enduring effects. Biddle and Berliner (2002) stated, “When it (CS) is planned thoughtfully and funded adequately, long-term exposure to small classes in the early grades generates substantial advantages for students in American schools, and those extra gains are greater the longer students are exposed to those classes" (p. 14). The present study demonstrated (again) that improper implementation produced little or no measurable achievement gain for students who attended small classes as opposed to those students who attended PTR classes.

The positive, enduring effects on student achievement are but one measure of the positive impact of small classes on students. Research (e.g., Johnson, 1990; Biddle & Berliner, 2002) has shown that small classes can create a supportive context for learning and an increased sense of community among students. Sarason (1974) coined the term *psychological sense of community* (PSOC) to describe the need of humans to be part of a community. Goodenow (1993) examined PSOC in terms of classrooms and found it to be correlated with a student's academic success. According to Mosteller (1995), "reducing [the size of classes in early grades] reduces the distractions in the room and gives the teacher more time to devote to each child" (p. 125). Observational data, displayed in Table 9, demonstrated that students in PTR classes had from 2.5 to 4.33 times as many disruptions as did students in CS classes, which confirmed what Bateman (2002) found, that "students in small classrooms [classes] report lower levels of antisocial behavior and higher affective evaluations of their peers" (p. 73). The vast majority of the disruptions observed in the present study were attributable to student behavioral issues caused either by a single student's or groups of students' disrupting the class. Students in small classes worked better with classmates in small groups, contributed to common class goals, and became valued members of the classroom community. Data in Table 9 also show that students in CS classes had 1.25 to 2.6 times more square footage per student than did those students in PTR settings. Tanner (2000b) reported that "if there are too many people in a given space, we usually react negatively. Children react both by withdrawing, physically and socially, and by acting aggressively" (p. 5). The observational data showed that all three of the observed CS settings provided greater square footage per student, by more than 3 to 30 square feet per student, than the 49 square feet per student recommended by Tanner (2000a).

The observational data support research (e.g., Finn, Pannozzo, & Achilles, 2003) indicating the positive impact small classes have on student behaviors. Students in the observed small class classrooms displayed significantly fewer student behavioral issues (2.5 to 4.3 times fewer) than did students in the corresponding PTR classrooms, suggesting that students feel a greater sense of community with their classmates and fill their need for PSOC in small classes. Small classes also provided significantly more room per student (1.25 to 2.6 times more square footage per student) than in the corresponding PTR classrooms. Crowding influences behaviors; Tanner showed that little students need about 49 ft2 each to learn well.

The data and analyses in this study partially supported findings regarding the positive influences that small classes can have on student achievement and behaviors, but the lack of implementation fidelity may have hidden any positive findings of the short-term achievement effects of CS. The finding that there was no difference between student-achievement-outcome scores of students in CS and PTR settings should not be a surprise if small classes are not implemented according to steps determined by research as important for successful outcomes. The study also showed that small classes lead to a greater sense of community among classmates, which results in fewer observed instances of negative student behavior.

Because participation in the study was voluntary, gathering data of sufficient quantity and quality was difficult. This fact was a serious constraint and limitation on the study, particularly among CS sites. One site supplied student-achievement-outcome data through a standardized test other than the MEAP test; efforts to secure MEAP results from the site were not successful. Some participating sites supplied student-achievement-outcome data for either reading or mathematics but not for both reading and mathematics. These omissions on the part of participating CS and PTR sites caused difficulty in the study. There was reluctance from some building principals to supply information regarding student participation in the National School Free and Reduced Lunch Program as well as the ethnicity of participating students. These omissions by participating sites inhibited the researcher's ability to disaggregate the data.

Recommendations for Policy and Practice

School administrators and policy makers need to seek and develop policies to ensure that the implementation of class-size reforms follows those steps, found by research, necessary for sustained short-term and long-term positive, enduring effects on student achievement and behaviors. Toward that end I recommend that

- Policy makers and administrators understand the sources of data used in decision making. Confusion regarding CS and PTR terms is promulgated by individuals' and publications' (e.g., Hanushek [1999] and Johnson [2002]) reporting CS results using PTR data to support the findings. Administrators and policy makers need to invest the time necessary to understand the background of the author(s) of the research they are reviewing, as well as locate referenced data tables to ensure that what is referenced is what is presented in the original source.

- Class-size (CS) policy needs to be clear and concise in the definition of CS, which is "The number of students for whom a teacher is primarily responsible during a school year" (Lewit & Baker, 1997, p. 113), arrived at by adding the number of students present in the classroom.
- Class-size (CS) implementation should follow the steps identified through the STAR experiment: early intervention, sufficient duration, intense treatment (including coherence), random assignment, and employment of a cohort model.
- Implementation of CS should begin with a student's entry into the school environment, be it in Pre-K, K, or first-grade.

Recommendations for Future Research and Study

The present study supported existing research found that "small classes are not simply hiring teachers and doing business as usual" (Achilles. 2005, p. 15) and that proper implementation is paramount to achieving positive short and enduring effects on student behavior and achievement. Further research on class-size implementation and outcomes is needed, such as the following:

- Research conducted between schools that have correctly implemented small classes and can document the implementation structure identified by research, as delineated by Ramey and Ramey (1989) or Achilles (2002), and schools that have just hired additional teachers and provided instruction as usual. This would clarify the difference that properly and improperly implemented small classes have on student-achievement-outcomes.

- The classroom environmental impact that small classes have on student behaviors warrants considerable research (space, air quality, room arrangements).
- Study of small-class implementation in differing school settings within states needs to relate to and compare student achievement gains (e.g., rural, small-town, suburban, urban, and inner-city).

Summary Statement

The study provided limited support for existing research on the impact of class size (CS) on student achievement. However, the limited support may be related to the fact that the class-size effort in the districts studied in this research were not initiated and implemented in accordance with the growing body of CS work. The theoretical framework of the study set forth the necessary characteristics for CS success; without proper implementation there may be no measurable advantage in student-achievement-outcome scores for those students in CS class settings in the short or long run when compared to outcome scores for students in PTR settings. There must be consistent implementation of small classes, as prescribed by research, in order to bring clarity to the benefits of having small classes in a school. The consistent implementation of small classes and clarity in reporting of CS and PTR data may allow policy makers and parents alike to make informed decisions, using correct data, and to provide the best possible learning environments for young children. The confusion surrounding CS and PTR classroom settings inhibits the implementation of CS, a school improvement initiative that no study has shown as detrimental to student achievement or behavior.

The influence of small classes goes beyond improved student achievement, however, as small classes positively influence student behaviors and students' feelings of belonging to a classroom community (e.g., PSOC). The present study demonstrated that even improperly implemented, small classes provide an environment featuring fewer negative student behavioral issues and an increased communal learning atmosphere.

The study confirmed that there are no negative student impacts from the implementation of small classes and that when small classes are implemented correctly, students experience the positive enduring effects of improved student achievement.

References

Achilles, C .M. (2005, April). Why hasn't class-size research been used appropriately (Or even used)? Paper Presented at the American Educational Research Association Annual Conference, Montreal, Canada.

Achilles, C .M. (2002, December). The impact of class-size reduction (CSR). Paper presented at the 2002 meeting of the SERVE Research and Policy Symposium on Class-Size Reduction and Beyond, Raleigh, NC.

Achilles, C .M., (1999) Let's put kids first, finally. Thousand Oaks, CA: Corwin Press Inc.

Achilles, C. M., & Finn, J. D. (2002, April). The role of school and district leadership in reform: A case of validity as mistaken identity. Paper presented at the annual conference of the American Educational Research Association (AERA), New Orleans, LA.

Achilles, C. M., & Sharp, M. (1998, Fall). Solve your puzzles using class-size (CS) and pupil-teacher ratio (PTR) differences. Catalyst for Change, 28(1), 5-10.

Bateman, H. V. (2002) Students' sense of community. In J. Finn & M. Wang (Eds.), Taking small classes one step further (pp. 63-75). Greenwich, CT: Information Age Publishing, Inc.

Biddle, B. J., & Berliner, D. C. (2002) What research says about small classes & their effects. San Francisco: WestEd Policy Perspectives.

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

Blatchford, P., Bassett. P., Brown, P., Martin, C., & Russell, A. (2004, December). The effects of class-size on attainment and classroom process in English primary schools (years 4 to 6) 2000-2003. Research Brief. RBX13-04. London, England: University of London, Institute of Education.

Bloom, B .S. (1984, June/July). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6) 3-16.

Bohrnstedt, G. W., Stecher, B. M., & Wiley, E. W. (2000). The California class size reduction evaluation: Lessons learned. In M. Wang & J. Finn (Eds.), How small classes help teachers do their best (pp. 201-226). Philadelphia, PA: Temple University Center for Research in Human Development and Education.

Boozer, M., & Rouse, C. (1995, June). Intraschool variation in class-size: Patterns and implications. (Working Paper No. 344). Washington, DC: National Bureau of Economic Research, Industrial Relations Section.

Bracey, G. W. (1999, November). Reducing class size: The findings, the controversy. PhiDelta Kappan, 81(3), 246-247.

Carnegie Corporation of New York. (1995). Great transitions: Preparing adolescents for a new century. Carnegie Council on Adolescent Development. New York: Carnegie Corporation of New York.

Chase, C. I., Muller, D. J. & Walden, J. D. (1986) PRIME TIME: Its impact on instruction and achievement. Final Report, December. Indianapolis: Indiana Department of Education

Duncanson, E. (2003). Classrooms are not roadways. They are parking lots! The Science Teachers Bulletin, 66(2), 1-4.

Egelson, P., & Harmon, P. (2000) Ten years of small class size in Burke County, North Carolina. In M. Wang & J. Finn (Eds.), How small classes help teachers do their best (pp. 279-298). Philadelphia, PA: Temple University Center for Research in Human Development and Education.

Finn, J. D., & Achilles. C. M. (1999, Summer). Tennessee's class-size study: Findings, implications, misconceptions. Educational Evaluation and Policy Analysis, 21(2), 97-109.

Finn, J. D., Fox, J. D., McClellan, M., Achilles, C. M., & Boyd-Zaharias, J. (2006). Small classes in the early grades and course taking in high school. International Journal of Education Policy and Leadership, 1(1), Retrieved July 14, 2006, from http://journals.sfu.ca/ijepl/index.php/ijepl/article/view/17

Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2001, April). The enduring effects of small classes. Teachers College Record, 103(2), 145-183.

Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005) Small classes in the early grades, academic achievement, and graduating from high school. Journal of Educational Psychology, 97(2), 214-223.

Finn, J. D., Pannozzo, G .M., & Achilles, C. M. (2003, Fall). The “why’s” of class size: Student behavior in small classes. Review of Educational Research, 73(3), 321-359.

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. Journal of Applied Psychology, 82(2), 221-234.

Goodenow, C. (1993). Classroom belonging among early adolescent students. Journal of Early Adolescence, 13(1), 21-43.

Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). Improving student achievement: What state NAEP test scores tell us. Santa Monica, CA: Rand Education.

Grissmer, D. W., Flanagan, A., Kawata, J., & Williamson, S. (1998). Why did Black test scores rise rapidly in the 1970’s and 1980’s? In C. Jencks & M. Phillips (Eds.), The Black-White test score gap (pp.182-226). Washington, DC: Brookings Institution.

Hanushek, E. A. (2002). Evidence, politics, and the class-size debate. In L. Mishel & R. Rothstein (Eds.), The class-size debate (pp. 37-66). Washington, DC: The Economic Policy Institute.

Hanushek, E. A. (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.

Hanushek, E. A. (1998, September). The evidence on class-size. Rochester, NY: University of Rochester, W.A. Wallis Institute of Political Economy.

Hauser, R. M. (1998). Trends in Black-White test-score differentials. In U. Neisser (Ed.), *The rising curve: Long-term changes in IQ and related measures* (pp. 219-250). Washington, DC: American Psychological Association.

Hedges, L. V., & Nowell, A. (1998) Black-White test score convergence since 1965. In C. Jencks & M. Phillips (Eds.), The Black-White test score gap (pp. 149-181). Washington, DC: Brookings Institution.

Johnson, B. (2001, March). Toward a new classification of non-experimental quantitative research. Educational Researcher, 30(2), 3-13.

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

Johnston, J. M. (1990). What are teachers' perceptions of teaching in different classroom contexts? Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service N0. ED320867)

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Reinhart & Winston.

Krueger, A. B. (1999). Experimental estimates of educational production functions. Quarterly Journal of Economics, CXIV, 497-532.

Krueger, A. B. (1998). Reassessing the view that American schools are broken. Federal Reserve Bank of New York Economic Policy Review, 4(1), 29-41.

Krueger, A., & Whitmore, D. (2001) The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. Economic Journal, 111, 1-28.

Ladson-Billings, G., & Gomez, M. L. (2001) Just showing up: Supporting early literacy through teachers' professional communities. Phi Delta Kappan, 82(9), 675-680.

Lain, S. W. M., & Ward, J. G. (Eds.) (2000). Using what we know, a review of the research on implementing class-size reduction initiatives for state and local policymakers. Oakbrook, IL: NCREL (Esp. Chapters 1-4 and 6)

Lewit, E. M., & Baker, L. S. (1997). Class size. The Future of Children, 7(3), 112-121.

Lofland, J., & Lofland, L. H. (1995). Analyzing social settings: A guide to qualitative observations and analysis (3rd ed.). Belmont, CA: Wadsworth Publishing.

McNamara, J .F. (1992, April). The effect size criterion in school improvement research. International Journal of Education Reform, 1(2), 91-202.

McRobbie, J., Finn, J. D., & Harmon, P. (1998, August). Class-size reductions: Lessons learned from experience. (Policy Briefs, No. 23). San Francisco, CA: West Ed.

Molnar, A. (1998). Smaller classes, not vouchers, increase student achievement. Harrisburg, PA: Keystone Research Center.

Molnar, A., Smith, P., Zahorik, J., Halbach, A., Ehrle, K., Hoffman, L .M., & Cross, B. (2001, December). 2000-2001 evaluation results of the Student Achievement Guarantee in Education (SAGE) program. Milwaukee: University of Wisconsin.

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

National Center for Education Statistics (NCES). (2000). Digest of Education Statistics 1999. Washington DC: U.S. Department of Education, Office of Educational Research & Improvement.

Norusis, M. J. (2000). SPSS 10.0 guide to data analysis, Upper Saddle River, NJ: Prentice Hall.

Nye, B., Hedges, L., & Konstantonpoulos, S. (1999, Summer). The long-term effects of small classes: A five-year follow-up of the Tennessee class study experiment. Educational Evaluation and Policy Analysis, 21(2), 143-163.

Nye, B., Hedges, L., & Konstantonpoulos, S. (1994, Special Issue). The long-term effects of small classes: A five-year follow-up of the Tennessee Class Size Experiment. Educational Evaluation and Policy Analysis, 21(2), 127-142.

Olson, M. N. (1971). Research notes: Ways to achieve quality in school classrooms: Some definitive answers. Phi Delta Kappan, 53(1), 63-65.

Ramey, C. T., & Ramey, S. (1989). Early intervention and early experience. American Psychologist, 53, 109-120.

Salkind, N. (2000). Statistics for people who (think they) hate statistics. Thousand Oaks, CA: Sage Publications, Inc.

Sarason, S. B. (1974). The PSOC: Prospects for a community psychology. San Francisco: Jossey-Bass.

Sharp, M. A. (2002). An analysis of pupil-teacher ratio and class-size: Differences that make a difference. Unpublished doctoral dissertation, Eastern Michigan University, Ypsilanti.

Tanner, C. (2000a). Minimum classroom size and number of students per classroom. Retrieved June 12, 2007, from www.coe.uga.edu/sdpl/research/territoriality.html.

Tanner, C. (2000b). [Review of the book Caring spaces, learning places: Children's environments that work].Retrieved June 12, 2007, from www.coe.uga.edu/sdpl/articlesandpapers/reviewcaringspaces/caringspace_s/html.

Weinstein, C. (1979, Fall). The physical environment of the school: A review of the research. Review of Education Research, 49(4), 577-610.

Word, E., Johnston, J., Bain, H. P., Fulton, B. D., Zaharias, J. B., Lintz, M. N., Achilles, C. M., Folger, J., & Breda, C. (1990). *Student/teacher achievement ratio (STAR)*, Tennessee's K-3 class-size study: Final summary report, 1985-1990. Nashville, TN: Tennessee State Department of Education.

Appendices

Appendix A: Texas House Bill 72

68th LEGIS-2nd CALLED SESSION CH 28, SED. IV-E1

PART C, CLASS-SIZE (P. 167)

SECTION 1. Section 16.054, Education Code, is amended to read as follows:

Sec. 16.054 STUDENT/TEACHER RATIOS; CLASS-SIZE, (a) *Except as provided by Subsection (b) of this section, each* [~~EACH~~] school district must employ a sufficient number of certified teachers to maintain an average ratio of not less than one teacher for each 20 [~~25~~] students in daily attendance.

(b) Beginning with the 1985-1986 school year, a school district may not enroll more than 22 students in a kindergarten, first, or second-grade class. Beginning with the 1988-1989 school year, a school district man not enroll more than 22 students in a third or fourth-grade class. This requirement shall not apply during the last 12 weeks of any school year.

(c) In determining the number of students to enroll in any class, a district shall consider the subject to be taught, the teaching methodology to be used, and any need for individual instruction.

(d) On application of a school district, the commissioner may except the district from the limits in Subsection (b) of this section if the commissioner finds the limits work an undue hardship on the district. An exception expires at the end of the semester for which it is granted, and the commissioner may nor grant an exception for no more than one semester at a time.

(e) The commissioner shall report to the legislature each biennium regarding compliance with this section. The report must include:

(1) a statement of the number of school districts granted an exception under Subsection (d) of this section; and

(2) an estimate of the total cost incurred by school districts in that biennium in complying with this section.

TEXAS H.B. 72, 1984

Appendix B: University Human Subjects Approval

Bob Becker – UHSRC approval

From: Patrick Melia Patrick.melia@emich.edu
To: beckerr@harpercreek.net
Date: 1/30/04 12:17PM
Subject: UHSRC approval

Mr. Becker,

The University Human Subjects committee has reviewed your protocol "Student Achievement: Class-size vs. Pupil-Teacher Ratio" and has recommended approval as the protocol is written. This is to inform you that you are approved to begin your data collection effective immediately.

If you change your protocol and decide to interview any human subjects then you will have to return to the committee and provide a copy of the Consent Agreement to be used and a complete copy of all questions to be asked of the participants. Currently this approval only covers your proposed review of student test scores.

You will receive an official letter of approval following our next meeting of the UHSRC on February 20th. Please be sure to place a copy of this letter in an appendix of your dissertation when it is turned into the Graduate School for review.

With best wishes,

Dr. Patrick Melia
Associate Graduate Dean
Administrative Co-Chair
UHSRC

CC: <Charles.Achilles@emich.edu>, <Steve.Pernecky@emich.edu>

Appendix C: University Human Subjects Extension

Bob Becker – UHSRC approval

From: Patrick Melia Patrick.melia@emich.edu
To: beckerr@harpercreek.net
Date: 1/28/0 11:03AM
Subject: UHSRC Extension approval

Mr. Becker,

This is to let you know that the UHSRC committee has approved your request for a year long extension to your previously approved protocol "Student Achievement: Class-size vs. Pupil-Teacher Ratio." You will be receiving an official letter following our next UHSRC meeting on February 25th but this is to let you know of the approval.

With best wishes,

Dr. Patrick Melia
Associate Graduate Dean
Administrative Co-Chair
UHSRC

CC: <Charles.Achilles@emich.edu>

Appendix D: Mean and Standard Deviation Information for Third-Grade

Class-size

Iowa Basic

Reading

Mean 186.4
Standard Deviation 21.7

Mathematics

Mean 185.7
Standard Deviation 17.7

NWEA

Reading

Mean 194.3
Standard Deviation 16.5

Mathematics

Mean 188.6
Standard Deviation 13.2

Pupil-Teacher Ratio

Terra Nova

Reading

Mean 624
Standard Deviation 41.10

Mathematics

Mean 595
Standard Deviation 37.54

Appendix E: Letter to School Superintendents and School and District Information Sheet

March 17, 2004

Dear Superintendent,

My name is Bob Becker, I am a doctoral student in education leadership at Eastern Michigan University. I am conducting a research study on student achievement as a factor of class-size (CS) and pupil-teacher ratio (PTR): CS is the number of students in a classroom. PTR is the number of students in a building divided by the number of certified staff. Confusion surrounding these two terms may influence student achievement and resulting policies. Many critics of public education use PTR data to show that class-size does not positively impact student achievement. Often education agencies do not routinely collect accurate class-size data.

This study will analyze standardized test scores from third and fourth-grade public school students in a sample of Michigan classrooms. The CS and PTR settings will be comparable in order to determine positive or negative impacts of either setting. Third-grade standardized test information will be gathered from whatever test the district is currently using. Fourth-grade data will be the 2003 M.E.A.P. test. I will use the demographic information of gender, race, and free and reduced lunch eligibility to determine if CS or PTR effect students differently in any pattern. All information regarding participating students, schools and districts will be confidential. At any time, you are free to withdraw from the study.

Please complete the attached form indicating your willingness to assist in this study on the impact of CS and PTR on student achievement and return it in the enclosed stamped envelope. From those respondents returning the enclosed building-level data form I will look for schools with similar characteristics such as state foundation grant amounts, school populations, and CS/PTR similarities. Final participants will be selected based on how closely school characteristics match CS and PTR classroom numbers. I will work with the selected respondents on the easiest method of data transfer and reimburse any costs incurred. Participants will receive an abstract of findings.

Should you have any questions, please call or E-mail me as shown below. I want to reiterate: all information will be confidential; no names will be stated in the study. Thank you for assisting in this important study. The building-level data form is enclosed.

Sincerely,

Bob Becker

Bob Becker
108 Shadowood Lane
Battle Creek, MI 49014
W 269/979-1135
H 269/660-8369
beckerr@harpercreek.net

Charles M. Achilles, Ed.D.
304 Porter Building
Eastern Michigan University
Ypsilanti, MI 48197
734/487-0255

Student Achievement Analysis: Class-size and Pupil-Teacher Ratio

Elementary School: ________________________ ***Total School Enrollment:*** __________

Building Principal/Contact: ____
Building Address: ____
Building Phone Number: ____
School District: ____________________ **Superintendent:** _______________________

Definitions of Importance

FTE *– Full Time Equivalence, based on hours worked as a full time teacher representing 1.0*

Regular Classroom Teachers – Teachers who work with students in their classroom each day. These can be inclusionary classrooms.
Specials Teachers – Teachers of Art, Music, Physical Education, Foreign Language, etc.
Support Teacher – Any certified teacher supporting regular classrooms (e.g. LD, EI, EMI, Title I, Resource Room, or any Special Education teacher who supports the regular program.

- **District Data** (Use FTE where appropriate) **Total (N) FTE**

  Total Enrollment __________

  Total Certified Staff (Regular, Specials, and Support) __________

  District Per Student Foundation Grant **$**__________

- **Building Data** (Use FTE where appropriate) **Total (N) FTE**

  Number of Students __________

  Percentage of Students Receiving Free Lunch __________

  Percentage of Students Receiving Reduced Lunch __________

  Regular Classroom Teachers __________

  Specials Teachers __________

  Support Teachers __________

  Administrators, Counselors, Psychologists, etc. __________

- **Select One Third-grade Classroom**
  Number of Students __________

- **Select One Fourth-grade Classroom**
  Number of Students __________

- **Third-grade Standardized Test Used** ____________________________

May I visit, for observation only, the selected classrooms? __________

Appendix F: Reminder Letter to School Superintendents

April 16, 2004

Dear Superintendent,

My name is Bob Becker, I am a doctoral student in education leadership at Eastern Michigan University. I am conducting a research study on student achievement as a factor of class-size (CS) and pupil-teacher ratio (PTR). Last month I sent you the attached letter and information form, as well as a stamped self addressed envelope to return the information to me. I know how busy this time of year can be so I am enclosing the letter with the information form printed on the opposite side. I would ask that you please read my letter and ask one of your elementary school principals to complete the information form and return it to me by May 14, 2004. I would like to select schools and have the data collected prior to end of the school year.

I know the time is short; your assistance is greatly appreciated. Increasing student achievement is in the forefront of education today, it is my hope that this study will be of assistance to the educational community, and your help will make that possible.

Sincerely,

Bob Becker

Bob Becker
108 Shadowood Lane
Battle Creek, MI 49014
W 269/979-1135
H 269/660-8369
beckerr@harpercreek.net

Charles M. Achilles, Ed.D.
304 Porter Building
Eastern Michigan University
Ypsilanti, MI 48197
734/487-0255

Appendix G: Letter to Participating School Principals

June 14, 2004

Dear Principal,

Thank you for agreeing to provide me standardized test data for my research study on student achievement as a factor of class-size (CS) and pupil-teacher ratio (PTR). This study will analyze standardized test scores from third and fourth-grade public school students in a sample of Michigan classrooms. All information regarding participating students, schools and districts will be confidential. The CS and PTR settings will be comparable in order to determine positive or negative impacts of either setting.

Please feel free to send me data any number of ways; any of the options below will work:

- I am enclosing a disc with an EXCEL spreadsheet as one possible method. Should you choose to use the EXCEL spread sheet you can email the data back to me.
- For the fourth-grade MEAP test data you print the comprehensive report list by student and the only information you would need to add would be whether the student received a free or reduced lunch and the ethnicity of each student.
- If your third-grade test results come in a similar format to the MEAP and it is easiest to make a copy of the results and send them to me that is also fine.
- If the information is already in electronic media, email is another option.

In short, whatever method is the easiest for you to send data to me, use it.

Should you have any questions, please call or E-mail me as shown below. I want to reiterate: all information will be confidential; no names will be stated in the study. Thank you for assisting in this important study. The building-level data form is enclosed.

Sincerely,

Bob Becker

Bob Becker
108 Shadowood Lane
Battle Creek, MI 49014
W 269/979-1135
H 269/660-8369
beckerr@harpercreek.net

Appendix H: Third-Grade CS Data

Setting	Student	Gender	Ethnicity	Lunch	Math	Reading	Math Z	Reading Z
1	3101	1	5	1	151	152	-1.96045	-1.58525
1	3102	1	5	2	186	185	0.016949	-0.06452
1	3103	1	5	1	164	166	-1.22599	-0.94009
1	3104	1	5	2	182	173	-0.20904	-0.61751
1	3105	1	5	1	178	185	-0.43503	-0.06452
1	3106	1	5	2	202	196	0.920904	0.442396
1	3107	2	5	2	213	207	1.542373	0.949309
1	3108	2	5	2	213	199	1.542373	0.580645
1	3109	1	5	2	181	176	-0.26554	-0.47926
1	3110	2	5	2	196	208	0.581921	0.995392
1	3111	1	5	2	196	195	0.581921	0.396313
1	3112	2	5	1	176	164	-0.54802	-1.03226
1	3113	1	5	2	182	179	-0.20904	-0.34101
1	3114	2	5	2	184	184	-0.09605	-0.1106
1	3115	1	5	2	187	196	0.073446	0.442396
1	3116	1	5	2	196	200	0.581921	0.626728
1	3117	2	2	2	202	191	0.920904	0.211982
1	3118	2	5	2	187	184	0.073446	-0.1106
1	3119	1	5	3	165	153	-1.78788	-2.50303
1	3120	2	5	3	170	172	-1.40909	-1.35152
1	3121	2	3	3	185	183	-0.27273	-0.68485
1	3122	1	3	3	178	184	-0.80303	-0.62424
1	3123	2	5	3	190	188	0.106061	-0.38182
1	3124	1	3	3	181	193	-0.57576	-0.07879
1	3125	1	5	3	190	197	0.106061	0.163636
1	3126	2	5	3	199	199	0.787879	0.580645
1	3127	2	5	3	194	200	0.409091	0.626728
1	3128	2	5	3	203	203	1.090909	0.764977
1	3129	1	5	3	205	207	1.242424	0.949309
1	3130	1	5	3	203	215	1.090909	1.317972
1	3131	2	5	2	192	179	0.355932	-0.34101
1	3132	2	5	1	169	155	-0.9435	-1.447
1	3133	2	5	1	168	158	-1	-1.30876
1	3134	2	5	1	184	189	-0.09605	0.119816
1	3135	2	4	1	198	169	0.694915	-0.80184
1	3136	1	5	1	176	184	-0.54802	-0.1106
1	3137	2	5	2	156	165	-1.67797	-0.98618
1	3138	1	5	1	170	175	-0.88701	-0.52535
1	3139	1	5	1	226	199	2.276836	0.580645
1	3140	1	5	1	174	186	-0.66102	-0.01843
1	3141	1	5	1	172	204	-0.77401	0.81106
1	3142	2	3	1	173	179	-0.71751	-0.34101
1	3143	1	5	1	175	176	-0.60452	-0.47926
1	3144	1	5	1	174	171	-0.66102	-0.70968

Appendix I: Third-Grade PTR Data

Setting	Student	Gender	Ethnicity	Lunch	Math	Reading	Math Z	Reading Z
2	3201	1	0	1	643	632	-0.37044	0.11576
2	3202	1	0	2	721	683	1.72217	1.52402
2	3203	1	0	2	684	740	0.72952	3.09796
2	3204	2	0	2	678	643	0.56855	0.4195
2	3205	2	0	2	647	635	-0.26312	0.1986
2	3206	2	0	2	701	638	1.1856	0.28144
2	3207	2	0	2	688	617	0.83684	-0.29843
2	3208	2	0	2	666	653	0.24661	0.69563
2	3209	2	0	2	750	650	2.50019	0.61279
2	3210	2	0	1	621	596	-0.96066	-0.8783
2	3211	1	0	1	673	644	0.43441	0.44712
2	3212	1	0	2	612	588	-1.20211	-1.09921
2	3213	1	0	1	625	626	-0.85334	-0.04992
2	3214	1	0	1	618	612	-1.04114	-0.4365
2	3215	1	0	1	622	580	-0.93383	-1.32011
2	3216	1	0	2	602	557	-1.47039	-1.95521
2	3217	1	0	2	624	619	-0.88017	-0.24321
2	3218	2	0	2	668	633	0.30027	0.14337
2	3219	1	0	1	658	641	0.03199	0.36428
2	3220	2	0	1	651	655	-0.15581	0.75086
2	3221	1	0	2	613	590	-1.17528	-1.04398
2	3222	2	0	1	629	584	-0.74603	-1.20966
2	3223	2	0	2	628	639	-0.77286	0.30905
2	3224	2	0	2	688	631	0.83684	0.08815
2	3225	1	0	1	683	633	0.70269	0.14337
2	3226	1	0	1	684	604	0.72952	-0.6574
2	3227	1	0	2	564	555	-1.8552	-1.67496
2	3228	1	0	2	586	577	-1.19155	-0.9657
2	3229	2	0	2	636	642	0.31674	1.12983
2	3230	1	0	2	623	640	-0.07541	1.06535
2	3231	2	0	2	639	600	0.40724	-0.22421
2	3232	2	0	2	636	629	0.31674	0.71072
2	3233	2	0	2	629	598	0.10558	-0.28869
2	3234	1	0	2	653	623	0.82956	0.51729
2	3235	1	0	2	670	649	1.34238	1.3555
2	3236	1	0	2	645	616	0.58823	0.29162
2	3237	2	0	2	625	570	-0.01508	-1.19138
2	3238	1	0	2	702	648	2.30769	1.32326
2	3239	1	0	2	620	621	-0.16591	0.45281

Appendix I: Third-Grade PTR Data

Setting	Student	Gender	Ethnicity	Lunch	Math	Reading	Math Z	Reading Z
2	3240	2	0	2	614	574	-0.34691	-1.06242
2	3241	2	0	2	638	610	0.37707	0.09818
2	3242	1	0	2	629	621	0.10558	0.45281
2	3243	1	0	2	623	626	-0.07541	0.61401
2	3244	1	0	2	589	601	-1.10105	-0.19197
2	3245	2	0	2	654	636	0.85973	0.93639
2	3246	2	0	2	571	548	-1.64404	-1.90063
2	3247	1	0	2	641	610	0.46757	0.09818
2	3248	2	0	2	574	559	-1.55354	-1.546
2	3249	2	1	5	0	528	0	0
2	3250	2	2	5	0	450	0	0
2	3251	2	2	5	0	500	0	0
2	3252	2	2	5	0	520	0	0
2	3253	2	2	5	0	573	0	0
2	3254	2	1	5	0	560	0	0
2	3255	2	1	5	0	524	0	0
2	3256	2	2	5	0	456	0	0
2	3257	2	2	5	0	505	0	0
2	3258	2	1	5	0	483	0	0
2	3259	2	2	5	0	439	0	0
2	3260	2	1	5	0	497	0	0
2	3261	2	1	5	0	489	0	0
2	3262	2	2	5	0	500	0	0
2	3263	2	2	5	0	528	0	0
2	3264	2	1	5	0	538	0	0
2	3265	2	1	5	0	485	0	0
2	3266	2	2	5	0	507	0	0
2	3267	2	2	5	0	524	0	0
2	3268	2	2	5	0	495	0	0
2	3269	2	1	5	0	502	0	0
2	3270	2	1	5	0	491	0	0
2	3271	2	2	5	0	495	0	0
2	3272	2	2	5	0	483	0	0
2	3273	2	1	5	0	532	0	0

Appendix J: Fourth-Grade CS Data

Student	Setting	Gender	Ethnicity	Lunch	Math	Reading
4101	1	1	5	1	503	491
4102	1	1	5	1	551	576
4103	1	2	5	3	567	602
4104	1	2	5	1	536	576
4105	1	1	4	1	511	477
4106	1	1	5	1	554	582
4107	1	2	5	3	564	564
4108	1	2	5	3	564	602
4109	1	1	5	3	507	513
4110	1	2	5	2	513	537
4111	1	1	4	1	533	519
4112	1	1	5	3	551	556
4113	1	1	5	1	519	506
4114	1	2	5	3	592	560
4115	1	1	5	1	548	506
4116	1	2	5	3	561	550
4117	1	2	5	1	519	528
4118	1	1	5	3	542	602
4119	1	2	5	3	526	528
4120	1	1	4	3	557	576
4121	1	2	5	3	546	588
4122	1	1	3	3	536	567
4123	1	1	5	3	556	550
4124	1	2	5	3	549	588
4125	1	1	5	1	503	493
4126	1	2	4	3	542	572
4127	1	2	5	3	539	504
4128	1	2	5	1	517	506
4129	1	2	5	3	575	553

Appendix J: Fourth-Grade CS Data

Student	Setting	Gender	Ethnicity	Lunch	Math	Reading
4130	1	1	5	3	562	567
4131	1	2	5	2	526	537
4132	1	1	5	3	522	544
4133	1	2	5	3	0	215
4134	1	2	5	3	0	198
4135	1	1	4	3	0	151
4136	1	1	5	1	0	224
4137	1	2	1	2	0	246
4138	1	1	3	1	0	169
4139	1	2	5	3	0	224
4140	1	1	3	1	0	158
4141	1	2	3	1	0	166
4142	1	1	5	1	0	169
4143	1	1	1	1	0	193
4144	1	2	5	1	0	166
4145	1	1	5	3	0	240
4146	1	1	5	3	0	196
4147	1	1	4	1	0	172
4148	1	2	5	3	0	200
4149	1	2	5	3	0	193

Appendix K: Fourth-Grade PTR Data

Student	Setting	Gender	Ethnicity	Lunch	Math	Reading
4201	2	2	5	3	542	560
4202	2	1	5	3	571	576
4203	2	2	5	3	546	576
4204	2	2	5	3	566	576
4205	2	1	5	3	575	594
4206	2	1	5	3	595	560
4207	2	1	5	3	614	588
4208	2	2	3	2	551	588
4209	2	1	5	3	559	537
4210	2	1	5	2	575	556
4211	2	1	5	3	562	564
4212	2	2	5	1	575	560
4213	2	1	5	3	581	560
4214	2	1	5	1	536	528
4215	2	2	5	2	539	560
4216	2	1	5	3	567	532
4217	2	1	5	3	557	547
4218	2	1	5	1	556	488
4219	2	2	5	3	571	542
4220	2	1	5	3	592	567
4221	2	1	5	3	561	539
4222	2	1	5	3	587	611
4223	2	1	4	2	597	602
4224	2	2	5	3	567	567
4225	2	2	5	3	526	532
4226	2	1	5	3	554	500
4227	2	2	5	3	597	560
4228	2	1	5	3	575	560
4229	2	1	5	3	539	588
4230	2	2	5	3	557	594
4231	2	2	5	3	545	553
4232	2	2	5	3	519	539
4233	2	1	3	1	529	530
4234	2	2	5	1	536	553
4235	2	2	5	3	507	493
4236	2	1	5	1	562	576
4237	2	2	5	3	529	537
4238	2	2	5	3	567	602
4239	2	1	5	3	579	594

Appendix K: Fourth-Grade PTR Data

Student	Setting	Gender	Ethnicity	Lunch	Math	Reading
4240	2	1	5	3	585	567
4241	2	2	5	1	539	560
4242	2	2	5	1	529	556
4243	2	2	5	1	539	560
4244	2	2	5	3	597	588
4245	2	1	5	3	610	588
4246	2	1	5	1	507	530
4247	2	1	5	3	589	567
4248	2	1	5	3	530	545
4249	2	1	5	0	554	582
4250	2	1	5	0	528	542
4251	2	2	5	0	597	623
4252	2	2	5	0	528	550
4253	2	1	5	0	542	539
4254	2	2	5	0	548	547
4255	2	2	5	0	548	542
4256	2	2	5	0	526	539
4257	2	1	5	0	528	544
4258	2	2	5	0	554	587
4259	2	2	5	0	536	576
4260	2	1	5	0	536	502
4261	2	2	5	0	581	539
4262	2	1	5	0	526	521
4263	2	2	5	0	526	532
4264	2	1	5	0	536	511
4265	2	2	5	0	564	572
4266	2	1	4	0	533	493
4267	2	2	5	0	522	519
4268	2	1	5	0	603	560
4269	2	1	5	0	556	567
4270	2	2	4	0	559	542
4271	2	1	5	0	500	532
4272	2	1	5	0	522	532
4273	2	2	5	0	603	602
4274	2	1	5	0	557	537
4275	2	2	1	0	551	511
4276	2	1	5	0	553	560

Appendix L: CS-PTR Classroom Observation Form

Type of Room Observed: CS_____ PTR_____

Number of Students in the Classroom:_____

Number of Adults (Teacher & Aides) in the Classroom:_______

Number of Student Desks or Tables in the Classroom:________

Size of Student Desks or Tables:________

Number of Teacher Desks:_________ Size of Teacher Desks:_________

Number of Non-Student Tables:_______ Size of Non-Student Tables:________

Size of Classroom:________ Square Footage:____________

Available Square Footage for Circulation:_________

Duration of Observation:__________

Number of Interruptions:________

Interruptions of Questions:________

- Repeating of Instructions:______
- Clarification of Instructions:______
- Repeating of Content:______
- Clarification of Content:______

Interruptions of Student Conduct:_________

- Individual Student:______
- Student-to-Student:______

Appendix M: Analysis of Third-Grade Reading Z-Scores

Group Statistics

Z-score(READING)	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	44	-.1940194	.7811959	.1177697
	PTR	48	-.1399027	.8560073	.1235540

Independent Samples Test

Levine's Test for Equality of Variances

t test for Equallity of Means

Z-score(READING)		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.067	.796	-.316	90	.753	-5.412E-02	.1713763	-.39455857	.2863523

Appendix N: Analysis of Third-Grade Mathematics Z-Scores

Group Statistics

Z-score(MATH)	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	44	-.1505755	.9451965	.1424937
	PTR	48	1.2642071	1.0242924	.1478439

Independent Samples Test

Z-score(MATH)		Levine's Test for Equality of Variances		t test for Equallity of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.081	.776	-6.866	90	.000	-1.4147825	.2060597	-1.8241561	-1.0054090

Appendix O: Analysis of Female Third-Grade Reading Z-Scores

Group Statistics

Z-Scores Reading	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	20	-.2441365	.7525762	.1682812
	PTR	22	-.2079191	.7820482	.1667332

Independent Samples Test

Levine's Test for Equality of Variances

t test for Equallity of Means

Z-Scores Reading		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.111	.741	-.153	40	.879	-3622E-02	.2373378	-.5158950	.4434602

Appendix P: Analysis of Female Third-Grade Mathematics Z-Scores

Group Statistics

Z-Scores Math	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	20	-.394E-03	.9223550	.2062448
	PTR	22	1.3815564	1.0633807	.2267135

Independent Samples Test

		Levine's Test for Equality of Variances		t test for Equallity of Means						
Z-Scores Math		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.032	.858	-4.489	40	.000	-1.3854979	.3086109	-2.0092238	-.7617720

Appendix Q: Analysis of Male Third-Grade Reading Z-Scores

Group Statistics

Reading Z-Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	24	-.15225	.8179607	.1669655
	PTR	26	-.27277	.9253686	.1814797

Independent Samples Test

		Levine's Test for Equality of Variances		t test for Equallity of Means						
Reading Z-Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.024	.878	-.282	48	.779	-6.990E-02	.2478407	-.5682210	.4284134

Appendix R: Analysis of Male Third-Grade Mathematics Z-Scores

Group Statistics

Math Z-Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	16	-.27277	.7815906	.179532
	PTR	41	1.1649	1.0131814	.132258

Independent Samples Test

		Levine's Test for Equality of Variances		t test for Equallity of Means						
Math Z-Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	1.002	.321	-3.152	55	.003	.89213	.047274	-.387291	.826743

Appendix S: Analysis of Third-Grade Reading Z-Scores for Regular Lunch Students

Group Statistics

Reading Z-Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	28	-.244E-02	.7892205	.1491487
	PTR	37	-.1440127	.9229604	.1517338

Independent Samples Test

		Levine's Test for Equality of Variances		t test for Equallity of Means						
Reading Z-Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.381	.539	.550	63	.584	.1195731	.2174610	-.3149881	.5541342

Appendix T: Analysis of Third-Grade Mathematics Z-Scores for Regular Lunch Students

Group Statistics

Math Z-Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	28	.1104361	.8696219	.1643431
	PTR	37	1.2354389	1.1131958	.1830083

Independent Samples Test

		Levine's Test for Equality of Variances		t test for Equallity of Means						
Math Z-Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	1.348	.250	-4.421	63	.000	-1.1250028	.2544856	-1.6335517	-.6164540

Appendix U: Analysis of Third-Grade Reading Z-Scores for Free and Reduced Lunch Students

Group Statistics

Reading Z-Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	16	-.4907825	.6932896	.1733224
	PTR	11	-.1260782	.6139725	.18512197

Independent Samples Test

Reading Z-Scores		Levine's Test for Equality of Variances: F	Sig.	t test for Equallity of Means: t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.049	.826	-1.405	25	.172	-.3647043	.2595640	-.8992864	.1698777

Appendix V: Analysis of Third-Grade Mathematics Z-Scores for Free and Reduced Lunch Students

Group Statistics

Math Z-Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	16	-.6073444	.9220180	.2305045
	PTR	11	1.3609718	.6757358	.2037420

Independent Samples Test

		Levine's Test for Equality of Variances		t test for Equallity of Means						
Math Z-Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.005	.945	-6.038	25	.000	-1.9683162	.3259896	-2.6397043	-1.2969281

Appendix W: Analysis of Fourth-Grade Reading Scores

Group Statistics

Reading Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	32	540.344	22.367	3.954
	PTR	76	555.329	26.418	3.030

Independent Samples Test

		Levine's Test for Equality of Variances		t test for Equality of Means						
Reading Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	1.059	.306	-2.811	106	.006	-14.985	5.332	-25.556	-4.415

Appendix X: Analysis of Fourth-Grade Mathematics Scores

Group Statistics

Math Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	32	547.500	35.428	6.263
	PTR	76	554.645	28.747	3.298

Independent Samples Test

Levine's Test for Equality of Variances

t test for Equallity of Means

Math Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	3.457	.066	-1.099	106	.274	-7.145	6.501	-20.034	5.745

Appendix Y: Analysis of Female Fourth-Grade Mathematics Scores

Group Statistics

Math Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	16	555.938	31.150	7.788
	PTR	35	558.200	27.194	4.597

Independent Samples Test

Math Scores		Levine's Test for Equality of Variances		t test for Equallity of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.875	.354	-.263	49	.793	-2.263	8.590	-19.524	14.999

Appendix Z: Analysis of Female Fourth-Grade Reading Scores

Group Statistics

Reading Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	16	546.000	22.955	5.739
	PTR	35	551.057	24.349	4.116

Independent Samples Test

Levine's Test for Equality of Variances

t test for Equallity of Means

Reading Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.028	.868	-.700	49	.487	-5.057	7.222	-19.570	9.456

Appendix AA: Analysis of Male Fourth-Grade Mathematics Scores

Group Statistics

Math Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	16	539.063	38.364	9.591
	PTR	41	551.610	30.007	4.686

Independent Samples Test

		Levine's Test for Equality of Variances		t test for Equallity of Means						
Math Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	3.314	.073	-1.310	55	1.96	-12.547	9.580	-31.746	6.651

Appendix AB: Analysis of Male Fourth-Grade Reading Scores

Group Statistics

Reading Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	16	534.688	20.947	5.237
	PTR	41	558.976	27.840	4.348

Independent Samples Test

Reading Scores		Levine's Test for Equality of Variances		t test for Equallity of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	1.002	.321	-3.152	55	.003	-24.288	7.706	-39.731	-8.846

Appendix AC: Analysis of Fourth-Grade Mathematics Scores for Regular Lunch Students

Group Statistics

Math Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	19	562.421	28.040	6.433
	PTR	34	561.853	27.078	4.644

Independent Samples Test

		Levine's Test for Equality of Variances		t test for Equallity of Means						
Math Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.012	.912	.072	51	.943	.568	7.854	-15.200	6.336

Appendix AD: Analysis of Fourth-Grade Reading Scores for Regular Lunch Students

Group Statistics

Reading Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	19	550.421	19.854	4.553
	PTR	34	565.235	25.876	4.438

Independent Samples Test

		Levine's Test for Equality of Variances		t test for Equallity of Means						
Reading Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	1.498	.227	-2.162	51	.035	-14.814	6.852	-28.570	-1.058

Appendix AE: Analysis of Fourth-Grade Mathematics Scores for Free and Reduced Lunch Students

Group Statistics

Math Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	13	525.692	34.575	9.589
	PTR	14	553.357	28.114	7.514

Independent Samples Test

Levine's Test for Equality of Variances

t test for Equallity of Means

Math Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	1.285	.268	-2.289	25	.031	-27.665	12.087	-52.559	-2.771

Appendix AF: Analysis of Fourth-Grade Reading Scores for Free and Reduced Lunch Students

Group Statistics

Reading Scores	SETTING	N	Mean	Std. Deviation	Std. Error Mean
	Small Class	13	525.615	17.491	4.851
	PTR	14	547.857	23.307	6.229

Independent Samples Test

		Levine's Test for Equality of Variances		t test for Equallity of Means						
Reading Scores		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
	Equal variances assumed	.788	.383	-2.787	25	.010	-22.242	7.981	-38.678	-5.806